

**Taller #8
Econometría 06216**

Profesor: Julio César Alonso
Monitor: Manuel Serna Cortés

Notas:

- o Recuerde que sólo tres preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

El departamento de seguimiento laboral de cierta universidad, desea conocer que variables incrementan la probabilidad de que un egresado consiga un puesto gerencial en una multinacional. Para esto, se realizó una encuesta en entre 6702 egresados de sus carreras, obteniendo así los siguientes datos:

- Gerencial: Toma el valor de 1 cuando el individuo i tenga un cargo gerencial en una multinacional y 0 en caso contrario.
- S.Público: Número de puestos de trabajo en este sector que el individuo i ha tenido.
- Experiencia: Número de años de experiencia laboral que el individuo i ha tenido.
- Inversión: Valor en miles dólares de la inversión en estudios que ha realizado la persona a lo largo de su vida.
- Carreras: Número de carreras que el individuo i ha cursado.
- Edad
- Educación: años de educación que el individuo i ha cursado.
- TOEFL: Toma el valor de 1 si el individuo i obtuvo un puntaje del TOEFL sobresaliente y 0 en caso contrario.
- Casado: Toma el valor de 1 si el individuo i es casado y 0 en caso contrario
- Género: Toma el valor de 1 si el individuo i es mujer y 0 en caso contrario.
- Internacional: Toma el valor de 1 si el individuo i ha tenido alguna experiencia de estudio o trabajo internacional y 0 en el caso contrario.
- Honor: Número de condecoraciones académicas que el individuo i ha tenido en su formación.

El grupo de investigadores desea estimar un modelo para explicar los determinantes de si un individuo obtiene o no un cargo gerencial en una multinacional. Empleando la información disponible, responda las siguientes preguntas:

1. Uno de los investigadores del departamento propone asumir una función de densidad acumulativa normal para el modelo por estimar. Usted como su asistente, estime el modelo que cumpla con los requerimientos del investigador y reporte sus resultados en una tabla. No olvide mostrar en ésta cuál es la razón de máxima verosimilitud y el logaritmo de la función de máxima verosimilitud. Comente sobre la significancia conjunta e individual de los coeficientes.

2. De acuerdo a sus estimaciones responda las siguientes preguntas:
 - a) Un compañero asistente, analiza sus resultados e interpreta los coeficientes estimados del modelo de acuerdo con su significancia. Justifique por qué es correcto o no interpretar los coeficientes del modelo. En caso de estar en desacuerdo, encuentre los resultados del modelo que si se pueden interpretar para cada una de las variables y repórtelos en la Tabla que realizó en el punto anterior.
 - b) Otro compañero concluye que la variable más decisiva para obtener un cargo gerencial en una multinacional es la experiencia internacional y que la que perjudica más la probabilidad de obtener este cargo es el género. ¿Usted está de acuerdo con esta conclusión? Justifique. .
3. Un estudio similar realizado por una universidad vecina muestra que el efecto marginal promedio de una carrera adicional sobre la probabilidad de obtener un cargo gerencial es igual que el efecto marginal promedio de un punto adicional en el puntaje de TOEFL. Lo mismo se evidenció para los efectos de la variable edad y la variable educación. Pruebe si para este estudio se pueden hacer las mismas conclusiones de acuerdo con el modelo que estimó en el punto anterior.
4. Paralelamente, otra división del departamento realizó la estimación de un modelo que supone una función de distribución logística. Usted no está seguro de los resultados obtenidos por ellos, realice la estimación de ese modelo y reporte sus resultados en una tabla. Compare estos resultados con los obtenidos en el punto 1, tanto en su valor como su significancia. ¿Se comprueba la hipótesis de Amemiya?
5. De acuerdo con los resultados obtenidos en el punto anterior, responda las siguientes preguntas:
 - a) La interpretación de los resultados de este modelo en cuanto a las variables que más inciden positiva y negativamente en la probabilidad de obtener un cargo gerencial coinciden con los obtenidos en el literal b del punto 2.
 - b) Los resultados obtenidos por la universidad vecina descritos en el punto 3 se cumplen para este modelo. Justifique.
6. Argumente por qué teóricamente es correcto su análisis anterior con los modelos empleados. ¿Qué aportan estos modelos respecto a modelo de probabilidad lineal y por qué sería correcto o no estimar la ecuación empleada en la investigación mediante este último modelo?

