

# **Guía de Proyeccion Econométrica**

Francisco Orlando Rosales

2025-05-04

## **Table of contents**

# Preface

El presente libro nace de la necesidad de acercar a los estudiantes y profesionales del análisis económico a una segunda etapa en su formación econométrica: aquella que trasciende el modelo lineal clásico para enfrentar los retos empíricos del mundo real. En la práctica, los datos rara vez se ajustan a los supuestos ideales de linealidad, normalidad o independencia. Por ello, el presente texto se concentra en métodos aplicados y técnicas econométricas que permiten abordar con mayor realismo y rigor los problemas más comunes en la evaluación de políticas, el estudio del comportamiento económico y el diseño de proyecciones con fines prácticos.

A lo largo de mi experiencia docente, he constatado una creciente demanda por parte de los estudiantes de herramientas que les permitan no solo comprender la teoría, sino también aplicarla con propiedad en entornos de datos diversos. Este libro busca responder a esa necesidad con un enfoque aplicado, utilizando ejemplos reales, sintaxis en R, y una estructura clara que vincula la teoría con la práctica. Cada capítulo está diseñado para ser autónomo, pero también parte de una secuencia lógica que va desde la especificación de modelos hasta la identificación de efectos causales.

El texto cubre áreas clave como el modelado de variables dependientes limitadas, el uso combinado de cortes transversales, el análisis con datos de panel, y una introducción metodológica a la evaluación de impacto. Estos temas han sido seleccionados no solo por su relevancia metodológica, sino también por su aplicación directa en proyectos de investigación, tesis de grado y análisis institucional en organismos públicos y privados. La selección responde, además, a una visión integral del análisis econométrico como una herramienta de diagnóstico, predicción y evaluación en contextos complejos y reales.

Espero que este libro contribuya a fortalecer una práctica econométrica más crítica, más contextualizada y más comprometida con los problemas que enfrenta nuestra sociedad. Mi deseo es que los lectores no solo dominen las técnicas aquí presentadas, sino que también aprendan a cuestionar los supuestos detrás de los modelos y a interpretar sus resultados con responsabilidad. Si este texto logra acompañar a sus lectores en ese proceso, su propósito habrá sido cumplido.

# 1 Introducción

La econometría contemporánea ha evolucionado más allá del modelo clásico de regresión lineal, abarcando una diversidad de técnicas diseñadas para abordar preguntas empíricas cada vez más complejas. Este libro de Proyección Econométrica tiene como objetivo servir de puente entre los fundamentos clásicos y las herramientas intermedias y aplicadas que hoy constituyen el núcleo de muchas investigaciones empíricas en economía, ciencias sociales, salud pública y políticas públicas.

En particular, el libro se enfoca en cuatro grandes bloques temáticos que reflejan escenarios reales de análisis económico:

## 1.1 Modelos con Variable Dependiente Limitada

Muchos fenómenos de interés no se expresan naturalmente como variables continuas y sin restricciones. Por ejemplo, las decisiones de participación laboral, la elección entre múltiples opciones educativas, o el acceso a servicios financieros, son todos ejemplos donde el uso de modelos lineales tradicionales es inadecuado. Este capítulo introduce modelos como el Logit, Probit, Tobit y multinomial, proporcionando una base conceptual y práctica para modelar variables cualitativas o censuradas.

## 1.2 Combinación Independiente de Cortes Transversales

Ante la escasez de datos panel, los investigadores recurren a combinaciones de cortes transversales para analizar tendencias o realizar inferencias en contextos donde los datos longitudinales no están disponibles. Aquí se discuten estrategias para integrar y modelar múltiples encuestas independientes, cuidando problemas de heterogeneidad no observada y diferencias estructurales entre periodos.

## 1.3 Modelos con Datos de Panel

Los datos de panel permiten controlar la heterogeneidad individual no observable y estudiar dinámicas temporales con mayor precisión. Este capítulo desarrolla las herramientas fundamentales para trabajar con paneles balanceados y no balanceados, introduciendo modelos de

efectos fijos, efectos aleatorios y modelos dinámicos, junto con criterios de elección entre ellos.

## **1.4 Introducción a la Evaluación de Impacto**

Evaluar el efecto causal de intervenciones o políticas públicas es una tarea central en la econometría aplicada. Este capítulo ofrece una introducción a las metodologías de evaluación de impacto, como el enfoque de diferencias en diferencias (DiD), la asignación aleatoria, los modelos de regresión discontinua y las estrategias de variables instrumentales. Se enfatiza la identificación causal y la interpretación adecuada de los resultados.

A lo largo del texto, se pone especial énfasis en la implementación práctica de los métodos econométricos utilizando software estadístico, análisis de datos reales y ejercicios interpretativos. Cada capítulo está diseñado para proporcionar tanto el aparato teórico como las herramientas empíricas necesarias para enfrentar problemas aplicados de proyección y análisis económico.

Este libro está pensado como un recurso intermedio entre los cursos introductorios de econometría y los enfoques avanzados de inferencia causal y econometría estructural. Es ideal para estudiantes de pregrado avanzado, posgrado o profesionales interesados en fortalecer su capacidad analítica con herramientas modernas y relevantes.

## 2 Resumen

Este libro de Proyección Econométrica ofrece una guía intermedia-aplicada para abordar problemas empíricos más allá del modelo clásico de regresión. Está dirigido a estudiantes y profesionales que buscan aplicar técnicas como modelos con variable dependiente limitada, combinación de cortes transversales, datos de panel y evaluación de impacto. Con un enfoque práctico y apoyado en ejemplos reales en R, el texto vincula la teoría con la implementación, promoviendo una práctica econométrica rigurosa, contextualizada y comprometida con el análisis de fenómenos económicos complejos.

## 3 Variable dependiente binaria

### 3.1 El modelo de probabilidad lineal (MPL)

¿Qué ocurre cuando se desea usar la regresión múltiple para explicar eventos cualitativos?

El caso más sencillo es un evento del tipo binario, es decir, que  $y$  toma valores de cero y uno. Por ejemplo,  $y$  puede indicar si una persona trabaja o no trabaja, esta empleado o desempleado o, si una empresa es grande o pequeña. En cualquier caso se puede hacer que  $y = 1$  denota uno de los resultados o  $y = 0$  denota el otro resultado. También podría pensarse como éxito y fracaso. La demostración parte de la función de regresión poblacional.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u \quad [1]$$

Como  $y$  solo puede tomar dos valores, los  $\beta_j$  **no** pueden interpretarse como un cambio en  $y$  para un aumento de  $x_j$  **ceteris paribus**. Recordar que en este caso  $y$  cambia de cero a uno, o no cambia. Partiendo del supuesto de **media condicional cero**  $E(u|X_1, X_2, \dots, X_k) = 0$ , tenemos:

$$E(y|\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u \quad [2]$$

El punto clave es que  $y$  es una variable binaria que toma valores de cero y uno, entonces tenemos que  $P(y = 1|\mathbf{X}) = E(y|\mathbf{X})$ : la probabilidad de “**éxito**”. Por lo tanto, tenemos

$$P(y = 1|\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u \quad [3]$$

La ecuación [3] dice que, la probabilidad de éxito, es decir  $p(x) = P(y = 1|\mathbf{X})$ , es una función lineal de las variables  $x_j$ , también se le conoce como la **probabilidad de respuesta**. Dado que, las probabilidades deben sumar uno,  $P(y = 0|\mathbf{X}) = 1 - P(y = 1|\mathbf{X})$ , es también una función lineal de las  $x_j$

