



Universidad de San Andrés

ECONOMETRÍA AVANZADA

WALTER SOSA ESCUDERO
GASTÓN GARCÍA ZAVALETA

Trabajo Práctico 4

GARCÍA VASSALLO, HEDEMAN, IOLSTER, SURY

2022

Ejercicio 1

En el año 2006, la Argentina asume el compromiso de que todos los adolescentes y jóvenes del país completen el nivel de educación secundaria como parte de su educación básica y obligatoria (Ley N° 26.206). Si bien el acceso a dicho nivel se ha incrementado significativamente en los últimos veinte años, las estadísticas educativas indican que muchos adolescentes y jóvenes tienen aún serias dificultades para permanecer y completar el nivel secundario. La obligatoriedad de la secundaria constituye así un desafío ineludible e importante para el sistema educativo argentino.

La deserción escolar genera elevados costos sociales y privados. Cuando hablamos de costos sociales podemos mencionar los que derivan de disponer de una fuerza de trabajo menos competente y más difícil de calificar. La baja productividad del trabajo, y su efecto en el (menor) crecimiento de las economías, se considera también como un costo social del bajo nivel educacional que produce el abandono de la escuela durante los primeros años del ciclo escolar. Asimismo, representan un costo social los mayores gastos en los que es necesario incurrir para financiar programas sociales y de transferencias a los sectores que no logran generar recursos propios. Por otro lado, teniendo en cuenta los costos privados de la deserción escolar, estos pueden calcularse sobre la base de una estimación del menor ingreso futuro que obtienen las personas en el mercado de trabajo como consecuencia de completar un número menor de tiempo de estudios, en comparación con un nivel de escolaridad preestablecido. En concreto, los costos privados se refieren a los ingresos laborales que dejan de percibir durante su vida activa los jóvenes que abandonan con anticipación sus estudios.

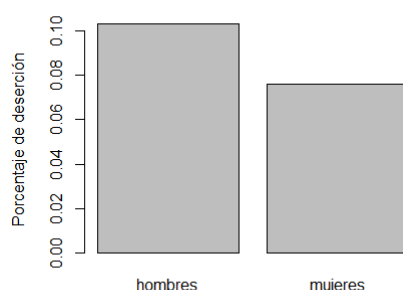
En este sentido, es importante que el gobierno lleve a cabo políticas de incentivos hacia los jóvenes para que estos terminen el secundario para así mejorar la situación de deserción escolar.

A continuación presentamos una tabla con estadísticas descriptivas sobre las variables de interés:

deserta	mujer	jmujer	educ_jefe	edad	ingreso_per_capita
0:4112	0:2340	0:2746	4 :2154	Min. :13.00	Min. : 0
1: 406	1:2178	1:1772	2 :1189	1st Qu.:14.00	1st Qu.: 3302
			7 : 586	Median :16.00	Median : 6833
			6 : 503	Mean :15.85	Mean : 9088
			8 : 50	3rd Qu.:17.00	3rd Qu.: 11857
			0 : 13	Max. :19.00	Max. :175000
			(Other): 23		

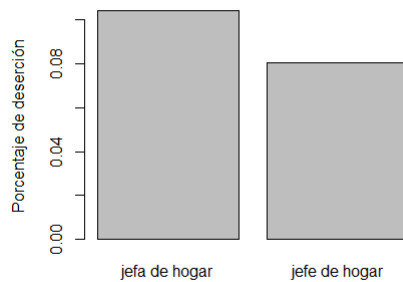
Lo que haremos será una descripción de las variables que consideramos relevantes.

- *mujer*: En la literatura hay evidencia de que, ceteris paribus, es más probable de que los hombres en edad de asistir al secundario deserten dado que el costo de oportunidad en términos laborales es mayor que en el caso de las mujeres. En nuestra base de datos hay 2340 hombres y 2178 mujeres.