Taller #8
Respuestas Sugeridas
Econometría 06216

Profesor: Julio César Alonso
Monitor: Manuel Serna Cortés

Notas:

- Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- Este taller es para ser entregado en los primeros 10 minutos de la clase.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

El departamento de seguimiento laboral de cierta universidad, desea conocer que variables incrementan la probabilidad de que un egresado consiga un puesto gerencial en una multinacional. Para esto, se realizó una encuesta en entre 6702 egresados de sus carreras, obteniendo así los siguientes datos:

- Gerencial: Toma el valor de 1 cuando el individuo i tenga un cargo gerencial en una multinacional y 0 en caso contrario.
- S.Público: Número de puestos de trabajo en este sector que el individuo i ha tenido.
- Experiencia: Número de años de experiencia laboral que el individuo i ha tenido.
- Inversión: Valor en miles dólares de la inversión en estudios que ha realizado la persona a lo largo de su vida.
- Carreras: Número de carreras que el individuo i ha cursado.
- Edad
- Educación: años de educación que el individuo i ha cursado.
- TOEFL: Toma el valor de 1 si el individuo i obtuvo un puntaje del TOEFL sobresaliente y 0 en caso contrario.
- Casado: Toma el valor de 1 si el individuo i es casado y 0 en caso contrario.
- Género: Toma el valor de 1 si el individuo i es mujer y 0 en caso contrario.
- Internacional: Toma el valor de 1 si el individuo i ha tenido alguna experiencia de estudio o trabajo internacional y 0 en el caso contrario.
- Honor: Número de condecoraciones académicas que el individuo i ha tenido en su formación.

El grupo de investigadores desea estimar un modelo para explicar los determinantes de si un individuo obtiene o no un cargo gerencial en una multinacional. Empleando la información disponible, responda las siguientes preguntas:

1. Uno de los investigadores del departamento propone asumir una función de densidad acumulativa normal para el modelo por estimar. Usted como su asistente, estime el modelo que cumpla con los requerimientos del investigador y reporte sus resultados en una tabla. No olvide mostrar en ésta cuál es la razón de máxima

verosimilitud y el logaritmo de la función de máxima verosimilitud. Comente sobre la significancia conjunta e individual de los coeficientes.

Se debe estimar un modelo Probit, pues éste asume una distribución acumulativa normal.

Este modelo se puede expresar de la forma:

$$P(\text{gerencial} = 1 | x_i) = \Phi(\beta^T x_i^T)$$

Donde $\Phi(z) = \int_{-\infty}^z \phi(v) dv = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{v^2}{2}} dv$, es decir, la función de densidad de la normal estándar, x_i corresponden al vector fila compuesto por algunas características de los individuos profesionales, y por consiguiente, el método de máximo verosimilitud para la estimación de los coeficientes viene definido por:

$$\begin{aligned} \text{Max}_{\beta} P(y|X) &= \text{Max}_{\beta} \left[\prod_{i=1}^n \Phi(\hat{\beta}^T x_i^T) \right] && \text{equivalente} && \text{a} \\ \text{Max}_{\beta} l(y|X) &= \text{Max}_{\beta} \left[\sum_{i=1}^n \ln[\Phi(\hat{\beta}^T x_i^T)] \right] \end{aligned}$$

Los resultados se muestran en la Tabla 1, al final del taller.

Los coeficientes estimados son significativos conjunta e individualmente, lo que se evidencia en los p-valores pues en la mayoría de los casos se encuentran entre el 0 y el 1% de significancia. Por su parte, la prueba de Wald, al tener un valor tan elevado, nos permite rechazar la hipótesis nula de que todos los coeficientes asociados a variables explicativas son conjuntamente iguales a cero.

