

Métodos Cuantitativos II

Tarea 2: Pauta

Profesor: Andrés Sagner

Ayudantes: Hiarela Aravena^{*}, Agustín Farías^{**}, Nicolás Leiva^{***} y Karl Vidal^{****}

Primavera 2023

Instrucciones

- El plazo de entrega de esta tarea es el **domingo 3 de septiembre hasta las 23:59 hrs.** Pasado este límite, la tarea tendrá una penalización de 1 punto, en la escala de 1.0 a 7.0, por día de atraso, con un máximo de 3 días. Posterior a este plazo, no se recibirán tareas y la calificación será un 1.0.
- Cada alumno debe subir su reporte INDIVIDUAL en formato electrónico en la plataforma CANVAS.
- La extensión del reporte no debe sobrepasar las 5 páginas. En este sentido, las respuestas deben ser concisas.
- No es necesario el envío de códigos o las bases de datos empleados para el desarrollo de la tarea.
- Si bien no existe un formato de entrega específico requerido, se exige que la entrega del informe sea pulcra. En este sentido, no entregar *screenshots* de códigos, tablas, desarrollos a mano, etc.. Las tablas y/o el desarrollo de ecuaciones se debe hacer con un procesador de texto para mayor claridad y para facilitar la corrección de esta evaluación (Por ejemplo L^AT_EX o Word).

Los documentos y bases de datos a utilizar en esta tarea se encuentran [aquí](#).

^{*} haravenaro@fen.uchile.cl

^{**} afariasl@fen.uchile.cl

^{***} nleivad@fen.uchile.cl

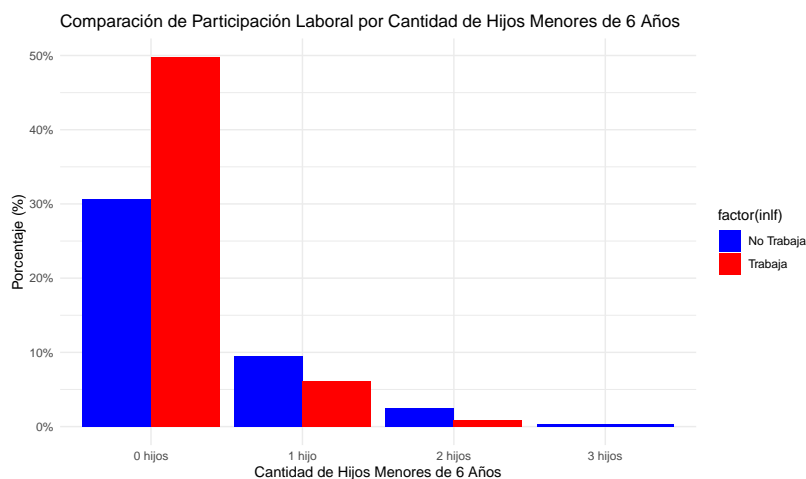
^{****} kvidalm@fen.uchile.cl

1. Participación laboral de las mujeres casadas en Estados Unidos: Modelos Probit y Logit

La base de datos Mroz es un conjunto de datos que contiene información sobre las decisiones laborales de las mujeres casadas en Estados Unidos en el año 1975. La base de datos fue utilizada por el economista Thomas A. Mroz en su artículo “The Sensitivity of an Empirical Model of Married Women’s Hours of Work to Economic and Statistical Assumptions” publicado en 1987. La base de datos tiene 753 observaciones y 22 variables, entre las que se encuentran el salario por hora, las horas trabajadas, la educación, la edad, el número de hijos, el ingreso no laboral y la participación laboral. Para realizar esta tarea, deberá descargar la base *MROZ.csv*, que se encuentra disponible en el siguiente [link](#). Responda las siguientes preguntas:

1. Genere un gráfico de barras que ilustre cómo varía la participación laboral de mujeres casadas en Estados Unidos en 1975 en relación a la cantidad de hijos menores de 6 años. Cada barra en el gráfico deberá representar la proporción de mujeres que trabajan y no trabajan, con el eje horizontal mostrando diferentes categorías de cantidad de hijos menores de 6 años y el eje vertical expresado en porcentaje. ¿Qué puede inferir acerca de las diferencias en las proporciones observadas en el gráfico? (5 puntos)

Respuesta:



Generar el gráfico correctamente con los datos de la base Mroz (2 puntos).