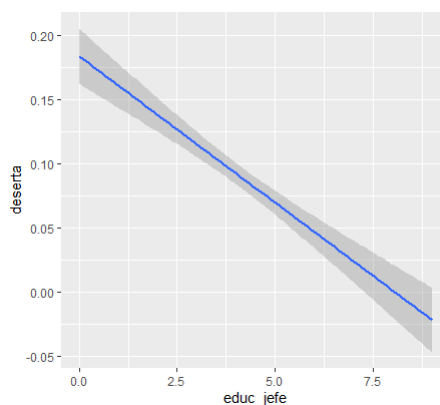


La correlación entre la dummy que toma como valor 1 si es mujer y la deserción escolar toma como valor -0.04758354

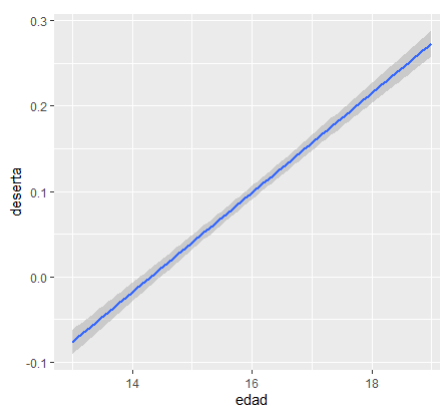
- *jmujer*: Esta variable toma valor 1 si se declara que la persona al mando del hogar es una mujer. Consideramos que esta variable es relevante ya que pensamos que es un buen indicador de monoparentalidad, es decir, la ausencia del padre en el hogar. De este modo, consideramos que es un indicador de un contexto vulnerable en el que probablemente sea muy costoso invertir en la educación de los hijos y por lo tanto el costo de oportunidad de asistir al secundario es mayor. En nuestra base de datos hay 2746 hogares donde los jefes son hombres y 1772 hogares donde las jefas son mujeres.



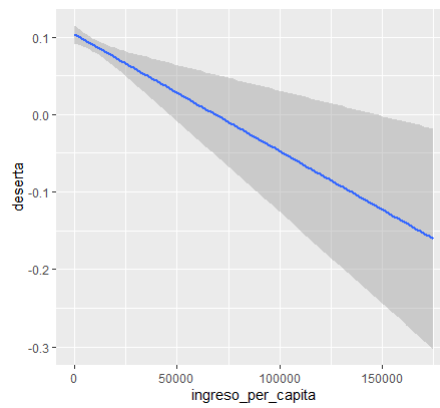
- *educ_jefe*: En la literatura hay evidencia de que hay una alta correlación entre la educación de los padres y la educación de los hijos. Naturalmente, si el jefe del hogar termino el secundario es mucho mas probable, ceteris paribus, que se priorice más la educación de los hijos en comparación a los casos en los que el jefe del hogar no tiene el secundario completo. En otras palabras, en aquellos hogares en los que hay una mayor cultura de estudio, terminar el secundario está mucho más valorado. De este modo, esperamos encontrar que los hogares en los que los hijos no desertan hay un mayor nivel de los jefes del hogar.



- *edad*: Consideramos que a mayor edad mayor será la probabilidad de no terminar los estudios secundarios. Esto es así ya que a mayor edad, aumenta el costo de oportunidad de estudiar dado que se presentan mayores oportunidades laborales. En particular, creemos que hay un salto importante al pasar los 18 años (mayoría de edad). En nuestra base de datos, la edad mínima es 13 mientras que la máxima es 19.



- *ingreso_per_cpita*: Naturalmente, en los hogares de menores ingresos es más probable el abandono escolar. Una de las razones principales corresponde a que en muchos casos, dado que los ingresos de los padres no alcanzan, es necesario que los hijos en edad de asistir al secundario salgan a trabajar. En nuestra base de datos el ingreso per capita mínimo es 0 mientras que el máximo es 175000 y el promedio es 9088.