2. De acuerdo a sus estimaciones responda las siguientes preguntas:
 - a) Un compañero asistente, analiza sus resultados e interpreta los coeficientes estimados del modelo de acuerdo con su significancia. Justifique por qué es correcto o no interpretar los coeficientes del modelo. En caso de estar en desacuerdo, encuentre los resultados del modelo que si se pueden interpretar para cada una de las variables y repórtelos en la Tabla que realizó en el punto anterior.
 - b) Otro compañero concluye que la variable más decisiva para obtener un cargo gerencial en una multinacional es la experiencia internacional y que la que perjudica más la probabilidad de obtener este cargo es el género. ¿Usted está de acuerdo con esta conclusión? Justifique. .

Respuesta sugerida:

a) Es incorrecto interpretar los coeficientes pues corresponden a un componente del efecto marginal de la respectiva variable sobre la probabilidad. Para tener una interpretación adecuada, se debe considerar la interpretación de los efectos marginales que resultan de emplear las fórmulas vistas en clase. Pero, es importante resaltar que el efecto marginal será diferente para cada individuo, por eso se acostumbra interpretar el promedio de los efectos marginales. Estos efectos marginales promedio se encuentran en la Tabla 1.

b) Según los resultados obtenidos en la Tabla 1, se encuentra que la variable que más incrementa la probabilidad de obtener un puesto gerencial (con 14.11 puntos porcentuales) cuando se incrementa una unidad de ella es la variable *Casado*, es decir que en promedio, si un profesional se casa, se aumenta su probabilidad de obtener un puesto gerencial. Por tanto, el compañero se equivoca al concluir como la más determinante a la variable experiencia internacional.

Por otra parte, la variable que más disminuye (con 17.86 puntos porcentuales) la probabilidad de ser propietario es la variable género, es decir, que cuando el individuo es mujer, entonces se la probabilidad de obtener un cargo gerencial disminuye en 17.86 puntos porcentuales. Por tanto, el compañero está en lo correcto al concluir sobre la variable género.

- Un estudio similar realizado por una universidad vecina muestra que el efecto marginal promedio de una carrera adicional sobre la probabilidad de obtener un cargo gerencial es igual que el efecto marginal promedio de un punto adicional en el puntaje de TOEFL. Lo mismo se evidenció para los efectos de la variable edad y la variable educación. Pruebe si para este estudio se pueden hacer las mismas conclusiones de acuerdo con el modelo que estimó en el punto anterior.

Respuesta sugerida:

Es importante aclarar que las afirmaciones se hacen sobre los efectos marginales promedio y no sobre los coeficientes, como normalmente se trabaja, no obstante, la prueba de hipótesis es equivalente pues tenemos que:

El efecto marginal es:

$$\frac{\partial[y_i | \mathbf{x}_i]}{\partial X_j} = \phi(z_i) \beta_j$$

Y en promedio es:

$$\frac{\sum_{i=1}^n \phi(z_i) \beta_j}{n}$$

Por tanto debemos comparar el efecto marginal promedio de la variable carreras con la variable TOEFL que corresponden a los coeficientes β_4 y β_6 respectivamente.

$$\frac{\sum_{i=1}^n \phi(z_i) \beta_4}{n} = \frac{\sum_{i=1}^n \phi(z_i) \beta_6}{n}$$

$$\sum_{i=1}^n \phi(z_i) \beta_4 = \sum_{i=1}^n \phi(z_i) \beta_6$$

$$\beta_4 \sum_{i=1}^n \phi(z_i) = \beta_6 \sum_{i=1}^n \phi(z_i)$$

De esta manera nuestra prueba de hipótesis corresponde a:

$$H_0: \beta_4 = \beta_6$$

Ha: No Ho.

El resultado de esta prueba con un test de Wald, muestra que el estadístico es 0.37 con un p-valor de 0.54031, concluyendo así que no existe suficiente evidencia para afirmar que los efectos marginales promedio de las variables carreras y TOEFL son estadísticamente diferentes.