Por lo tanto, a un modelo de regresión lineal múltiple en el que la variable dependiente es una variable binaria se le conoce como: **El modelo de probabilidad lineal (MPL)**, porque la probabilidad de respuesta es lineal a los parámetros  $\beta_j$ . En el MPL, los  $\beta_j$  miden la variación de la probabilidad de éxito de variar  $x_j$  **ceteris paribus**:

$$\Delta P(y = 1|X) = \beta_j \Delta x_j \quad [4]$$

### 3.1.1 Ejemplo en la clase

Los determinantes del desempleo

$$P(de = 1|educ) = \beta_0 + \beta_1 educ + u \quad ej1$$

$$P(de = 1|educ) = 0.23 - 1.04educ + u \quad ej2$$

Donde:

- de: Desempleo, 1 si estas desempleado. 0 otro caso
- educ: años de educación

El modelo de regresión lineal múltiple permite estimar el efecto de diversas variables explicativas sobre un evento cualitativo. Entonces la mecánica de los **MCO** es la misma de siempre:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k \quad [5]$$

Donde:

- $\hat{y}$ : es la probabilidad de éxito predicha
- $\hat{\beta}_0$ : es la probabilidad de éxito cuando cada una de las  $x_j = 0$
- $\hat{\beta}_1$ : mide la variación de la probabilidad de éxito predicha cuando  $x_1$  varia en una unidad, mientras las demás permanecen constantes.
- $\hat{\beta}_j$ : mide la variación de la probabilidad de éxito predicha cuando  $x_j$  varia en una unidad, mientras las demás permanecen constantes.

**¡¡Concepto clave: para interpretar correctamente un MPL, debe saberse qué es lo que constituye el éxito!!**

**Recomendación:** la variable dependiente debe describir el nombre del evento cuando  $y = 1$

Por ejemplo, si estudiamos los determinantes del desempleo, la variable  $y$  debe llamarse desempleo



### 3.2 Ejemplo 1: Determinantes de la denegación en solicitudes de hipoteca en el mercado inmobiliario (Stock and Watson 2012)

- $denegar = 1$  le negaron la hipoteca y 0 otro caso

Variable explicativa:

- $\frac{P}{I}$ : Ratio Pagos-ingresos

```
pacman::p_load(AER,
                stargazer,
                sandwich)

data("HMDA")
str(HMDA)
```

```
'data.frame':  2380 obs. of  14 variables:
 $ deny      : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ pirat     : num  0.221 0.265 0.372 0.32 0.36 ...
 $ hirat     : num  0.221 0.265 0.248 0.25 0.35 ...
 $ lvrat     : num  0.8 0.922 0.92 0.86 0.6 ...
 $ chist     : Factor w/ 6 levels "1","2","3","4",...: 5 2 1 1 1 1 1 2 2 2 ...
 $ mhist     : Factor w/ 4 levels "1","2","3","4": 2 2 2 2 1 1 2 2 2 1 ...
 $ phist     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ unemp     : num  3.9 3.2 3.2 4.3 3.2 ...
 $ selfemp   : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ insurance : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ condomin  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 2 1 1 1 ...
 $ afam      : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ single    : Factor w/ 2 levels "no","yes": 1 2 1 1 1 1 2 1 1 2 ...
 $ hschool   : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
```

El objetivo es estimar la siguiente ecuación

$$P(Y = 1|x_1) = \beta_0 + \beta_1 x_1 + u \quad [6]$$

ahora usando nuestra ejemplo:

$$P(deny = 1|P/I) = \beta_0 + \beta_1 P/I + u \quad [7]$$

Miremos las variables que ingresan. Primero hacemos binaria a la variable *deny*

```
HMDA$deny <- ifelse(HMDA$deny=="yes",
                    1,
                    0)
# Estimar el modelo de probabilidad lineal

denymod1 <- lm(deny~pirat,
               HMDA)
stargazer(denymod1,
          type = "text")
```

```
=====
                        Dependent variable:
-----
                        deny
-----
pirat                    0.604***
                        (0.061)

Constant                -0.080***
                        (0.021)

-----
Observations              2,380
R2                        0.040
Adjusted R2              0.039
Residual Std. Error      0.318 (df = 2378)
F Statistic              98.406*** (df = 1; 2378)
=====
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

Seria  $\widehat{\beta}_1 = 0.604 \times 0.01 \approx 0.06$

Una vez que tengo el MPL, puedo graficar

```
plot(x = HMDA$pirat,
     y = HMDA$deny,
     main = "Gráfico de dispersión de las denegaciones de hipoteca y el ratio pagos-ingresos",
     xlab = "ratio P/I",
     ylab = "Denegar",
     pch = 20,
     ylim = c(-0.4,1.4),
```

```

    cex.main = 0.8)

# Añadir las líneas horizontales

abline(h=1, lty = 2, col = "darkred")
abline(h=0, lty = 2, col = "darkred")

text(2.5, 0.9,
     cex = 0.8,
     "Hipoteca denegada")

text(2.5, -0.1,
     cex = 0.8,
     "Hipoteca concedida")

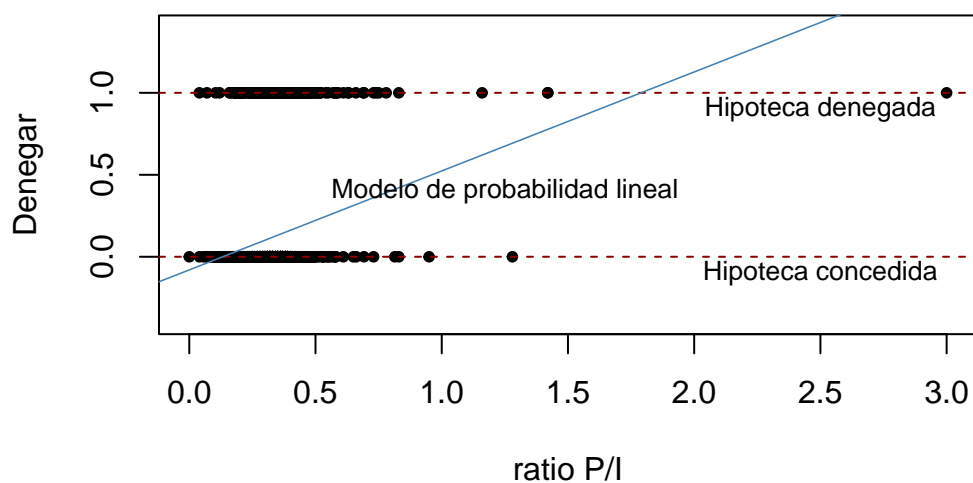
# Añadiendo la línea del MPL

abline(denymod1,
      lwd = 0.8,
      col = "steelblue")

text(1.25, 0.4,
     cex = 0.8,
     "Modelo de probabilidad lineal")

```

### Gráfico de dispersión de las denegaciones de hipoteca y el ratio pagos-ingresos



Presentación de la regresión