Se puede inferir que la participación laboral de las mujeres casadas en Estados Unidos en 1975 está inversamente relacionada con la cantidad de hijos menores de 6 años. A medida que aumenta el número de hijos menores de 6 años, disminuye el porcentaje de mujeres que trabajan y aumenta el porcentaje de mujeres que no trabajan. Esto sugiere que las mujeres casadas con hijos pequeños enfrentan mayores costos de oportunidad y/o restricciones para ingresar al mercado laboral, como por ejemplo el cuidado infantil, la preferencia por el hogar o la división del trabajo familiar (3 puntos).

2. Estime un modelo de probabilidad lineal, en el cual la variable dependiente sea la participación laboral de las mujeres casadas, que es una variable dicotómica que toma el valor 1 si la mujer trabaja y 0 si no. La variable explicativa principal es el ingreso no laboral de la familia (*nwifeinc*), que mide el efecto de la renta sobre la decisión laboral. Esta variable representa la suma de los ingresos del esposo y de otras fuentes distintas al trabajo de la mujer. Además incorpore como regresores: el nivel educativo, la experiencia laboral, la edad y el número de hijos menores de 6 años¹. Presente los resultados en una tabla e interprete los coeficientes estimados. (10 puntos)

Respuesta:

Estimar correctamente el modelo con los datos de la base Mroz y presentar los resultados en una tabla (4 puntos).

Para interpretar los coeficientes estimados, se debe tener en cuenta que en el modelo de probabilidad lineal, el valor predicho de la variable dependiente se interpreta como la probabilidad predicha de que la variable dependiente sea igual a 1, y el coeficiente estimado de cada variable independiente se interpreta como el cambio en la probabilidad predicha producto de un cambio unitario en dicha variable, manteniendo constantes las demás variables.

- **nwifeinc:** El coeficiente estimado de *nwifeinc* es -0.003, lo que significa que un aumento de una unidad en el ingreso no laboral de la familia reduce la probabilidad predicha de que la mujer casada trabaje en 0.003 puntos porcentuales, manteniendo constantes las demás variables.
- **educ:** El coeficiente estimado de *educ* es 0.039, lo que significa que un aumento de un año en el nivel educativo aumenta la probabilidad predicha de que la mujer casada trabaje en 0.039 puntos porcentuales, manteniendo constantes las demás variables.
- **exper:** El coeficiente estimado de *exper* es 0.022, lo que significa que un aumento de un año en la experiencia laboral aumenta la probabilidad predicha de que la mujer casada trabaje en 0.022 puntos porcentuales, manteniendo constantes las demás variables.
- **age:** El coeficiente estimado de *age* es -0.019, lo que significa que un aumento de un año en la edad reduce la probabilidad predicha de que la mujer casada trabaje en 0.019 puntos porcentuales, manteniendo constantes las demás variables.
- **kidslt6:** El coeficiente estimado de *kidslt6* es -0.275, lo que significa que un aumento de un hijo menor de 6 años reduce la probabilidad predicha de que la mujer casada trabaje en 0.275 puntos porcentuales, manteniendo constantes las demás variables.
- (6 puntos).

¹Para encontrar las variables que se adaptan mejor a lo que se pide, consulte el libro de códigos.

Cuadro 1: Modelo de probabilidad lineal

<i>Dependent variable:</i>	
inlf	
nwifeinc	-0.003** (0.001)
educ	0.039*** (0.007)
exper	0.022*** (0.002)
age	-0.019*** (0.002)
kidslt6	-0.275*** (0.033)
Constant	0.770*** (0.135)
Observations	753
R ²	0.253
Adjusted R ²	0.248
Residual Std. Error	0.430 (df = 747)
F Statistic	50.611*** (df = 5; 747)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

3. Calcule e interprete los efectos marginales del modelo de probabilidad lineal estimado. Además, comente la significancia de los efectos marginales. Muestre los resultados en una tabla. (6 puntos)

Respuesta:

Calcular correctamente los efectos marginales con los datos y presentar los resultados en una tabla (2 puntos).

Los efectos marginales son el cambio en la probabilidad predicha de que la variable dependiente sea igual a 1 producto de un cambio unitario en cada variable independiente, manteniendo constantes las demás variables. En el modelo de probabilidad lineal, los efectos marginales son iguales a los coeficientes estimados (2 puntos). Por ejemplo:

El efecto marginal de nwifeinc es -0.003, lo que significa que un aumento de una unidad en el ingreso no laboral de la familia reduce la probabilidad predicha de que la mujer casada trabaje en 0.003 puntos porcentuales, manteniendo constantes las demás variables.