Para la edad y educación tenemos:

Ho: $\beta_2 = \beta_5$

Ha: No Ho.

El resultado de esta prueba con un test de Wald, muestra que el estadístico es 1.65 con un p-valor de 0.19912, concluyendo así que no existe suficiente evidencia para afirmar que los efectos marginales promedio de las variables edad y educación son estadísticamente diferentes.

4. Paralelamente, otra división del departamento realizó la estimación de un modelo que supone una función de distribución logística. Usted no está seguro de los resultados obtenidos por ellos, realice la estimación de ese modelo y reporte sus resultados en una tabla. Compare estos resultados con los obtenidos en el punto 1, tanto en su valor como su significancia. ¿Se comprueba la hipótesis de Amemiya?

Recuerden que la diferencia fundamental entre los modelos Logit y Probit es el supuesto sobre la distribución de los errores (normal en Probit, Logística en Logit). Así, tenemos:

$$P(\text{gerencial} = 1 | x_i) = \Lambda(\beta^T x_i^T)$$

Donde $\Lambda(z) = \frac{e^z}{1 + e^z}$, es decir, la función de densidad logística, x_i corresponden al vector fila compuesto por las mismas características de los individuos profesionales i que fueron consideradas en el ejercicio anterior, y por consiguiente, el método de máximo verosimilitud para la estimación de los coeficientes viene definido por:

$$\text{Max}_{\hat{\beta}} P(y|X) = \text{Max}_{\hat{\beta}} \left[\prod_{i=1}^n \Lambda(\hat{\beta}^T x_i^T) \right]$$

$$\text{equivalente a } \text{Max}_{\hat{\beta}} l(y|X) = \text{Max}_{\hat{\beta}} \left[\prod_{i=1}^n \ln[\Lambda(\hat{\beta}^T x_i^T)] \right]$$

Los resultados se muestran en la Tabla 1.

La comparación entre los estimadores del modelo Probit y el Logit, arroja que en la mayoría de las ocasiones los coeficientes de Probit son menores a los de Logit. Igualmente, se observa que los estimadores conservan el mismo signo así como su significancia. En cuanto a Amemiya, se encuentra que sí parecen comportarse de la forma: $\hat{\beta}_{\text{Probit}} \approx 0.625 \hat{\beta}_{\text{Logit}}$, pues el promedio de las diferencias entre ambos, el estimador aquí hallado y el predicho por Amemiya, arroja un valor de 0.01, lo cual

es en realidad bastante bajo. Recuerde que esto es posible debido a la similitud entre las funciones de densidad.

5. De acuerdo con los resultados obtenidos en el punto anterior, responda las siguientes preguntas:
- La interpretación de los resultados de este modelo en cuanto a las variables que más inciden positiva y negativamente en la probabilidad de obtener un cargo gerencial coinciden con los obtenidos en el literal b del punto 2.
 - Los resultados obtenidos por la universidad vecina descritos en el punto 3 se cumplen para este modelo. Justifique.

Respuesta sugerida:

- a) Los resultados se muestran en la Tabla 1.

Se puede observar que los valores obtenidos para ambas estimaciones son muy similares en su valor y conservan el signo. De manera similar, se conserva nuestra conclusión anterior de que, la variable que al incrementarse en una unidad genera el mayor cambio en la probabilidad de obtener un puesto gerencial, es el estar casado (con un incremento de 13.28 puntos porcentuales) y la que mayor decrecimiento genera, es la de pertenecer al género femenino (con un valor de 18.64 puntos porcentuales).

- b) Realizando el mismo análisis del punto 3 pero para el efecto marginal del modelo logit, tenemos las siguientes pruebas de hipótesis:

$$H_0: \beta_4 = \beta_6$$

Ha: No H_0 .

El resultado de esta prueba con un test de Wald, muestra que el estadístico es 0.60 con un p-valor de 0.43691, concluyendo así que no existe suficiente evidencia para afirmar que los efectos marginales promedio de las variables carreras y TOEFL son estadísticamente diferentes.