```
library(stargazer)
library(sandwich)
eshr <- list(sqrt(diag(vcovHC(denymod1,
                             type = "HC1"))))

stargazer(denymod1,denymod1,
          type = "text",
          se = eshr)
```

Dependent variable:		
	deny	
	(1)	(2)
pirat	0.604*** (0.098)	0.604*** (0.061)
Constant	-0.080** (0.032)	-0.080*** (0.021)

```

-----
Observations                2,380          2,380
R2                          0.040          0.040
Adjusted R2                 0.039          0.039
Residual Std. Error (df = 2378) 0.318          0.318
F Statistic (df = 1; 2378)    98.406***      98.406***
=====
Note:                       *p<0.1; **p<0.05; ***p<0.01

```

Como los modelos de regresión lineal simple poseen el problema de sesgo de variable omitida, y de debido a que el gráfico muestra comportamientos que no son solo explicados por la variable independiente (*pirat*), se añade otra variable que puede ayudar a explicar el fenómeno. La variable es respecto a la consecución de hipoteca a las personas negras (*afam*)

```

colnames(HMDA)[13] <- "negra"

denymod2 <- lm(deny ~
               pirat +
               negra,
               data = HMDA)

eshr <- list(sqrt(diag(vcovHC(denymod1,
                              type = "HC1"))),
             sqrt(diag(vcovHC(denymod2,
                              type = "HC1"))))

stargazer(denymod1,
          denymod2,
          type = "text",
          se = eshr,
          df = F,
          report = "vcts*")

```

```

=====
Dependent variable:
-----
               deny
(1)              (2)
-----

```

pirat	0.604 t = 6.128 (0.098)***	0.597 t = 6.247 (0.096)***
negrayes		0.047 t = 3.387 (0.014)***
Constant	-0.080 t = -2.500 (0.032)**	-0.096 t = -3.105 (0.031)***

Observations	2,380	2,380
R2	0.040	0.045
Adjusted R2	0.039	0.044
Residual Std. Error	0.318	0.318
F Statistic	98.406***	55.614***

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Miremos los valores ajustados

$\hat{y}_i$

```
# modelo con un solo regresor
summary(denymod1$fitted.values)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.07991	0.08908	0.11926	0.11975	0.14340	1.73070

```
# modelo con dos regresores
summary(denymod2$fitted.values)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.09614	0.08800	0.11874	0.11975	0.14859	1.69459

### 3.2.1 Porcentaje predicho correctamente

Mirada a los valores ajustados

