

Examen Final - Soluciones

1. Interpretación de coeficientes

A continuación tendrán que interpretar de la forma más específica posible los valores estimados de los coeficientes que se les indican. Todas estas preguntas hacen referencia a la Tabla 2 del documento que recibieron via email, excepto en el caso de Logit Multinomial.

1. (5 puntos) Coeficiente de “Respuesta” en la columna (1)

Respuesta: *Caeteris paribus*, las personas que sí respondieron al email más reciente tienen una probabilidad 19.8 puntos porcentuales mayor de haber donado este año que aquellas que no respondieron.

2. (5 puntos) Coeficiente de “Semanas resp” en la columna (2)

Respuesta: *Caeteris paribus*, un aumento en semanas desde la última respuesta del individuo a emails enviados está relacionado a una disminución en la probabilidad de haber donado este año.

3. (5 puntos) Coeficiente de “Mails” en la columna (3)

Respuesta: *Caeteris paribus*, un aumento de un mail enviado el año pasado incrementa en 21.37 % el ratio de la probabilidad de haber donado este año respecto a no haber donado.

4. (5 puntos) Coeficiente de “Donacion prom” en la columna (4)

Respuesta: *Caeteris paribus*, un aumento de un euro en las donaciones promedio hechas previamente está asociado a un aumento de la probabilidad de estar en el cuartil más alto del monto donado el último año y a una reducción en la probabilidad de estar en el cuartil más bajo.

5. (5 puntos) Interpreta de la forma mas especifica posible el valor estimado del coeficiente “Age” de la probabilidad de elegir la estufa *gr* si tu estufa de referencia es *hp*

Respuesta: *Caeteris paribus*, un aumento de un año de edad del jefe del hogar está relacionado con un aumento de 1.9 % en el ratio de la probabilidad de elegir la estufa *gr* respecto a elegir la estufa *hp*.

2. Modelo de Probabilidad Lineal

Las siguientes preguntas se relacionan con la Tabla 1 en el documento que recibieron via email.

6. (10 puntos) Indica cuál sería el valor estimado del coeficiente de “respuesta” en una estimación de un modelo de probabilidad lineal de la siguiente especificación:

$$Dono_i = \beta_0 + \beta_1 Respuesta_i + U_i$$

Respuesta: Empecemos por definir los coeficientes:

$$Pr(Dono_i = 1 | Respuesta = 0) = \beta_0$$

$$Pr(Dono_i = 1 | Respuesta = 1) = \beta_0 + \beta_1$$

Con la información disponible en la Tabla 1 sabemos que de las personas que donaron, el 58 %¹ de los individuos sí respondieron al email más reciente (i.e. $Respuesta = 1$). Dado que hay 1,707 individuos que sí donaron, sabemos que 990 de ellos si respondieron al email más reciente y 717 no lo hicieron. Similarmente, si nos fijamos en la columna de los que no donaron sabemos que el 21 % de los individuos sí respondieron al email más reciente. Habiendo 2,561 individuos que no donaron, sabemos que 531 de ellos si respondieron el email más reciente y 2,030 no lo hicieron.

Con estos datos, ya sabemos el total de individuos que sí respondieron al email más reciente ($990 + 531 = 1521$) y el total que no lo hicieron: ($717 + 2030 = 2746$). Y podemos entonces calcular las probabilidades que establecimos arriba:

$$\begin{aligned} Pr(Dono_i = 1 | Respuesta = 0) &= \hat{\beta}_0 = \frac{717}{2746} = 0.26 \\ Pr(Dono_i = 1 | Respuesta = 1) &= \hat{\beta}_0 + \hat{\beta}_1 = \frac{990}{1521} = 0.65 \\ \implies \hat{\beta}_1 &= 0.65 - 0.26 = 0.39 \end{aligned}$$

7. (3 puntos) Con solo ver la Tabla 1, ¿podrías determinar cuáles serían los valores estimados de los coeficientes de un modelo de probabilidad lineal que estimaría la siguiente especificación?

$$Dono_i = \beta_0 + \beta_1 Respuesta_i + \beta_2 Mails_i + U_i$$

¹Esta proporción variaba entre distintas versiones de examen.