Para la edad y educación tenemos:

$$H_0: \beta_2 = \beta_5$$

Ha: No H_0 .

El resultado de esta prueba con un test de Wald, muestra que el estadístico es 0.58 con un p-valor de 0.44448, concluyendo así que no existe suficiente evidencia para afirmar que los efectos marginales promedio de las variables edad y educación son estadísticamente diferentes.

6. Argumente por qué teóricamente es correcto su análisis anterior con los modelos empleados. ¿Qué aportan estos modelos respecto a modelo de probabilidad lineal y por qué sería correcto o no estimar la ecuación empleada en la investigación mediante este último modelo?

Es correcto analizar el modelo empleado con logit y probit, pues ambas consideran que la probabilidad de que un suceso ocurra no se una función lineal de variables

explicativas, es decir que no asume una restricción tan fuerte como decir que los incrementos en la probabilidad debido a cambios en los regresores es constante.

Al asumir una función de distribución normal y logística, la variación ya no es constante, lo cual es ciertamente más realista que asumir un cambio constante. Un claro ejemplo son los años de educación y la probabilidad de obtener empleo, pues no es lo mismo un año más de educación, si se pasa de noveno a décimo, que si se pasa de undécimo a la universidad, claramente, el obtener el título de bachiller debe incrementar la probabilidad de obtener trabajo en una mayor proporción que pasar de noveno a décimo.

De esta forma, podemos decir que los modelos logit y probit aportan en comparación con MPL un análisis más realista de las probabilidades, pues no sería correcto estimar la ecuación por MPL, pues las probabilidades deberían cambiar con las variaciones de los regresores.

Tabla 1. Estimación modelos Probit y Logit.

VARIABLE DEPENDIENTE: Gerencial,
Estadísticos t entre paréntesis

	Probit EMV		Logit EMV	
	Coeficientes	Efecto Marginal	Coeficientes	Efecto Marginal
<i>constante</i>	-1.7941 (-13,07) ***	--	-3.0486 (-12,55) ***	--
<i>Edad_i</i>	0.0298 (23.00) ***	0.0079	0.0517 (22.40) ***	0.0079
<i>Experiencia_i</i>	0.0550 (1.92) *	0.0147	0.0695 (1.38)	0.0106
<i>Carreras_i</i>	0.2658 (10.77) ***	0.0711	0.4613 (10.58) ***	0.0706
<i>Educacion_i</i>	0.0384 (5.70) ***	0.0103	0.0607 (5.10) ***	0.0093
<i>TOEFL_i</i>	0.3337 (3.10) ***	0.0893	0.6168 (3.17) ***	0.0945
<i>Inversión_i</i>	0.0000 (12.00) ***	0.0000	0.0000 (11.44) ***	0.0000
<i>S.público_i</i>	0.0616 (3.23) ***	0.0165	0.1087 (3.29) ***	0.0166
<i>Casado_i</i>	0.5272 (11.94) ***	0.1411	0.8673 (11.31) ***	0.1328
<i>Género_i</i>	-0.6675 (-15,10) ***	-0.1787	-1.1865 (-14,99) ***	-0.1817
<i>Internacional_i</i>	0.1177 (2.64) ***	0.0314	0.1667 (2.16) **	0.0255
<i>Honor_i</i>	-0.1643 (-8,57) ***	-0.0439	-0.2748 (-8,30) ***	-0.0421
LRI	0.2621		0.2677	
Wald	1,656.14 ***		1,394.42 ***	
ln(L)	-3,154.56		-3,130.62	
# de Obs.		6702		6702

(*) nivel de significancia: 10%

(**) nivel de significancia: 5%

(***) nivel de significancia: 1%

LRI: Razón de Máxima verosimilitud

Wald: corresponde al estadístico de Wald que comprueba la significancia conjunta de todas las pendientes.

LN(L): Logaritmo de la función de máxima verosimilitud