Ejercicio 2

A continuación estimamos un modelo probit para la probabilidad de que un individuo abandone el secundario agregando como variables explicativas las analizadas en el inciso anterior. El modelo a estimar es el siguiente:

$$Deserta_i = \beta_0 + mujer_i \beta_1 + jmujer_i \beta_2 + educ_jefe_i \beta_3 + ingreso_per_capita_i \beta_4 + edad_i \beta_5 + \mu_i$$

En el cuadro 1 que se encuentra en el anexo podemos encontrar la reproducción del modelo probit. A partir de estos resultados, solo podemos interpretar los signos de los coeficientes y no sus magnitudes dado que los efectos marginales no son constantes a través de los valores de las variables explicativas.

En primer lugar puede observarse que la probabilidad de deserción escolar disminuye con el nivel educativo del jefe del hogar con respecto a tener sólo educación preescolar (excepto para el caso de EGB). En cuanto a la significativa, la única categoría significativa (al 5 %) es tener estudios universitarios. En segundo lugar, si la jefa del hogar es mujer, la probabilidad de desertar aumenta en comparación a aquellos hogares cuyos jefes son hombres y esta diferencia es significativa al 5 %. En tercer lugar, cuanto mayor la edad del individuo, mayor la probabilidad de abandonar el secundario, siendo esta variable significativa al 1 %. En cuarto lugar, al ser mujer disminuye la probabilidad de deserción en comparación al caso de ser hombre, siendo esto significativo al 10 %. Por último, el signo correspondiente a ingreso per cápita es negativo pero muy cercano a 0 y no resulta significativa. Consideramos que esto puede deberse a que la EPH permite que los hogares se rehusen a contestar cuál es su ingreso y los incluye en la muestra asignándoles un 0.

En conclusión, los signos de las variables independientes resultan ser los esperados.

Ejercicio 3

En el cuadro 2 computamos los efectos marginales en las medias para las variables de interés.

Tal como vimos en clase, al indicarle a R el uso de variables categóricas, al calcular el efecto marginal en el promedio, para este tipo de variables el efecto marginal es calculado en una de las categorías.

En primer lugar podemos observar que cuando la persona es mujer, en promedio la probabilidad de deserción escolar disminuye 0.9 % con respecto a cuando el individuo es hombre, ceteris paribus. Por otro lado, si el jefe de hogar es una mujer la probabilidad de deserción escolar en promedio aumenta un 1 % en comparación al caso en el que el jefe del hogar es hombre, ceteris paribus. Luego, el efecto del nivel de educación del jefe de hogar sobre la probabilidad de deserción escolar es mayor a medida que el nivel de educación del jefe de hogar aumenta. De este modo, la probabilidad de abandono disminuye en un 1,1 % si el jefe del hogar tiene el primario completo con respecto al caso de tener solo educación preescolar mientras que la probabilidad de abandono disminuye en un 4,1 % si el jefe del hogar tiene estudios universitarios con respecto al caso de tener solo educación preescolar. Con respecto al resto de las categorías, por ejemplo, la probabilidad de abandono disminuye en un 3,4 % si el jefe del hogar tiene el secundario completo con respecto al caso de tener solo educación preescolar, la probabilidad de abandono disminuye en un 1,8 % si el jefe del hogar alcanzó el polimodal con respecto al caso de tener solo educación preescolar, la probabilidad de abandono disminuye en un 2,9 % si el jefe del hogar tiene educación terciaria con respecto al caso de tener solo educación preescolar. Además, observando la variable edad, podemos concluir que para una persona con edad promedio si su edad aumenta en un 1 año, la probabilidad de desertar aumenta en promedio 3.4 %. Por ultimo, podemos observar que el ingreso per cápita promedio no afecta la deserción escolar.

Cabe destacar que dado que los efectos marginales no son constantes, al realizar un análisis de los efectos margi-

nales en la media puede ser el caso en que la media no sea un punto relevante en el cual analizar los efectos marginales. Es decir, que haya otros valores de las variables en los que analizar el efecto marginal resulta mas informativo. Por ejemplo, puede pensarse que en el caso de la edad, es relevante ver que pasa cuando se calcula el efecto marginal de pasar de 17 años a 18, es decir, alcanzar la mayoría de edad. Además, el promedio es muy sensible a los outliers, por lo tanto, proponemos calcular los efectos marginales en la mediana.