```
denymod2$fitted.values
```

1	2	3	4	5
0.0357740938	0.1088880900	0.1259074499	0.0948681447	0.1187445298
6	7	8	9	10
0.0471153745	0.1596254083	0.0709917596	0.0888990484	0.0581507716
11	12	13	14	15
0.1178417344	0.1357490232	0.0709917596	0.1187445298	0.0113007969
16	17	18	19	20
0.1187445298	0.0411462782	0.0948681447	0.0888990484	0.1715636008
21	22	23	24	25
0.1247136260	0.1366518186	0.1417181195	0.1366518186	0.1476872157
26	27	28	29	30
0.1178417344	0.1715636008	0.1297799269	0.0888990484	0.1059035418
31	32	33	34	35
0.1655945046	0.1247136260	0.1596254083	0.1715636008	0.1068063372
36	37	38	39	40
0.1306827223	0.1664973000	0.1127754335	0.1306827223	0.1715636008
41	42	43	44	45
0.1247136260	0.1485900111	0.1127754335	0.1485900111	0.1536563120
46	47	48	49	50
0.0769608559	0.0650226633	0.1715636008	0.1068063372	0.1655945046
51	52	53	54	55
0.1536563120	0.0650226633	0.0462125791	0.1835017934	0.1187445298
56	57	58	59	60
0.0912866846	0.1068063372	0.1488810396	0.1476872157	0.1888739891
61	62	63	64	65
0.1080001610	0.1751450495	0.0954650452	0.0703948477	0.0930774204
66	67	68	69	70
0.0739763077	0.1217290779	0.1050156129	0.1596254083	0.1608192321
71	72	73	74	75
0.1127754335	0.1091939848	0.1178417344	0.1294888985	0.0578597432
76	77	78	79	80
0.0339833581	0.0411462782	0.0739763077	0.0620381152	0.1655945046
81	82	83	84	85
0.0662164871	0.1655945046	0.0888990484	0.1649976041	0.1408301905
86	87	88	89	90
0.1068063372	0.0345802700	0.1417181195	0.0948681447	0.1178417344

91	92	93	94	95
0.1354579948	0.0999344456	0.1068063372	0.1417181195	0.0900928722
96	97	98	99	100
0.1476872157	0.0948681447	0.1306827223	0.0948681447	0.0411462782
101	102	103	104	105
0.0650226633	0.1381366594	0.1596254083	0.1417181195	0.1175507060
106	107	108	109	110
0.1596254083	0.1068063372	0.1357490232	0.0769608559	0.2679719366
111	112	113	114	115
0.0608442914	0.1070973656	0.1187445298	0.2252854673	0.0859145003
116	117	118	119	120
0.1799203447	0.0614412033	0.1008372410	0.2312545636	0.0172698931
121	122	123	124	125
0.1297799269	0.1187445298	0.1476872157	0.0879962530	0.1485900111
126	127	128	129	130
0.1187445298	0.1485900111	0.1008372410	0.1238108307	0.1485900111
131	132	133	134	135
0.1008372410	0.0411462782	0.1366518186	0.1366518186	0.0650226633
136	137	138	139	140
0.1799203447	0.1787265209	0.1247136260	0.1447026676	0.1882770659
141	142	143	144	145
0.1864863415	0.1936492616	0.1303768388	0.1685790527	0.1094850019
146	147	148	149	150
0.0638288395	0.0644257514	0.1948430854	0.0877052246	0.1363459351
151	152	153	154	155
0.1187445298	0.1366518186	0.0948681447	0.0113007969	0.1068063372
156	157	158	159	160
0.1464933919	0.1798009601	0.1056125134	0.0715886715	0.1848746828
161	162	163	164	165
0.1205352541	0.0505774499	0.1319362271	0.0948681447	0.0745732196
166	167	168	169	170
0.1094850019	0.0918835965	0.1325853981	0.1524624882	0.1691759532
171	172	173	174	175
0.1253105266	0.1312796228	0.1596254083	0.1068063372	0.0621574998
176	177	178	179	180
0.1596254083	0.1835017934	0.1835017934	0.0829299521	0.0769608559
181	182	183	184	185
0.1127754335	0.1835017934	0.1655945046	0.1059035418	0.1655945046
186	187	188	189	190
0.0650226633	0.1638037802	0.0829299521	0.1366518186	0.0888990484
191	192	193	194	195
0.1605282037	0.1426209149	0.1536563120	0.1462023635	0.0411462782
196	197	198	199	200



0.3864510667	0.1775326971	0.1793234214	0.0948681447	0.0787515802
201	202	203	204	205
0.0941444263	0.0665075042	0.1261984669	0.1787265209	0.1774133125
206	207	208	209	210
0.1043590087	0.1836808590	0.1851731331	0.0709917596	0.0888990484
211	212	213	214	215
0.1799800257	0.1500748634	0.1238108307	0.1482244353	0.1518058840
216	217	218	219	220
0.1081792266	0.1313915857	0.0943309251	0.0508759001	0.0745732196
221	222	223	224	225
0.1118726381	0.1190355582	0.1443445136	0.1115816097	0.1273922907
226	227	228	229	230
0.0947487600	0.1895305706	0.0687831890	0.1610580014	-0.0222455235
231	232	233	234	235
0.1536563120	0.0301631402	0.0725437260	0.0763639440	0.0829299521
236	237	238	239	240
0.1187445298	0.0948681447	0.0948681447	0.0590535670	0.0590535670
241	242	243	244	245
0.1417181195	0.2014090822	0.1536563120	0.1247136260	0.1238108307
246	247	248	249	250
0.1596254083	0.1894708897	0.1655945046	0.1306827223	0.1247136260
251	252	253	254	255
0.1954399859	0.1417181195	0.1008372410	0.1008372410	0.1238108307
256	257	258	259	260
0.1715636008	0.1655945046	0.1247136260	0.1297799269	0.1068063372
261	262	263	264	265
0.1545591074	0.0232389894	0.0829299521	0.1306827223	0.1536563120
266	267	268	269	270
0.0769608559	0.0948681447	0.1187445298	0.1485900111	0.1417181195
271	272	273	274	275
0.1306827223	0.0700889642	0.1178417344	0.0530844708	0.0769608559
276	277	278	279	280
0.2431927561	0.1008372410	0.1724663962	0.1954399859	0.1127754335
281	282	283	284	285
0.1127754335	-0.0483901659	0.0999344456	0.1903736851	0.1187445298
286	287	288	289	290
0.0948681447	0.0650226633	0.1068063372	0.1476872157	0.0709917596
291	292	293	294	295
0.1068063372	0.1715636008	0.1297799269	0.1187445298	0.1158196626
296	297	298	299	300
0.0999344456	-0.0125755882	0.1664973000	0.1724663962	0.1247136260
301	302	303	304	305
0.1068063372	0.0652614326	0.0411462782	0.1536563120	0.1476872157

306	307	308	309	310
0.1745481490	0.1042396240	0.0697382548	0.1485900111	0.1204755732
311	312	313	314	315
0.0881827518	0.1223259784	0.1386738790	0.0681265961	0.1127754335
316	317	318	319	320
0.1492988745	0.0948681447	0.0615605879	0.1066272716	0.1066272716
321	322	323	324	325
0.1191026837	0.1287726133	0.0835268640	0.2730382375	0.1133723340
326	327	328	329	330
0.1829048929	0.1605282037	0.1605282037	-0.0066064920	0.1536563120
331	332	333	334	335
0.1357490232	0.1247136260	0.0888990484	0.0829299521	-0.0006373957
336	337	338	339	340
0.1187445298	0.1187445298	0.1426209149	0.1715636008	0.0709917596
341	342	343	344	345
0.1605282037	0.1187445298	0.0471153745	0.1844045888	-0.0015401911
346	347	348	349	350
0.0471153745	0.0351771819	-0.0125755882	0.0292080857	-0.0961429361