Respuesta: Falso. Para poder determinar los coeficientes tendríamos que ver combinaciones de condiciones que la Tabla 1 no nos indica. Por ejemplo, tan solo la estimación de β_0 implicaría:

$$Pr(Dono_i = 1 | Respuesta = 0, Mails = 0) = \beta_0$$

Similarmente, para determinar β_1 tendríamos que poder observar la diferencia de probabilidad de donar entre individuos que responden y no responden al último email, pero esto para diferentes combinaciones de la variable *Mails*, pensando en el caeteris paribus. Por último, β_2 solo la podemos observar para pares de individuos que responden al último mail, pero uno de ellos recibió un email más el año pasado. De hecho, aun no controlando por *Respuesta*, no podríamos solo con la Tabla 1 estimar el coeficiente β_2 .

3. MPL, Probit y Logit

Las siguientes preguntas se refieren a la Tabla 2 del documento que recibiste por email.

8. (6 puntos) Indica TODAS las afirmaciones que sean correctas para el modelo de probabilidad lineal utilizando los resultados de la columna 1 de la Tabla 2.

Respuesta:

- Es posible interpretar el signo del coeficiente (excepto la constante): los coeficientes de signo positivo indican que un aumento en dicha variable esta relacionada a un aumento en la probabilidad de que un individuo haya hecho una Donacion ($Dono = 1$)

Cierto. De hecho no solo podemos interpretar el signo. Podemos además interpretar el valor específico, pero es cierto que el signo por si solo sí indica si la variable aumenta (signo positivo) o disminuye (signo negativo).

- El coeficiente de “semanas resp” indica el “efecto parcial promedio” sobre la probabilidad de donar que resulta de un aumento de una semana desde la ultima respuesta del individuo a emails recibidos.

Cierto. En el caso de MPL, el coeficiente de “semanas resp” corresponde a la derivada parcial de $Pr(Donar = 1 | X_i)$ respecto a “semanas resp”. Y en el caso de MPL, esta derivada entrega directamente el coeficiente. Por lo tanto, promediarlo sobre todos los individuos nos da el propio coeficiente.

- El coeficiente de “donacion prom” indica el “efecto parcial para la persona promedio” de la probabilidad de donar que resulta de aumentar en un Euro las donaciones promedio previas

Cierto. Esto es una extensión directa del argumento anterior. Como la derivada parcial no depende de X , el coeficiente representa el cambio en la probabilidad

de donar por incrementar las donaciones previas para cualquier individuo, y en particular para el individuo promedio.

- El coeficiente de “mails” indicaría el aumento en la probabilidad de donar por incrementar un mail enviado por año para CUALQUIER persona en la base de datos

Cierto. Nuevamente, es extensión de los argumentos previos.

- Podemos asegurar que el modelo no predecirá para ninguna persona en la base de datos una probabilidad de donar mayor a 1

Falso. Como vimos en clase, estamos aplicando un OLS y siendo que tenemos una línea con pendiente constante, podría darse el caso que tengamos alguna predicción que salga del intervalo $[0,1]$.

9. (10 puntos) Utilizando los resultados de la columna 2 de la Tabla 2, calcula el ratio $\frac{Pr(Donar_i=1|X_i)}{Pr(Donar_i=0|X_i)}$ que predeciría tu modelo para una persona que: (i) no respondió al email mas reciente, (ii) recibe en promedio 2 emails al año, (iii) en el pasado ha donado en promedio 20 Euros, y (iv) hace 50 semanas se recibió su última respuesta a un email.

NOTA: ¡no es un error que me refiera a la columna 2 del modelo probit! Puedes hacer tus cálculos en Excel o en una hoja usando tu calculadora. Carga el archivo donde este tu respuesta. Asegúrate que este indicado muy claramente.