Ejercicio 4

A continuación estimamos el siguiente modelo de probabilidad lineal:

$$Deserta_i = \beta_0 + mujer_i \beta_1 + jmujer_i \beta_2 + educ_jefe_i \beta_3 + ingreso_per_capita_i \beta_4 + edad_i + \beta_5 \mu_i$$

En el cuadro 3 que se encuentra en el anexo podemos encontrar la reproducción del modelo logit. A partir de estos resultados, al igual que con el modelo probit, solo podemos interpretar los signos de los coeficientes y no sus magnitudes. A continuación analizaremos los resultados del modelo logit y los compararemos con los resultados obtenidos en el modelo probit.

Entonces, en primer lugar se puede observar que la probabilidad de deserción escolar en el modelo logit disminuye cuando el individuo es mujer dado que el coeficiente asociado es negativo y significativo al 10 %, esto mismo ocurre con el modelo probit. En segundo lugar, al igual que en el modelo probit, la probabilidad de desertar aumenta si la jefa de hogar es mujer en comparación a los hogares donde el jefe de hogar es hombre. Esta diferencia es significativa al 5 %. En tercer lugar, la probabilidad de deserción escolar, al igual que en el modelo probit, disminuye con el nivel educativo del jefe del hogar con respecto a tener solo educación preescolar (excepto para el caso EGB). En cuanto a la significancia, la única categoría significativa es cuando el jefe de hogar tiene estudios universitarios completos y es significativo al 5 %. En cuarto lugar, en cuanto a la variable edad, al igual que en el modelo probit, en el modelo logit el coeficiente asociado a esta es positivo por lo que a mayor edad mayor probabilidad de deserción escolar y su significancia es del 1 %. Por ultimo, en cuanto al ingreso per capita, el signo correspondiente es negativo al igual que en el modelo probit. Sin embargo, esta variable no es significativa.

Los modelos Logit y Probit son modelos econométricos no lineales que se utilizan cuando la variable dependiente es binaria o dummy, es decir que sólo puede tomar dos valores. Se debe tener un especial cuidado al comparar los modelos Logit y Probit si en especial se tratan de diferencias, debido a que ya existe una diferencia principal. Y es que utilizan funciones de enlace que matemáticamente son diferentes (Logaritmo neperiano de los odds, y la inversa de la distribución normal estándar acumulada respectivamente.), debido a ello no se debe esperar que las estimaciones coincidan. Cabe resaltar que los coeficientes de correlación son ligeramente menores en los modelos Probit que los modelos obtenidos con el modelo Logit.

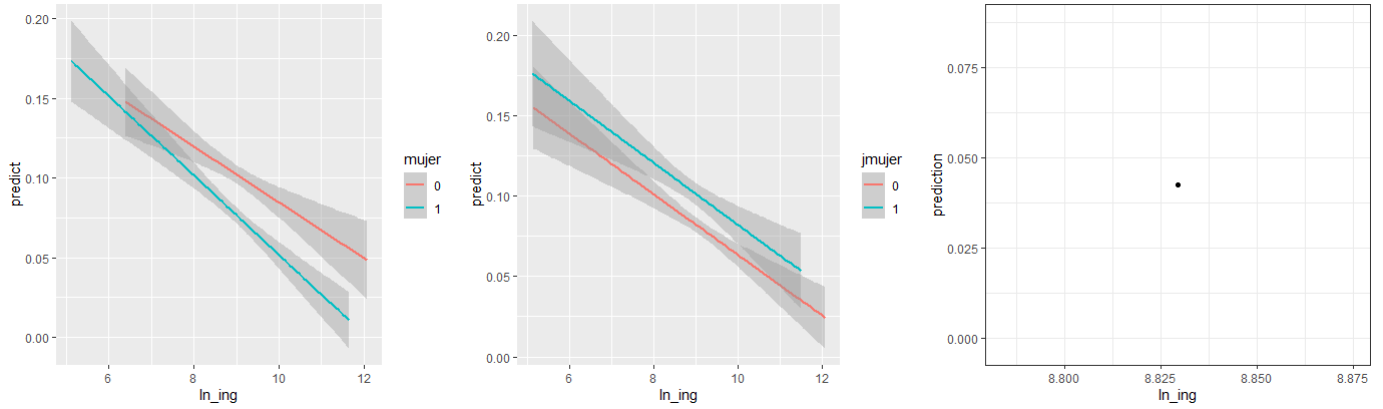
Por otro lado, una de las ventajas del modelo probit es que se pueden obtener estimaciones de probabilidad para la ocurrencia de un suceso. Aparte, si se utiliza un gran número de observaciones, el supuesto de distribución normal de los errores podría primar como criterio para la elección del modelo. Sin embargo, una de las desventajas es que los coeficientes no se pueden interpretar directamente.