Todos los efectos marginales son significativos estadísticamente al nivel del 5%, ya que sus valores p son menores que 0.05. Esto implica que todas las variables independientes tienen un efecto significativo sobre la probabilidad predicha de que la mujer casada trabaje (2 puntos).

Cuadro 2: Efectos marginales del modelo de probabilidad lineal

Variable	Efecto Marginal
(Intercept)	0.770
nwifeinc	-0.003
educ	0.039
exper	0.022
age	-0.019
kidslt6	-0.275

4. Estime un modelo Probit para la participación laboral de las mujeres casadas en Estados Unidos utilizando todas las variables explicativas mencionadas en la pregunta anterior, y presente sus estimaciones en una tabla. Discuta e interprete sus resultados para las variables. (10 puntos)

Respuesta:

Estimar correctamente el modelo Probit y presentar los resultados en una tabla (4 puntos).

Para interpretar los coeficientes estimados, se debe tener en cuenta que en el modelo Probit, el valor predicho de la variable dependiente se interpreta como la probabilidad predicha de que la variable dependiente sea igual a 1, y el coeficiente estimado de cada variable independiente se interpreta como el cambio en el índice de utilidad latente producto de un cambio unitario en dicha variable, manteniendo constantes las demás variables. Sin embargo, estos cambios no tienen una interpretación directa en términos de probabilidades, por lo que se requiere calcular los efectos marginales para obtener una medida más intuitiva del impacto de cada variable.

- **nwifeinc:** El coeficiente estimado de nwifeinc es -0.011, lo que significa que un aumento de una unidad en el ingreso no laboral de la familia reduce el índice de utilidad latente de trabajar en 0.011 unidades, manteniendo constantes las demás variables.
- **educ:** El coeficiente estimado de educ es 0.132, lo que significa que un aumento de un año en el nivel educativo aumenta el índice de utilidad latente de trabajar en 0.132 unidades, manteniendo constantes las demás variables.
- **exper:** El coeficiente estimado de exper es 0.069, lo que significa que un aumento de un año en la experiencia laboral aumenta el índice de utilidad latente de trabajar en 0.069 unidades, manteniendo constantes las demás variables.
- **age:** El coeficiente estimado de age es -0.058, lo que significa que un aumento de un año en la edad reduce el índice de utilidad latente de trabajar en 0.058 unidades, manteniendo constantes las demás variables.
- **kidslt6:** El coeficiente estimado de kidslt6 es -0.886, lo que significa que un aumento de un hijo menor de 6 años reduce el índice de utilidad latente de trabajar en 0.886 unidades, manteniendo constantes las demás variables.

- (6 puntos).

Todos los coeficientes estimados son significativos estadísticamente al nivel del 5 %, ya que sus valores p son menores que 0.05. Esto implica que todas las variables independientes tienen un efecto significativo sobre el índice de utilidad latente de trabajar.

Cuadro 3: Modelo Probit

	Dependent variable:
	inlf
nwifeinc	-0.011** (0.005)
educ	0.132*** (0.025)
exper	0.069*** (0.008)
age	-0.058*** (0.008)
kidslt6	-0.886*** (0.117)
Constant	0.765* (0.439)
Observations	753
Log Likelihood	-406.542
Akaike Inf. Crit.	825.084
Note: *p<0.1; **p<0.05; ***p<0.01	

5. En base al modelo Probit estimado, calcule e interprete los efectos marginales de cada variable, evaluados en la media, y comente si son significativos estadísticamente. Presente los resultados en una tabla. (6 puntos)

Respuesta:

Calcular correctamente los efectos marginales evaluados a la media del modelo Probit y presentar los resultados en una tabla (2 puntos).

En el modelo Probit, los efectos marginales no son iguales a los coeficientes estimados, sino que dependen de la función de distribución acumulada normal estándar y de los valores de las variables. Para interpretar los efectos marginales, se debe tener en cuenta que en el modelo Probit, el valor predicho de la variable dependiente se interpreta como la probabilidad predicha de que la variable dependiente sea igual a 1, y el efecto marginal de cada variable independiente se interpreta como el cambio en la probabilidad predicha producto de un cambio unitario en dicha variable, manteniendo constantes las demás variables y evaluando las variables en sus medias muestrales (2 puntos). Por ejemplo:

El efecto marginal de nwifeinc es -0.004, lo que significa que un aumento de una unidad en el ingreso no laboral de la familia reduce la probabilidad predicha de que la mujer casada trabaje

en 0.004 puntos porcentuales, manteniendo constantes las demás variables y evaluando las variables en sus medias muestrales.