Respuesta: Con el modelo de Probit y las características mencionadas podemos calcular:

$$\begin{aligned}
 Pr(Donar = 1|X_i) &= \Phi(-0.2228 + 0.4656 * (0) + 0.1311 * (2) + 0.0007 * (20) \\
 &\quad - 0.0087 * (50)) \\
 &= \Phi(-0.3816) = 0.3514 \\
 Pr(Donar = 0|X_i) &= 1 - Pr(Donar = 1|X_i) = 0.6486 \\
 \frac{Pr(Donar = 1|X_i)}{Pr(Donar = 0|X_i)} &= \frac{0.3514}{0.6486} = 0.5417
 \end{aligned}$$

10. (5 puntos) Repite la pregunta anterior con los resultados del modelo logit (columna 3).

Respuesta: Con el modelo logit simplemente tenemos que recordar que:

$$\begin{aligned}\ln \left(\frac{Pr(Donar = 1|X_i)}{Pr(Donar = 0|X_i)} \right) &= -0.3394 + 0.7424 * (0) + 0.2137 * (2) \\ &\quad + 0.0012 * (20) - 0.0146 * (50) \\ &= -0.618 \\ \exp \left(\ln \left(\frac{Pr(Donar = 1|X_i)}{Pr(Donar = 0|X_i)} \right) \right) &= \frac{Pr(Donar = 1|X_i)}{Pr(Donar = 0|X_i)} = \exp(-0.618) = 0.539\end{aligned}$$

11. (8 puntos) En los modelos (1) a (3): ¿consideras que sería adecuado agregar como control la variable “donacion anterior”? ¿Qué efecto crees que esto tendría sobre el valor estimado del coeficiente de “donacion prom”? Justifica claramente tus respuestas.

Respuesta: La clave aquí es la multicolinealidad. Podríamos imaginar que la variable “donacion anterior” esta altamente correlacionado con otro de los controles: “donacion prom”. Además ambas variables conceptualmente capturan lo mismo, el perfil de qué tan generoso ha sido el individuo en sus donaciones previas. Por lo tanto, si agregáramos esta variable seguramente haría que el valor estimado del coeficiente de “Donacion prom” se redujera, pero esto no necesariamente querría decir que estaría preveniendo un problema de sesgo por variables omitidas, sino que el efecto de la generosidad del individuo acabaría distribuida entre ambos controles. Por lo tanto, en un caso como este no sugeriría agregar esta variable.

4. Modelos de Probit y Logit ordenado

12. (5 puntos) ¿Por qué crees que solo obtenemos dos α 's (constantes) en estos modelos siendo que tenemos 4 cuartiles?

Respuesta: Si vemos la Tabla 1 podemos ver que más de la mitad de la muestra no lleva a cabo ninguna donación. Si nos fijamos en el número de observaciones de las columnas 4 y 5, podemos ver que estamos utilizando toda la muestra. Esto quiere decir que el primer y segundo cuartil no pueden ser distinguidos dado que en ambos casos tenemos solo individuos que no llevan a cabo donaciones. Es por ello, que el modelo únicamente distingue entre 3 distintos tipos de individuos: (i) aquellos en el primer o segundo cuartil, (ii) aquellos en el tercer cuartil y (iii) aquellos en el cuarto cuartil. Esto quiere decir, que solo necesitamos dos constantes para dividir el plano de la $X_i'\beta$ en tres intervalos.

13. (15 puntos) (**Opcion A**) Probit ordenado:

Considera a una persona con las siguientes características (mismo individuo de la pregunta Probit 2): (i) no respondió al email mas reciente, (ii) recibe en promedio 2 emails al año, (iii) en el pasado ha donado en promedio 20 Euros, y (iv) hace 50 semanas se recibió su ultima respuesta a un email.

