

Problem Set #1

Profesor: Tomás Rau
Ayudantes: Valentina Andrade y Nicolás Valle

Fecha de Entrega: 12 de Abril

El Problem Set 1 incluye tanto una sección teórica como una empírica. Ambas secciones se deben desarrollar de manera **individual**. Ud. puede intercambiar ideas, no respuestas ni códigos.

Las entregas deben realizarse exclusivamente a través de CANVAS. Cualquier entrega realizada después de la fecha límite será penalizada con 5 décimas por día de atraso.

En cuanto al formato, se requiere la entrega de un archivo comprimido que contenga: 1) documento pdf (escrito preferentemente en Latex) y 2) códigos para replicación (puede utilizar Matlab, R, STATA, Julia). Respuestas sin códigos tendrán la nota mínima. Los códigos entregados deben replicar los resultados de las respuestas.

Parte 1. Teórica (60 puntos)

Parte A

Suponga que un modelo económico sugiere que la variable dependiente latente y_i^* satisface el siguiente modelo clásico lineal

$$y_i^* = x_i' \beta_0 + \varepsilon_i,$$

pero no se puede observar y_i^* sobre todo su rango. Por el contrario, se puede observar una muestra aleatoria de tamaño N de y_i y x_i , donde

$$\begin{aligned} y_i &\equiv \tau_i(y_i^*) \\ &= 0 \quad \text{si } y_i^* \leq 0, \\ &= y_i^* \quad \text{si } 0 < y_i^* \leq L_i, \\ &= L_i \quad \text{si } L_i < y_i^* \leq U_i, \text{ y} \\ &= y_i^* \quad \text{si } U_i < y_i^*. \end{aligned}$$

Esto significa que, la variable latente y_i^* es observada a menos que esta sea menor a cero o en el intervalo (L_i, U_i) . Las variables de umbral L_i y $U_i > L_i > 0$ se asumen observables para todo i , y las dos son funciones lineales de regresores de x_i .

Asuma que ε_i es normalmente distribuida con media cero y con varianza desconocida σ_0^2 .

1. Derive la log-likelihood promedio para los parámetros desconocidos de este modelo. (10 puntos)
2. Discuta cómo la presencia de censura en este modelo (tanto inferior como superior) afecta la estimación de los parámetros β_0 y la interpretación de los efectos marginales de las variables explicativas x_i en la variable dependiente latente y_i^* . (10 puntos)

Parte B

Supongamos ahora que y_i^* nunca es observada, sino que solo se observa el rango en el que "cae". Más específicamente, la variable dependiente y_i ahora se define como

$$\begin{aligned} y_i &\equiv t_i(y_i^*) \\ &= 0 \text{ si } y_i^* \leq 0, \\ &= 1 \text{ si } 0 < y_i^* \leq L_i, \\ &= 2 \text{ si } L_i < y_i^* \leq U_i, \text{ and} \\ &= 3 \text{ si } U_i < y_i^*. \end{aligned}$$

Nuevamente asuma que ε_i es normalmente distribuido con media cero y varianza σ_0^2 desconocida.

1. Derive la log-likelihood promedio para los parámetros desconocidos de este modelo. (10 puntos)
2. ¿Es necesaria una normalización de β_0 o σ^2 , o son todos los parámetros identificables? ¿Por qué sí o por qué no? (10 puntos)

Parte C

Considere el modelo presentado en la parte B, en el que la variable dependiente y_i indica únicamente el rango en el que "cae" la variable dependiente latente y_i^* , clasificándola en una de cuatro categorías: 0, 1, 2, o 3.

A diferencia del caso anterior, en este escenario, los umbrales U_i y L_i que definen estos rangos **son desconocidos**.

1. Analice cómo la incertidumbre sobre los umbrales U_i y L_i afecta la identificación de los parámetros del modelo (β_0 y σ^2). (10 puntos)
2. Proponga estrategias para la estimación de los parámetros del modelo, incluidos U_i , L_i , β_0 y σ^2 , dada la incertidumbre sobre los umbrales. (10 puntos)

Parte 2. Empírica (135 puntos)

El Ministerio de Educación (MINEDUC) está interesado en realizar estudio sobre las trayectorias académicas de los estudiantes que ingresaron a la Educación Superior (ESUP) el año 2018. El ministerio le ha encargado a usted realizar este análisis, y requiere de responder algunas preguntas concretas.