351	352	353	354	355
0.0761774451	0.1393379165	0.1297799269	0.1187445298	0.0709917596
356	357	358	359	360
0.0590535670	0.1178417344	0.0855563577	0.0910479267	0.0843028415
361	362	363	364	365
0.0229405391	0.0677087612	0.1405242957	0.1724663962	0.1168941017
366	367	368	369	370
0.1023220818	0.0943234919	0.0709917596	0.1994989732	0.1826661236
371	372	373	374	375
0.1072838759	0.1118726381	0.1127754335	0.1127754335	0.1417181195
376	377	378	379	380
0.1417181195	0.1715636008	0.1596254083	0.2133472748	0.1426209149
381	382	383	384	385
0.1238108307	0.1954399859	0.0530844708	0.1894708897	0.1605282037
386	387	388	389	390
0.1485900111	0.1963427813	0.1954399859	0.0590535670	0.1697728765
391	392	393	394	395
0.1835017934	0.1715636008	0.1476872157	0.0709917596	0.1655945046
396	397	398	399	400
0.1127754335	0.0948681447	0.1306827223	0.1127754335	0.1187445298
401	402	403	404	405
0.0530844708	0.2312545636	0.0113007969	0.1238108307	0.0530844708
406	407	408	409	410
0.0590535670	0.0769608559	0.1238108307	0.1118726381	0.1238108307
411	412	413	414	415

0.0292080857	0.4709212099	0.0530844708	0.0351771819	0.1247136260
416	417	418	419	420
0.1426209149	0.1238108307	0.1655945046	0.1417181195	0.1306827223
421	422	423	424	425
0.1247136260	0.1426209149	0.0939653493	0.1247136260	0.1596254083
426	427	428	429	430
0.1485900111	0.1485900111	0.1417181195	0.1068063372	0.1835017934
431	432	433	434	435
0.0465184626	0.1014341415	0.1626099564	0.1479334069	0.0491971272
436	437	438	439	440
0.0871083241	0.0441308264	0.0906897727	0.1094850019	0.0417431901
441	442	443	444	445
0.0853175884	0.1109847092	0.1602223088	0.0954650452	0.1312796228
446	447	448	449	450
0.2324483874	0.2043936303	0.0942712328	-0.0075689797	0.1074032377
451	452	453	454	455
0.1121785330	0.1056125134	0.1402332673	0.1835017934	0.1148571862
456	457	458	459	460
0.1360474735	0.2300607398	0.1358161488	0.0483091983	0.3169185078
461	462	463	464	465
0.2476770002	0.1152227621	0.2366267593	0.1163568822	0.1673852289
466	467	468	469	470
0.0614412033	0.0539723997	0.1217290779	0.1924554378	0.0542782946
471	472	473	474	475
0.0987406218	0.1840986939	0.1196324587	0.1321675632	-0.0627159975
476	477	478	479	480
0.1799203447	0.2993022702	0.1076942662	0.1015461043	0.0924805084
481	482	483	484	485
0.1644006808	0.1241167255	0.0832209805	0.1017251699	0.0136884331
486	487	488	489	490
0.1894708897	0.1536563120	0.1319884976	0.1485900111	0.0581507716
491	492	493	494	495
0.1426209149	0.1357490232	0.2202191664	0.1451802062	0.1849343865
496	497	498	499	500
0.0382214224	0.1316974578	0.1521640380	0.0462125791	0.1297799269
501	502	503	504	505
0.0103980015	0.0232389894	-0.0015401911	0.0103980015	0.0053317006
506	507	508	509	510
0.1238108307	0.1715636008	0.1655945046	0.0530844708	0.1187445298
511	512	513	514	515
0.1366518186	0.1775326971	0.1894708897	0.1655945046	0.1306827223
516	517	518	519	520
0.1476872157	0.1247136260	0.0351771819	0.0829299521	0.1655945046

521	522	523	524	525
0.1485900111	0.1596254083	0.1417181195	0.1485900111	0.0948681447
526	527	528	529	530
0.1426209149	0.1476872157	0.0709917596	0.1127754335	0.1426209149
531	532	533	534	535
0.0829299521	0.1596254083	0.0232389894	0.0769608559	0.1187445298
536	537	538	539	540
0.1536563120	0.0590535670	0.0709917596	0.1605282037	0.0650226633
541	542	543	544	545
0.1178417344	0.0650226633	0.1655945046	0.1127754335	0.1476872157
546	547	548	549	550
0.0948681447	0.0411462782	0.1247136260	0.1015535261	0.1604013972
551	552	553	554	555
0.1766447682	0.1142080266	0.1995586542	0.0346399624	0.0731928970
556	557	558	559	560
0.2259420715	0.1107459399	0.0928983434	0.1263849657	0.3452642933
561	562	563	564	565
0.1164762668	0.1273997239	0.0912866846	0.1798606410	0.1195801996
566	567	568	569	570
0.1097237598	0.0961216494	0.0418028825	0.1682806024	0.1584987100
571	572	573	574	575
0.0768414712	0.4367108464	0.1723992707	0.0807810738	0.1148646081
576	577	578	579	580
0.1112831594	0.2489827868	0.1272803393	0.1830316994	0.0129647147
581	582	583	584	585
0.0965394844	0.1257209396	0.1041799431	0.1525892947	0.0575612930
586	587	588	589	590
0.1173716404	0.1824944798	0.0327895457	0.0692010353	0.0650226633
591	592	593	594	595
0.0948681447	0.0999344456	0.2052964257	0.0948681447	0.1626099564
596	597	598	599	600
0.1357490232	0.1297799269	0.0641198679	0.0113007969	-0.0483901659
601	602	603	604	605
0.1187445298	-0.0125755882	0.1068063372	0.1068063372	0.0948681447
606	607	608	609	610
0.0172698931	0.0948681447	0.3864510667	0.2500646478	0.1297799269
611	612	613	614	615
-0.0125755882	0.0590535670	0.1605282037	0.1844045888	0.0530844708
616	617	618	619	620
0.0829299521	0.0650226633	0.1187445298	0.0524875589	0.0327895457
621	622	623	624	625
0.5962722317	0.1187445298	0.0697979358	0.0948681447	0.0757670321
626	627	628	629	630

0.1921644094	0.0900928722	0.1279892026	0.1560439596	0.1199383536
631	632	633	634	635
0.0739763077	0.0966588690	0.1450085625	0.0924805084	0.1226170068
636	637	638	639	640
0.1175507060	0.0769608559	0.1127754335	0.1068063372	0.1187445298
641	642	643	644	645
0.0650226633	0.1724663962	0.0948681447	0.1476872157	0.1784354925
646	647	648	649	650
0.1596254083	0.1596254083	0.1366518186	0.1784354925	0.1306827223
651	652	653	654	655
0.0650226633	0.1306827223	0.0351771819	0.1775326971	0.1247136260
656	657	658	659	660
0.1247136260	0.0709917596	0.0999344456	0.0590535670	0.0650226633
661	662	663	664	665
0.1775326971	0.1127754335	0.0530844708	0.1306827223	0.1247136260
666	667	668	669	670
0.1118726381	0.0471153745	0.1187445298	0.0879962530	0.1008372410
671	672	673	674	675
0.0709917596	0.1476872157	0.1715636008	0.0948681447	0.1118726381
676	677	678	679	680
0.0709917596	0.1366518186	0.1306827223	0.1306827223	0.1775326971
681	682	683	684	685
0.0829299521	0.1366518186	0.1068063372	0.0172698931	0.1068063372
686	687	688	689	690
0.1476872157	0.0888990484	0.0769608559	0.0232389894	0.1724663962
691	692	693	694	695
0.0948681447	0.0232389894	0.0650226633	0.0283052903	-0.0066064920