Utilizando los resultados de la columna (4) calcula: ¿cuál es la probabilidad de que este individuo SI haga una donación? Contrasta tu resultado con el resultado que obtendrías usando la información de la columna (2)

Respuesta: Dado que el primer y segundo cuartil (que corresponden a la categoria mas baja) indican el no hacer donaciones, lo único que tenemos que hacer es calcular la probabilidad de estar en dicho cuartil y la probabilidad de realizar donaciones será el complemento. Por lo tanto:

$$\begin{aligned}
 Pr(Cuartil > 1|X_i) &= 1 - Pr(Cuartil = 1|X_i) \\
 &= \Phi(X_i'\beta - \alpha_1) \\
 &= \Phi(0.4042 * (0) + 0.1736 * (2) + 0.0006 * (20) \\
 &\quad - 0.0079 * (50) - 0.3362) \\
 &= \Phi(-0.372) = 0.3549
 \end{aligned}$$

Recordando nuestra respuesta de la pregunta 9, ahí obteníamos que esta probabilidad era igual a 0.3514. Por lo tanto, vemos que la diferencia es de solamente 0.35 puntos porcentuales.

14. (15 puntos) **(Opcion B)** Logit Ordenado:

Utilizando el resultado de la columna (5), calcula el “efecto parcial para la persona promedio” sobre la probabilidad de estar en el cuartil mas alto que resultaría de aumentar un email enviado por año.

Respuesta: Para poder resolver esto podemos utilizar la estimación exacta y calcular la diferencia entre dos predicciones: (a) la probabilidad del cuartil mas alto para la persona promedio y (b) la probabilidad del cuartil mas alto para la persona promedio incrementándole un email enviado al año. Para determinar a la persona promedio utilizamos las características de la Tabla 1 y empleando el número de observaciones calculamos un promedio ponderado para poder definir a la persona promedio. Si hacemos eso tendremos que las características de la persona promedio son: (i) una proporción de 0.36 responde a sus emails², (ii) recibe en promedio 2.05 emails al año, (iii) en el pasado ha donado en promedio 18.24 Euros, y (iv) hace 59.05 semanas se recibió su última respuesta a un email.

²esto puede variar para distintos tipos de examen

Empecemos con (a):

$$\begin{aligned}
 Pr(Cuartil = 4|\bar{X}) &= \Phi(\bar{X}'\beta - \alpha_2) \\
 &= \Phi(0.6603 * (0.36) + 0.2785 * (2.05) + 0.0009 * (18.24) \\
 &\quad - 0.0136 * (59.05) - 1.7027) \\
 &= \Phi(-1.683) = \frac{\exp(-1.683)}{1 + \exp(-1.683)} = 0.1567
 \end{aligned}$$

Para calcular (b) solo basta sumar el coeficiente de *mails* a la $\bar{X}'\beta$ y obtenemos:

$$\begin{aligned}
 Pr(Cuartil = 4|\bar{X}') &= \Phi(\bar{X}'\beta + \beta_{Mails} - \alpha_2) \\
 &= \Phi(-1.683 + 0.2785) = \frac{\exp(-1.405)}{1 + \exp(-1.405)} = 0.1971
 \end{aligned}$$

Y entonces tenemos que el efecto parcial para la persona promedio de incrementar un email enviado por año resulta en un aumento de 4.03 puntos porcentuales en la probabilidad de que la donación del último año este en el cuartil más alto.

También podríamos haber utilizado la estimación aproximada siguiendo las notas que indican que esta es:

$$\begin{aligned}
 \frac{\partial Pr(Cuartil = 4|\bar{X})}{\partial Mails} &= \Phi(\bar{X}'\beta - \alpha_2) * (1 - \Phi(\bar{X}'\beta - \alpha_2)) * \beta_{Mails} \\
 &= \Phi(-1.683) * (1 - \Phi(-1.683)) * 0.2785 \\
 &= (0.1567) * (1 - 0.1567) * 0.2785 = 0.0368
 \end{aligned}$$

Si siguiendo esta alternativa, tendríamos que el efecto parcial para la persona promedio de incrementar un email enviado por año resulta en un aumento de 3.68 puntos porcentuales en la probabilidad de que la donación del último año este en el cuartil más alto.

5. Multinomial Logit

En el modelo logit multinomial de la Tarea 3 modelamos la eleccion de la siguiente forma:

$$U_{si} = \beta_{ic}IC_{si} + \beta_{oc}OC_{si} + X_i'\gamma_s + \epsilon_{si}$$

donde -s- corresponde a la alternativa de estufa, que puede ser {gc,gr,ec,er,hp}, -i- indica el hogar. Las variables IC y OC son especificas a la estufa y el hogar, mientras que las variables "X" incluyen edad, ingreso, región y número de habitantes del hogar y son especificas del hogar -i-.