Para hacer estos análisis debe utilizar los datos abiertos del MINEDUC: <https://datosabiertos.mineduc.cl>. En la pestaña estudiantes encontrarán los datos de matrícula, pruebas de admisión, acceso a beneficios y créditos, y titulación.

Para simplificar los análisis, el MINEDUC le indica que solo estudiará la trayectoria universitaria de los estudiantes hasta un año después de su matrícula en la ESUP. Es decir, solo es necesario ocupar los datos del 2018 y 2019 para hacer su informe.

1. El MINEDUC quiere saber qué factores individuales son relevantes para la permanencia de los estudiantes universitarios en la ESUP. Para realizar este ejercicio

- a) Reporte una tabla con el porcentaje de estudiantes que permanecen en la ESUP. (5 puntos)
- b) Estime un modelo de respuesta binaria para la permanencia en la ESUP. Indique cuál es la ecuación a estimar. Puede utilizar cualquiera de las variables disponibles en los datos, **pero solo deben corresponder a regresores que varíen entre las personas**, por ejemplo: edad, sexo y tipo de establecimiento. Justifique su utilización. (10 puntos)
- c) Reporte los efectos marginales del modelo propuesto. Puede evaluarlos en el promedio. (10 puntos)
- d) Interprete sus resultados. (10 puntos)

2. Un asesor del MINEDUC plantea que las trayectorias educacionales son más complejas que solo permanecer en la ESUP. Se plantea que los estudiantes pueden **permanecer, cambiarse o desertar** de la ESUP. Para realizar estos análisis

- a) Reporte una tabla con el porcentaje de estudiantes en cada una de las categorías. Preocúpese que el número final de observaciones sean coherentes con la población de estudio. (5 puntos)
- b) Estime un modelo multinomial (Logit o Probit). Justifique su uso. (10 puntos)
- c) Reporte los efectos marginales de esta regresión e interprete los resultados. (10 puntos)
- d) Discuta sobre el supuesto de IIA y la factibilidad de que este supuesto se cumpla en este caso. (5 puntos)
- e) ¿Son comparables estos resultados con los del inciso anterior? Explique. (5 puntos)

3. El Subsecretario de ESUP está preocupado de que su modelo solo esté utilizando regresores que varían entre individuos. Por ello, le indica que debe utilizar regresores que varíen con las alternativas (permanecer, cambiarse y desertar) para entender la decisión de los estudiantes. En particular, le indican que existen dos regresores que varían entre alternativas

- Tiempo en realizar el trámite: si el estudiante decide desertar el trámite dura 1 semana, si decide cambiarse 2 días, y si decide permanecer no gasta tiempo.

- Impuesto del Estado al tomar la decisión: si el estudiante decide desertar debe pagar CLP 100.000, si se cambia CLP 30.000 y si permanece no debe pagar nada

Para resolver este problema

- a) Desde la teoría económica, discuta ventajas y desventajas que tiene una especificación como la planteada por el Subsecretario. (5 puntos)
- b) Cree los regresores propuestos por el MINEDUC en base a la siguiente tabla (5 puntos):

	Permanecer	Cambiarse	Desertar
Tiempo	0 días	2 días	7 días
Impuesto	0 CLP	30.000 CLP	100.000 CLP

- c) Estime un modelo Conditional (Logit o Probit). Justifique su uso. (10 puntos)
 - d) Reporte los efectos marginales del modelo propuesto e interprete los resultados. (10 puntos)
4. El Ministro de Educación no está convencido del ejercicio propuesto por el asesor y el Subsecretario. Él plantea que es necesario estimar la decisión de los estudiantes ocupando **ambos tipos de regresores**: aquellos que varían con la alternativa y persona.

Para realizar este ejercicio

- a) Discuta las ventajas y desventajas de utilizar esta especificación. (5 puntos)
- b) Estime un modelo Mixed (Logit o Probit). Justifique su uso. (10 puntos)
- c) Reporte los efectos marginales del modelo propuesto e interprete los resultados. (10 puntos)
- d) Compare sus resultados con los incisos 3 y 4. (10 puntos)