696	697	698	699	700
-0.0364519733	0.1306827223	0.0351771819	0.0351771819	-0.0185446845
701	702	703	704	705
-0.0075092874	0.1655945046	0.1655945046	0.0509952848	0.0345802700
706	707	708	709	710
0.0906897727	0.1288919980	0.0229331059	0.0942712328	0.1169538054
711	712	713	714	715
0.1763388733	0.0787515802	0.1312796228	0.0897869773	0.1220201063
716	717	718	719	720
0.0500999226	0.0799454040	0.1384425429	0.1008372410	0.0817361283
721	722	723	724	725
0.1444116392	0.1342641710	0.1787265209	0.1139692573	0.1333613870
726	727	728	729	730
0.1426805958	0.0650226633	0.1805172453	0.0304019095	0.0083162487
731	732	733	734	735
0.1295411576	0.0578597432	0.0626350271	0.0957560736	-0.0036219438

736	737	738	739	740
0.0829299521	0.1420240144	0.0983824792	0.1396363667	0.1655945046
741	742	743	744	745
0.1605282037	0.1366518186	0.1008372410	0.1127754335	0.1008372410
746	747	748	749	750
0.1605282037	0.1238108307	0.0709917596	0.1894708897	0.1247136260
751	752	753	754	755
0.1008372410	0.0709917596	0.1187445298	0.2193163710	0.0829299521
756	757	758	759	760
0.1775326971	0.1238108307	0.1008372410	0.0292080857	0.0888990484
761	762	763	764	765
0.0641198679	0.1655945046	0.1127754335	0.1715636008	0.1536563120
766	767	768	769	770
0.1008372410	0.1357490232	0.1571183988	0.1545591074	0.1775326971
771	772	773	774	775
0.1596254083	0.1476872157	0.0471153745	0.1183789540	0.1417181195
776	777	778	779	780
-0.0006373957	0.0530844708	0.1596254083	0.1775326971	0.1417181195
781	782	783	784	785
0.1596254083	0.1118726381	0.0530844708	0.0769608559	0.1125889347
786	787	788	789	790
0.1417181195	0.1357490232	0.1715636008	0.2969146226	0.1008372410
791	792	793	794	795
0.0462125791	0.1775326971	0.1068063372	0.1366518186	0.1724663962
796	797	798	799	800
0.2551309487	0.1605282037	0.1357490232	0.2252854673	0.0939653493
801	802	803	804	805
0.0351771819	0.0351771819	0.0581507716	0.3446673928	0.2560337441
806	807	808	809	810
0.1596254083	0.2611000449	0.1545591074	0.1485900111	0.1963427813
811	812	813	814	815
0.1784354925	0.0999344456	0.1835017934	0.1485900111	0.1059035418
816	817	818	819	820
0.1655945046	0.1835017934	0.1655945046	0.1017251699	0.4461420294
821	822	823	824	825
0.1068063372	0.0590535670	0.1178417344	0.1536563120	0.1715636008
826	827	828	829	830
0.0411462782	0.0888990484	0.1187445298	0.1485900111	0.1357490232
831	832	833	834	835
0.1664973000	0.1247136260	0.1366518186	0.1306827223	0.2193163710
836	837	838	839	840
0.1068063372	0.1536563120	0.1187445298	0.2261882627	0.1127754335
841	842	843	844	845

0.1008372410	0.1127754335	0.0888990484	0.0530844708	0.1247136260
846	847	848	849	850
0.0650226633	0.0650226633	0.1187445298	0.0292080857	0.1297799269
851	852	853	854	855
0.1068063372	0.0769608559	0.1894708897	0.0829299521	0.1059035418
856	857	858	859	860
0.0709917596	0.1366518186	0.0351771819	0.0715886715	0.1127754335
861	862	863	864	865
0.1199383536	0.1565214755	0.1211321774	0.0089131606	0.1271012737
866	867	868	869	870
0.1181476293	0.1202293706	0.1513283681	0.0560690189	0.1364056274
871	872	873	874	875
0.0984495933	0.1465530957	0.1306827223	0.0960619685	0.0924805084
876	877	878	879	880
0.0850117049	0.0841760350	0.0838178811	0.0515847635	0.2026029060
881	882	883	884	885
0.0985689780	0.0942712328	0.2133472748	0.1318765461	0.0853175884
886	887	888	889	890
0.1536563120	0.1429119433	0.1300858218	0.2918483217	0.1411212076
891	892	893	894	895
0.1470977371	0.1651169659	0.0206125838	0.1273922907	0.1638037802
896	897	898	899	900
0.1667883284	0.1333613870	0.1148571862	0.0873470820	0.1441057671
901	902	903	904	905
0.0399524544	0.1276384932	0.1444116392	0.1748540439	0.0697979358
906	907	908	909	910
0.0753491971	0.0626350271	0.1500748634	0.0545693116	0.1026279653
911	912	913	914	915
0.1026279653	0.1318765461	0.0017502405	0.0945622612	0.1279892026
916	917	918	919	920
0.1435088438	0.0769608559	0.1074032377	0.1978276336	0.0626350271
921	922	923	924	925
0.1055453992	0.1962831004	0.0626350271	0.0357740938	0.1448294741
926	927	928	929	930
0.0292080857	0.1321675632	0.1829048929	0.1276981742	0.1435088438
931	932	933	934	935
0.0942712328	0.0841237759	0.1664301744	0.0829299521	0.0286111738
936	937	938	939	940
0.1900677902	0.1008372410	0.0853175884	-0.0089941282	0.1715636008
941	942	943	944	945
0.0471153745	0.0292080857	0.0590535670	0.1775326971	0.0590535670
946	947	948	949	950
0.0113007969	0.0471153745	0.1178417344	0.0530844708	0.2193163710

951	952	953	954	955
0.1297799269	0.1068063372	0.1297799269	0.1178417344	0.1306827223
956	957	958	959	960
0.1068063372	0.1127754335	0.1476872157	0.1417181195	0.0820271567
961	962	963	964	965
0.1297799269	0.0590535670	0.1536563120	0.1178417344	0.1476872157
966	967	968	969	970
0.1426209149	0.0590535670	0.0232389894	0.1536563120	0.1008372410
971	972	973	974	975
0.1127754335	0.1655945046	0.1417181195	0.0948681447	0.0351771819
976	977	978	979	980
0.0709917596	0.0709917596	0.1357490232	0.1247136260	0.0948681447
981	982	983	984	985
0.1366518186	0.1366518186	0.0581507716	0.1775326971	0.1775326971
986	987	988	989	990
0.1059035418	0.1655945046	0.1118726381	0.0471153745	0.0530844708
991	992	993	994	995
0.1178417344	0.1366518186	0.1306827223	-0.0066064920	0.0411462782
996	997	998	999	1000
0.0232389894	0.0948681447	0.1127754335	0.2252854673	0.0292080857
1001	1002	1003	1004	1005
0.0888990484	0.1238108307	0.1357490232	0.1008372410	0.1008372410
1006	1007	1008	1009	1010
0.1605282037	0.1247136260	0.1536563120	0.2073781785	0.1655945046
1011	1012	1013	1014	1015
0.3694465733	0.1775326971	0.1247136260	0.1247136260	0.1545591074
1016	1017	1018	1019	1020
0.1008372410	0.1954399859	0.1306827223	0.1476872157	0.1835017934
1021	1022	1023	1024	1025
0.1118726381	0.3267601040	0.1605282037	0.1476872157	0.1775326971
1026	1027	1028	1029	1030
0.1247136260	0.0760580605	0.0760580605	0.0411462782	0.2909455263
1031	1032	1033	1034	1035
0.1118726381	0.1008372410	0.1844045888	0.1417181195	0.0769608559
1036	1037	1038	1039	1040