La regresión logística es más fácil de implementar, interpretar y muy eficiente de entrenar. No hace suposiciones sobre distribuciones de clases en el espacio de características. Puede extenderse fácilmente a múltiples clases (regresión multinomial) y una vista probabilística natural de las predicciones de clases. La regresión logística es menos propensa a sobreajustar, pero puede sobreajustarse en conjuntos de datos de alta dimensión. Sin embargo, una de las desventajas es que si el número de observaciones es menor que el número de características, no se debe utilizar la regresión logística; de lo contrario, puede provocar un sobreajuste. Además, la principal limitación de la regresión logística es el supuesto de linealidad entre la variable dependiente y las variables independientes.

Ejercicio 5

La resolución de este inciso se encuentra en el script de R.

Ejercicio 6



En el primer gráfico se puede observar como a medida que aumenta el ingreso el gap entre hombres y mujeres se agranda, los hombres siempre teniendo una mayor predicción de desertión ante iguales ingresos.

En el segundo gráfico se notan prácticamente las mismas pendientes a medida que el ingreso aumenta entre hogares con jefe o jefa de hogar. Los hogares con jefa de hogar ante mismos ingresos tienen siempre una predicción más alta de desertión.

Por último, en el tercer gráfico se ve la predicción de la persona con mediana en todas las variables.

Ejercicio 7

Lo que queremos es un subsidio para que la probabilidad que un individuo deserte dado que el jefe del hogar es hombre sea igual a la probabilidad que un individuo deserte dado que la jefa del hogar es mujer ceteris paribus. Dado que tener una jefa de hogar mujer tiene efectos negativos sobre la escolaridad, es decir, tener como jefa de hogar una mujer hace que su hijo/a tenga una mayor probabilidad de desertar, habría que darles a los hogares con jefa de hogar mujer un subsidio para que estos disminuyan su probabilidad de desertar.

$$\hat{\beta}_{\ln_ing} X = P(Y = 1 | jmujer = 1) - P(Y = 1 | jmujer = 0)$$

$$X = \frac{P(Y = 1 | jmujer = 1) - P(Y = 1 | jmujer = 0)}{\hat{\beta}_{\ln_ing}}$$

Con el modelo estimado el X que sería la cantidad de aumento necesaria en ln_ing para que los hogares con jefa de hogar igualen a la probabilidad de desertar a los que tienen un jefe de hogar ceteris paribus es de 18,25.