Todos los efectos marginales son significativos estadísticamente al nivel del 5 %, ya que sus valores p son menores que 0.05. Esto implica que todas las variables independientes tienen un efecto significativo sobre la probabilidad predicha de que la mujer casada trabaje, evaluadas en la media (2 puntos).

Cuadro 4: Efectos marginales del modelo Probit (evaluados en la media)

Variable	Efecto Marginal
nwifeinc	-0.004
educ	0.051
exper	0.027
age	-0.023
kidslt6	-0.346

6. Estime un modelo Logit, que incluya las mismas variables que los modelos anteriores. Presente los resultados en una tabla, interprete los coeficientes estimados y comente la significancia del modelo. (10 puntos)

Respuesta:

Estimar correctamente el modelo Logit y presentar los resultados en una tabla (4 puntos).

Para interpretar los coeficientes estimados, es fundamental tener en cuenta que en el modelo Logit, el valor predicho de la variable dependiente se corresponde con la probabilidad estimada de que la variable dependiente sea igual a 1. Además, cada coeficiente estimado de las variables independientes refleja el cambio en el índice de utilidad latente debido a un incremento unitario en esa variable, manteniendo las demás variables constantes (4 puntos). Por ejemplo:

El coeficiente estimado de educ es 0.223, lo que significa que un aumento de un año en el nivel educativo aumenta el índice de utilidad latente de trabajar en 0.223 unidades, manteniendo constantes las demás variables.

Todos los coeficientes estimados son significativos estadísticamente al nivel del 5 %, ya que sus valores p son menores que 0.05. Esto implica que todas las variables independientes tienen un efecto significativo sobre el índice de utilidad latente de trabajar (2 puntos).

Cuadro 5: Modelo Logit

	<i>Dependent variable:</i>
	inlf
nwifeinc	-0.020** (0.008)
educ	0.223*** (0.043)
exper	0.118*** (0.013)
age	-0.095*** (0.013)
kidslt6	-1.464*** (0.200)
Constant	1.153 (0.742)
Observations	753
Log Likelihood	-406.458
Akaike Inf. Crit.	824.917
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

7. Calcule e interprete los efectos marginales del modelo Logit estimado. Muestre los resultados en una tabla. (6 puntos)

Respuesta:

Calcular correctamente los efectos marginales evaluados a la media del modelo Logit y presentar los resultados en una tabla (2 puntos).

En el modelo Logit, los efectos marginales no son iguales a los coeficientes estimados, sino que dependen de la función de distribución acumulada logística estándar y de los valores de las variables. Para comprender los efectos marginales, es esencial considerar que en el modelo Logit, el valor estimado de la variable dependiente se interpreta como la probabilidad estimada de que la variable dependiente sea igual a 1. Del mismo modo, el efecto marginal de cada variable independiente se interpreta como la variación en la probabilidad estimada resultado de un incremento unitario en dicha variable. Esto ocurre al mantener constantes las restantes variables y al evaluar las variables en sus valores medios muestrales (4 puntos). Por ejemplo:

El efecto marginal de age es -0.023, lo que significa que un aumento de un año en la edad reduce la probabilidad predicha de que la mujer casada trabaje en 0.023 puntos porcentuales, manteniendo constantes las demás variables y evaluando las variables en sus medias muestrales.

Todos los efectos marginales son significativos estadísticamente al nivel del 5%, ya que sus valores p son menores que 0.05. Esto implica que todas las variables independientes tienen un efecto significativo sobre la probabilidad predicha de que la mujer casada trabaje, evaluadas en la media.

Cuadro 6: Efectos marginales del modelo Logit (evaluados en la media)

Variable	Efecto Marginal
nwifeinc	-0.005
educ	0.054
exper	0.029
age	-0.023
kidslt6	-0.355

8. Compare los resultados de los tres modelos estimados con la base de datos Mroz: el modelo de probabilidad lineal, el modelo Probit y el modelo Logit. Explique por qué son diferentes los resultados entre los tres modelos, considerando las suposiciones, las formas funcionales y las distribuciones que asume cada modelo. Analice tanto los coeficientes estimados como los efectos marginales de cada modelo. Utilizando el estadístico pseudo- R^2 y la matriz de confusión para predecir la participación o no de las mujeres casadas en el mercado laboral de Estados Unidos, ¿cuál modelo se ajusta mejor a los datos y por qué? (7 puntos)