15. (10 puntos) Imagina que observas a un hogar con las siguientes características: (i) edad del jefe del hogar 35, (ii) nivel de ingreso 5, (iii) 5 habitantes del hogar y (iv) que vive en la region "valley". Utilizando los resultados de tu estimación, indica qué alternativa de estufa predeciría tu modelo que este hogar elegiría si es que el costo de instalación fuera de 500 USD y el costo anual de operación fuera de 200 USD para todas las alternativas de estufa.

Respuesta: Al final del examen agregué la Tabla 1 que muestra los resultados de las estimaciones del logit multinomial. Para entender qué estufa predeciría el modelo que elegiría el hogar con las características indicadas tendríamos que comparar la utilidad que predice el modelo para cada modelo. La utilidad para el hogar -i- de elegir la estufa -s- se indica en la función de arriba. En el caso particular que estamos analizando, tenemos que, dado que estamos comparando una situación donde los costos de operación (OC) y de instalación (IC) son iguales, para determinar la alternativa que da mayor utilidad únicamente tendríamos que comparar el $X_i' \gamma_s$. Usando los resultados de la Tabla 1:

$$U_{i,gc} = +1.1669 + 0.0151 * (35) - 0.0092 * (5) - 0.6743 * (1) + 0.6162 * (1) = 1.59$$

$$U_{i,gr} = -0.9138 + 0.0184 * (35) - 0.0174 * (5) - 0.0749 * (1) + 0.3996 * (1) = -0.03$$

$$U_{i,ec} = +0.4939 + 0.0196 * (35) + 0.0461 * (5) - 0.4154 * (1) + 0.6257 * (1) = 1.62$$

$$U_{i,er} = +1.4788 - 0.0062 * (35) + 0.0177 * (5) - 0.0803 * (1) + 0.7804 * (1) = 2.05$$

Por lo tanto, este hogar elegiría comprar una estufa **er**.

Cuadro 1: Logit Multinomial

VARIABLES	(1) gc	(2) gr	(3) ec	(4) er
Coeficientes comunes:				
ic	-0.00155** (0.0006)	-0.00155** (0.0006)	-0.00155** (0.0006)	-0.00155** (0.0006)
oc	-0.00695*** (0.0016)	-0.00695*** (0.0016)	-0.00695*** (0.0016)	-0.00695*** (0.0016)
Coeficientes específicos:				
age	0.01511 (0.0108)	0.01838 (0.0122)	0.01962 (0.0137)	-0.00617 (0.0131)
num_people	-0.00919 (0.0861)	-0.01742 (0.0972)	0.04609 (0.1096)	0.01766 (0.1039)
Constante	1.16694 (0.7284)	-0.91380 (0.8226)	0.49389 (1.0133)	1.47884 (0.9280)
Dummies region:				
valley	-0.67426 (0.4397)	-0.07488 (0.5036)	-0.41542 (0.5598)	-0.08027 (0.5393)
scostl	-0.42494 (0.3918)	0.21972 (0.4457)	-0.20216 (0.4903)	0.15361 (0.4765)
mountn	-0.68862 (0.5105)	-0.02816 (0.5792)	-0.34154 (0.6492)	-0.09413 (0.6217)
Dummies income:				
Income1	0.53802 (0.5173)	0.61830 (0.5886)	0.23579 (0.6884)	0.58042 (0.6444)
Income2	-0.04258 (0.4630)	0.75348 (0.5203)	0.31073 (0.6001)	0.45885 (0.5835)
Income3	1.41516** (0.6575)	1.49740** (0.7110)	1.32941* (0.7736)	1.47442* (0.7532)
Income4	0.10648 (0.4268)	-0.04492 (0.5134)	0.00198 (0.5826)	0.27640 (0.5479)
Income5	0.61621 (0.4928)	0.39963 (0.5744)	0.62569 (0.6269)	0.78039 (0.6035)
Observations	4,500	4,500	4,500	4,500

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1