Anexo

```

# TP4
setwd("~/Desktop/Econometria Tutorial /TPs/TP4")
library(mfx) # para evaluar el efecto marginal en las medias
library(margins) # para evaluar el efecto marginal en un valor
cualquiera
library(ggplot2) # gráficos, dentro de tidyverse
library(stargazer)

install.packages('haven')
library(haven)
mydata <- read_dta("~/Desktop/Econometria Tutorial /TPs/TP4/
cuarto_trim_2019.dta")

names(mydata)

#Nos quedamos con las variables que creemos relevantes
mydata2 <- subset(mydata, select= c( "deserta", "mujer", "jmujer",
"educ_jefe", "edad", "ingreso_per_capita" ) )

mydata2$deserta <- factor(mydata2$deserta)
mydata2$jmujer<- factor(mydata2$jmujer)
mydata2$mujer<- factor(mydata2$mujer)
mydata2$educ_jefe<- factor(mydata2$educ_jefe)

###1
str(mydata2)
sum <- summary(mydata2)
sum
print(xtable(as.table(sum.test), type = "latex"), file = "test.tex")
barplot(c(hombres=0.103,mujeres=0.0758),ylab="Porcentaje de
deserción")

cor(mydata$mujer,mydata$deserta)

jefa = mydata %>% filter(jmujer == 1)
pro_jefa = sum(jefa$deserta)/count(jefa)
no_jefa = mydata %>% filter(jmujer == 0)
pro_nojefa = sum(no_jefa$deserta)/count(no_jefa)

barplot(c("jefa de hogar"=0.104, "jefe de hogar" =
0.0805),ylab="Porcentaje de deserción")

ggplot(mydata, aes(educ_jefe,deserta)) +
geom_smooth(aes(educ_jefe,deserta),method = "lm", formula = y~x)

ggplot(mydata, aes(edad,deserta)) +
geom_smooth(aes(edad,deserta),method = "lm", formula = y~x)

```

```

limp = mydata %>% filter(ln_ing != 0)

ggplot(limp, aes(ingreso_per_capita,deserta)) +
geom_smooth(aes(ingreso_per_capita,deserta),method = "lm", formula =
y~x)

###2
names(mydata2)

myprobit <- glm(deserta ~ mujer + jmujer + educ_jefe + edad +
ingreso_per_capita, family = binomial(link = "probit"),
               data = mydata2)

summary(myprobit)

stargazer(myprobit,
           type="latex",
           dep.var.labels=c("Desercion"),
           out="myprobit")

###3

marginal_media <- probitmfx(deserta ~ mujer + jmujer + educ_jefe +
edad + ingreso_per_capita, data = mydata2 ,
                           atmean = TRUE, robust = TRUE)

marginal_media

stargazer(marginal_media,
           type="latex",
           dep.var.labels=c("Mediamg"),
           out="marginal_media")

###4 si te pide lm esta

mylogit <- glm(deserta ~ mujer + jmujer + educ_jefe + edad +
ingreso_per_capita, family = binomial(link = "logit"),
               data = mydata2)
summary(mylogit)

stargazer(mylogit,
           type="latex",
           dep.var.labels=c("mylogit"),
           out="mylogit")

```


###5 ESTA

```
mydata2$ln_ing<- ifelse(mydata2$ingreso_per_capita>1,  
                        log(mydata2$ingreso_per_capita),0)
```

###6

```
mydata2$predict = predict(myprobit,mydata2, type="response")
```

```
limpio = mydata2 %>% filter(ln_ing != 0)  
ggplot(limpio) +  
  #geom_point( aes(x = ln_ing, y = predict), size = 1.5) +  
  geom_smooth(aes(ln_ing,predict,color = mujer),method = "lm",  
formula = y~x)
```

```
ggplot(limpio) +  
  #geom_point( aes(x = ln_ing, y = predict), size = 1.5) +  
  geom_smooth(aes(ln_ing,predict,color = jmujer),method = "lm",  
formula = y~x)
```

```
myprobit <- glm(deserta ~ mujer + jmujer + educ_jefe + edad +  
ln_ing,  
                family = binomial(link = "probit"),  
                data = mydata2)
```

```
newvalues <- with(mydata2,  
data.frame(educ_jefe=median(as.numeric(educ_jefe)),
```

```
mujer=median(as.numeric(mujer)),  
                                                edad=median(edad),  
jmujer=median(as.numeric(jmujer)),  
                                                ln_ing=median(ln_ing)))  
newvalues
```

```
newvalues$jmujer<- factor(newvalues$jmujer)
```

```
newvalues$mujer<- factor(newvalues$mujer)
```

```
newvalues$educ_jefe<- factor(newvalues$educ_jefe)
```

```
predict(myprobit, newvalues, type="response")
```

```
newvalues[, "prediction"] <- predict(myprobit, newvalues, type =  
"response")
```

```
ggplot(newvalues) +  
  geom_point( aes(x = ln_ing, y = prediction), size = 1.5) +
```

```
#facet_wrap(~ln_ing) +  
theme_bw()  
  
###7  
  
attach(mydata3)  
  
myprobit2 <-glm(deserta ~ ln_ing + mujer + hermanos + jmujer +  
educ_jefe,  
                family = binomial(link = "probit"),  
                data = mydata3)  
stargazer(myprobit2, type="text")  
  
0.146/0.008
```