0.1844045888	0.2202191664	0.0888990484	0.1366518186	0.1187445298
1041	1042	1043	1044	1045
0.2014090822	0.1187445298	0.1596254083	0.1059035418	0.0351771819
1046	1047	1048	1049	1050
-0.0125755882	0.0948681447	0.1178417344	0.1715636008	0.0709917596
1051	1052	1053	1054	1055
0.1238108307	0.0650226633	0.0999344456	0.0590535670	0.0829299521
1056	1057	1058	1059	1060



0.2023118776	0.1306827223	0.1426209149	0.1068063372	0.0948681447
1061	1062	1063	1064	1065
0.0590535670	0.0232389894	0.0769608559	0.1417181195	0.1835017934
1066	1067	1068	1069	1070
0.0641198679	0.1247136260	0.1485900111	0.1775326971	0.1008372410
1071	1072	1073	1074	1075
0.1590285078	0.1330703699	0.0888990484	0.1118726381	0.1402332673
1076	1077	1078	1079	1080
0.0590535670	0.1241167255	0.0888990484	0.1145661578	0.0686041234
1081	1082	1083	1084	1085
0.0888990484	0.1775326971	0.1247136260	0.1187445298	0.1127754335
1086	1087	1088	1089	1090
0.1655945046	0.1402332673	0.1655945046	0.1187445298	0.1118726381
1091	1092	1093	1094	1095
0.0888990484	0.1068063372	0.0769608559	0.1545591074	1.6945859460
1096	1097	1098	1099	1100
0.1476872157	0.1417181195	0.0829299521	0.0769608559	0.0709917596
1101	1102	1103	1104	1105
0.1247136260	0.1247136260	0.0641198679	0.1187445298	0.4461420294
1106	1107	1108	1109	1110
0.1238108307	0.1715636008	0.0292080857	0.0760580605	0.1238108307
1111	1112	1113	1114	1115
0.0760580605	0.2014090822	0.0351771819	0.0650226633	0.2073781785
1116	1117	1118	1119	1120
0.1247136260	0.1238108307	0.0053317006	0.1536563120	-0.0006373957
1121	1122	1123	1124	1125
-0.0185446845	0.0888990484	0.0411462782	0.0700889642	0.0292080857
1126	1127	1128	1129	1130
0.0471153745	0.0351771819	0.0999344456	0.1536563120	0.1426209149
1131	1132	1133	1134	1135
0.0769608559	0.0769608559	-0.0125755882	0.1485900111	0.1715636008
1136	1137	1138	1139	1140
0.2073781785	0.0590535670	0.0879962530	0.0333864576	0.1715636008
1141	1142	1143	1144	1145
0.1020310648	0.1297799269	0.1253105266	0.1127754335	0.0521816754
1146	1147	1148	1149	1150
0.0601280062	0.0345802700	0.1115816097	0.1187445298	0.1605282037
1151	1152	1153	1154	1155
0.0650226633	0.1187445298	0.0650226633	0.1357490232	0.1127754335
1156	1157	1158	1159	1160
0.1655945046	0.0590535670	0.1187445298	0.0948681447	0.0709917596
1161	1162	1163	1164	1165
0.1536563120	0.0999344456	0.1127754335	0.0223361940	0.1008372410

1166	1167	1168	1169	1170
0.0590535670	0.0700889642	0.1357490232	0.0700889642	0.1247136260
1171	1172	1173	1174	1175
0.1357490232	0.1775326971	0.0769608559	0.1127754335	0.1894708897
1176	1177	1178	1179	1180
0.1068063372	0.1008372410	0.0888990484	0.1954399859	0.0471153745
1181	1182	1183	1184	1185
0.0939653493	0.1306827223	0.1476872157	0.0948681447	0.1247136260
1186	1187	1188	1189	1190
0.1306827223	0.1357490232	0.1357490232	0.1127754335	0.0888990484
1191	1192	1193	1194	1195
0.0939653493	0.1127754335	0.1715636008	0.0829299521	0.1357490232
1196	1197	1198	1199	1200
0.0053317006	0.1247136260	0.0471153745	0.1596254083	0.1596254083
1201	1202	1203	1204	1205
0.1417181195	0.1008372410	0.1127754335	0.1835017934	0.1178417344
1206	1207	1208	1209	1210
0.1426209149	0.2312545636	0.1775326971	0.1835017934	0.0948681447
1211	1212	1213	1214	1215
0.1247136260	0.1963427813	0.0053317006	0.1008372410	0.1008372410
1216	1217	1218	1219	1220
0.1118726381	0.1417181195	0.1127754335	0.1008372410	0.0769608559
1221	1222	1223	1224	1225
0.1417181195	0.1715636008	0.1059035418	0.1596254083	0.0590535670
1226	1227	1228	1229	1230
0.0888990484	0.1596254083	0.2790073338	0.0888990484	0.0471153745
1231	1232	1233	1234	1235
0.1536563120	0.0829299521	0.1008372410	0.1715636008	0.0530844708
1236	1237	1238	1239	1240
0.0769608559	0.1417181195	0.1187445298	0.0939653493	-0.0364519733
1241	1242	1243	1244	1245
0.1068063372	0.2073781785	0.1835017934	0.1357490232	0.1715636008
1246	1247	1248	1249	1250
0.0888990484	0.1715636008	0.0471153745	0.1247136260	0.1596254083
1251	1252	1253	1254	1255
0.0769608559	0.0769608559	-0.0543592622	0.1187445298	0.0530844708
1256	1257	1258	1259	1260
0.0292080857	-0.0185446845	0.1536563120	0.1366518186	0.1297799269
1261	1262	1263	1264	1265
0.0769608559	0.0053317006	0.1247136260	0.1664973000	0.1247136260
1266	1267	1268	1269	1270
0.0530844708	0.1059035418	0.3983892592	0.0551662235	0.0113007969
1271	1272	1273	1274	1275

0.1127754335	0.1154540981	0.1178417344	0.1008372410	0.0581507716
1276	1277	1278	1279	1280
0.2372236599	0.0172698931	0.1894708897	0.2321573590	0.1835017934
1281	1282	1283	1284	1285
0.3169185078	0.1462023635	0.1966338097	0.0829299521	0.1115816097
1286	1287	1288	1289	1290
0.1447026676	0.1181476293	0.0829299521	0.0632319390	0.1476872157
1291	1292	1293	1294	1295
0.0859145003	0.1417181195	0.1247136260	0.0495030107	0.0972557923
1296	1297	1298	1299	1300
0.1127754335	0.1306827223	0.0939653493	0.0900928722	0.0948681447
1301	1302	1303	1304	1305
0.1655945046	0.0089131606	0.1357490232	0.0650226633	0.1297799269
1306	1307	1308	1309	1310
0.1485900111	0.0172698931	0.0232389894	0.1894708897	0.1417181195
1311	1312	1313	1314	1315
0.1127754335	0.1715636008	0.1127754335	0.1008372410	-0.0066064920
1316	1317	1318	1319	1320
0.1068063372	0.1118726381	0.1068063372	0.0829299521	0.1417181195
1321	1322	1323	1324	1325
0.6679013869	0.1178417344	0.0471153745	0.0939653493	0.1306827223
1326	1327	1328	1329	1330
0.0948681447	0.1596254083	0.0999344456	0.2133472748	-0.0245137808
1331	1332	1333	1334	1335
0.0760580605	0.0471153745	0.1187445298	0.1417181195	0.1545591074
1336	1337	1338	1339	1340
0.1715636008	0.1476872157	0.0650226633	0.1297799269	0.2372236599
1341	1342	1343	1344	1345
0.0172698931	0.0172698931	0.0888990484	0.1187445298	0.0292080857
1346	1347	1348	1349	1350
0.1008372410	0.1655945046	0.1715636008	0.1127754335	0.2014090822
1351	1352	1353	1354	1355
0.1068063372	0.0471153745	0.1775326971	0.1306827223	0.0113007969
1356	1357	1358	1359	1360
0.1068063372	0.1476872157	0.0530844708	0.1306827223	0.1127754335
1361	1362	1363	1364	1365