Respuesta:

Los resultados entre los tres modelos son diferentes porque cada modelo tiene suposiciones, formas funcionales y distribuciones distintas. El modelo de probabilidad lineal (MPL) para la estimación utiliza mínimos cuadrados ordinarios (MCO). El problema de este modelo es que puede dar valores predichos fuera del intervalo $[0,1]$, lo que no tiene sentido para una probabilidad. El modelo Probit se basa en una distribución acumulada normal estándar vinculada a una variable latente que representa la utilidad de trabajar o no. El modelo Probit se estima maximizando la función de verosimilitud, que es la probabilidad conjunta de observar los valores de la variable dependiente dados los valores de las variables independientes. El modelo Logit se basa en una distribución acumulada logística estándar vinculada a una variable latente que representa la utilidad de trabajar o no. El modelo Logit se estima maximizando la función de verosimilitud, que es la misma que en el modelo Probit, pero con una función de distribución diferente.

(2 puntos)

En los modelos Probit y Logit, el coeficiente estimado de cada variable independiente se interpreta como el cambio en el índice de utilidad latente producto de un cambio unitario en dicha variable, manteniendo constantes las demás variables. Sin embargo, estos cambios no tienen una interpretación directa en términos de probabilidades, por lo que se requiere calcular los efectos marginales para obtener una medida más intuitiva del impacto de cada variable.

Para analizar los efectos marginales, se debe tener en cuenta que son el cambio en la probabilidad predicha producto de un cambio unitario en cada variable independiente, manteniendo constantes las demás variables. En el modelo de probabilidad lineal, los efectos marginales son iguales a los coeficientes estimados. En los modelos Probit y Logit, los efectos marginales dependen de la función de distribución acumulada y de los valores de las variables.

(2 puntos)

Para elegir el mejor modelo entre los tres, se debe observar el valor del estadístico pseudo-R² y los elementos de la matriz de confusión. Un mayor valor del estadístico pseudo-R² indica un mejor ajuste del modelo a los datos. Un mayor número de verdaderos positivos y verdaderos negativos indica una mayor precisión del modelo para predecir correctamente la variable dependiente. Un menor número de falsos positivos y falsos negativos indica una menor tasa de error del modelo para predecir incorrectamente la variable dependiente.

Al evaluar los resultados de los modelos, observamos que tanto el modelo Logit como el modelo Probit presentan métricas similares en términos de precisión, sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. Estas métricas indican que ambos modelos tienen una capacidad razonable para predecir la participación laboral de mujeres casadas en el mercado laboral. En contraste, el modelo de probabilidad lineal (LPM) muestra un desempeño inferior en comparación con los modelos Logit y Probit. Sus valores de sensibilidad y valor predictivo positivo son significativamente más bajos, lo que indica una menor capacidad para identificar correctamente los casos de participación laboral.

En base a estos resultados, podemos concluir que tanto el modelo Logit como el modelo Probit son opciones adecuadas para predecir la participación laboral de mujeres casadas en el mercado laboral de Estados Unidos. Dado que ambas metodologías logran un rendimiento similar en términos de métricas de evaluación, la elección entre ellas puede depender de consideraciones teóricas o preferencias metodológicas. Por otro lado, el modelo de probabilidad lineal (LPM) parece no ser la mejor opción en este caso, ya que muestra un rendimiento inferior en la mayoría de las métricas evaluadas.

Sin embargo, no hay una forma definitiva de decidir cuál de los dos modelos, Probit o Logit, es preferible, ya que ambos tienen supuestos diferentes sobre la distribución de la variable latente que afecta la probabilidad de trabajar. Algunas recomendaciones sugieren usar el modelo Logit si se cree que la distribución es más aguda que la normal, y usar el modelo Probit si se cree que la distribución es similar o más plana que la normal. También se pueden considerar otros factores, como la facilidad de interpretación o la disponibilidad computacional.

(3 puntos)

Cuadro 7: Matriz de Confusión: Modelo de probabilidad lineal

Clase Real / Clase Predicha	0	1
0	210	115
1	83	345

Cuadro 8: Matriz de Confusión: Modelo Probit

Clase Real / Clase Predicha	0	1
0	210	115
1	83	346

Cuadro 9: Matriz de Confusión: Modelo Logit

Clase Real / Clase Predicha	0	1
0	211	114
1	83	345