Cuadro 1: Modelo Probit

	<i>Dependent variable:</i>
	Desercion
mujer	−0.119* (0.063)
jmujer	0.154** (0.064)
educ_jefe2	−0.159 (0.467)
educ_jefe3	0.044 (0.769)
educ_jefe4	−0.455 (0.466)
educ_jefe5	−0.330 (0.806)
educ_jefe6	−0.581 (0.475)
educ_jefe7	−1.036** (0.485)
educ_jefe8	−0.414 (0.581)
educ_jefe9	−4.241 (92.127)
edad	0.447*** (0.022)
ingreso_per_capita	−0.00000 (0.00000)
Constant	−8.421*** (0.616)
Observations	4,518
Log Likelihood	−984.077
Akaike Inf. Crit.	1,994.154
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Cuadro 2: Efectos marginales en la media

	dF/dx	Std. Err.	z	P> z
mujer	-0.009	0.005	-1.882	0.060
jmujer	0.012	0.005	2.227	0.026
educ_jefe2	-0.011	0.024	-0.452	0.651
educ_jefe3	0.003	0.056	0.061	0.952
educ_jefe4	-0.034	0.029	-1.179	0.238
educ_jefe5	-0.018	0.022	-0.825	0.409
educ_jefe6	-0.029	0.013	-2.303	0.021
educ_jefe7	-0.041	0.009	-4.382	0.00001
educ_jefe8	-0.021	0.016	-1.308	0.191
educ_jefe9	-0.033	0.005	-6.713	0
edad	0.033	0.002	14.360	0
ingreso_per_capita	-0.00000	0.00000	-1.214	0.225

Cuadro 3: Modelo logit

	<i>Dependent variable:</i>
	mylogit
mujer	−0.236* (0.121)
jmujer	0.243** (0.121)
educ_jefe2	−0.256 (0.826)
educ_jefe3	0.395 (1.431)
educ_jefe4	−0.785 (0.825)
educ_jefe5	−0.303 (1.502)
educ_jefe6	−1.015 (0.844)
educ_jefe7	−1.994** (0.876)
educ_jefe8	−0.660 (1.056)
educ_jefe9	−11.963 (324.745)
edad	0.984*** (0.049)
ingreso_per_capita	−0.00001 (0.00001)
Constant	−18.319*** (1.224)
Observations	4,518
Log Likelihood	−958.988
Akaike Inf. Crit.	1,943.976
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Cuadro 4: Modelo punto 7

	<i>Dependent variable:</i>
	deserta
ln_ing	−0.008 (0.008)
mujer1	−0.194*** (0.054)
hermanos	0.008 (0.018)
jmujer1	0.146*** (0.055)
educ_jefe2	−0.305 (0.387)
educ_jefe3	−0.602 (0.653)
educ_jefe4	−0.659* (0.387)
educ_jefe5	−0.678 (0.660)
educ_jefe6	−0.786** (0.394)
educ_jefe7	−1.310*** (0.403)
educ_jefe8	−0.786 (0.478)
educ_jefe9	−3.750 (57.937)
Constant	−0.669* (0.397)
Observations	4,518
Log Likelihood	−1,304.808
Akaike Inf. Crit.	2,635.616
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01