0.1775326971	0.1715636008	0.1306827223	0.1187445298	0.1068063372
1366	1367	1368	1369	1370
0.1187445298	0.0411462782	0.1297799269	0.0769608559	0.1366518186
1371	1372	1373	1374	1375
0.1187445298	0.1178417344	0.1306827223	0.0590535670	0.1664973000
1376	1377	1378	1379	1380
0.1068063372	0.1068063372	0.1306827223	0.1306827223	0.1596254083

1381	1382	1383	1384	1385
-0.0006373957	0.0888990484	0.1187445298	0.1357490232	0.2193163710
1386	1387	1388	1389	1390
0.1835017934	0.0700889642	0.0471153745	0.1724663962	0.3148219114
1391	1392	1393	1394	1395
0.1118726381	0.2321573590	0.1417181195	0.0521816754	0.1844045888
1396	1397	1398	1399	1400
-0.0125755882	0.1775326971	0.0888990484	0.2202191664	0.1059035418
1401	1402	1403	1404	1405
0.0232389894	0.1476872157	0.0700889642	0.2142500702	0.1476872157
1406	1407	1408	1409	1410
0.0641198679	0.0793484921	0.1256015550	0.1518655877	0.0238359013
1411	1412	1413	1414	1415
0.1082911781	0.1176626574	0.2133472748	0.0634110046	0.1954399859
1416	1417	1418	1419	1420
0.0736181651	0.1297799269	0.1118726381	0.1492988745	0.0872873897
1421	1422	1423	1424	1425
0.1223856821	0.2128100552	0.0223361940	0.1187445298	0.1485900111
1426	1427	1428	1429	1430
0.1476872157	0.1306827223	0.1417181195	0.1417181195	0.1685790527
1431	1432	1433	1434	1435
0.0709917596	0.1127754335	0.0417431901	0.0888990484	-0.0280952408
1436	1437	1438	1439	1440
0.0829299521	0.1220201063	0.1605282037	0.1127754335	0.1476872157
1441	1442	1443	1444	1445
0.1545591074	0.2491618524	0.1894708897	0.1775326971	0.1476872157
1446	1447	1448	1449	1450
0.0820271567	0.1306827223	0.1306827223	0.1306827223	0.1775326971
1451	1452	1453	1454	1455
0.1306827223	0.1306827223	0.1306827223	0.1306827223	0.1306827223
1456	1457	1458	1459	1460
0.0411462782	0.0590535670	-0.0066064920	0.0053317006	0.0709917596
1461	1462	1463	1464	1465
0.1297799269	0.1297799269	0.0521816754	0.1775326971	0.1247136260
1466	1467	1468	1469	1470
0.1596254083	0.2482739235	0.1715636008	0.1229229017	0.1945520570
1471	1472	1473	1474	1475
0.1420240144	0.1199383536	0.1545591074	0.1611251042	0.1187445298
1476	1477	1478	1479	1480
0.1715636008	0.1536563120	0.1297799269	0.1775326971	0.1306827223
1481	1482	1483	1484	1485
0.1417181195	0.0530844708	0.1118726381	0.1894708897	0.1187445298
1486	1487	1488	1489	1490

0.0948681447	0.1775326971	0.0650226633	0.0829299521	0.1476872157
1491	1492	1493	1494	1495
0.0709917596	0.0769608559	0.0948681447	0.0888990484	0.1835017934
1496	1497	1498	1499	1500
0.0951069140	0.1228631980	0.1485303302	0.0590535670	0.1835017934
1501	1502	1503	1504	1505
0.0140988405	0.0948681447	0.1187445298	0.1306827223	0.0709917596
1506	1507	1508	1509	1510
0.1306827223	0.1357490232	0.1596254083	-0.0006373957	0.1476872157
1511	1512	1513	1514	1515
0.1835017934	0.0590535670	0.1187445298	0.1068063372	0.1187445298
1516	1517	1518	1519	1520
0.1476872157	0.1187445298	0.1247136260	0.0351771819	0.1068063372
1521	1522	1523	1524	1525
0.1068063372	0.1426209149	-0.0364519733	0.1894708897	0.0709917596
1526	1527	1528	1529	1530
0.1187445298	0.1835017934	0.0820271567	0.1655945046	0.1068063372
1531	1532	1533	1534	1535
0.1417181195	0.1366518186	0.1357490232	0.1247136260	0.1545591074
1536	1537	1538	1539	1540
0.0650226633	0.1715636008	0.1417181195	0.1238108307	0.1247136260
1541	1542	1543	1544	1545
0.1655945046	0.1187445298	0.1008372410	0.1238108307	0.0590535670
1546	1547	1548	1549	1550
0.0769608559	0.0948681447	0.0292080857	0.1417181195	0.0769608559
1551	1552	1553	1554	1555
0.1655945046	0.1187445298	0.0820271567	0.0530844708	0.1596254083
1556	1557	1558	1559	1560
0.1059035418	0.1476872157	0.0888990484	0.1068063372	0.0829299521
1561	1562	1563	1564	1565
0.1008372410	0.1366518186	0.1655945046	0.0709917596	0.1059035418
1566	1567	1568	1569	1570
0.1596254083	0.0769608559	0.1068063372	0.0829299521	0.0999344456
1571	1572	1573	1574	1575
0.0709917596	0.1655945046	0.1596254083	0.0888990484	0.0590535670
1576	1577	1578	1579	1580
0.0650226633	0.0948681447	0.1357490232	0.1306827223	0.0650226633
1581	1582	1583	1584	1585
0.0650226633	0.0879962530	0.1008372410	0.1297799269	0.1247136260
1586	1587	1588	1589	1590
0.1596254083	0.0829299521	0.3625746816	0.0471153745	0.0530844708
1591	1592	1593	1594	1595
0.1118726381	0.1238108307	0.1068063372	0.1008372410	0.2491618524

1596	1597	1598	1599	1600
0.2014090822	0.0948681447	0.1664973000	0.1775326971	0.1536563120
1601	1602	1603	1604	1605
0.0829299521	0.1835017934	0.0709917596	0.1417181195	0.1485900111
1606	1607	1608	1609	1610
0.1844045888	0.1596254083	0.0948681447	0.0769608559	0.1417181195
1611	1612	1613	1614	1615
0.0999344456	0.1068063372	0.1655945046	0.1476872157	0.0999344456
1616	1617	1618	1619	1620
0.0999344456	0.0650226633	0.0999344456	0.0411462782	0.0641198679
1621	1622	1623	1624	1625
0.0530844708	0.1059035418	0.0113007969	0.1068063372	0.1655945046
1626	1627	1628	1629	1630
0.1059035418	0.1775326971	0.1426209149	0.1297799269	0.0351771819
1631	1632	1633	1634	1635
0.0172698931	0.1835017934	0.1118726381	0.1655945046	0.0650226633
1636	1637	1638	1639	1640
0.1476872157	0.1417181195	0.1426209149	0.1187445298	0.0888990484
1641	1642	1643	1644	1645
-0.0364519733	0.1118726381	0.1118726381	0.1715636008	0.1366518186
1646	1647	1648	1649	1650
0.0948681447	0.1306827223	0.1247136260	0.1536563120	0.1366518186
1651	1652	1653	1654	1655
0.0888990484	0.1127754335	0.0829299521	0.1118726381	0.0530844708
1656	1657	1658	1659	1660
0.1247136260	0.1068063372	0.1187445298	0.1476872157	0.1306827223
1661	1662	1663	1664	1665
0.1357490232	0.1297799269	0.1178417344	0.0888990484	0.0888990484
1666	1667	1668	1669	1670
0.1775326971	0.1068063372	0.1118726381	0.1178417344	0.0939653493
1671	1672	1673	1674	1675
0.1178417344	0.1357490232	0.0760580605	0.1536563120	0.0700889642
1676	1677	1678	1679	1680
0.1417181195	0.1426209149	0.1178417344	0.0888990484	0.0888990484
1681	1682	1683	1684	1685
0.1596254083	0.1247136260	0.0948681447	0.1068063372	0.1068063372
1686	1687	1688	1689	1690
0.1008372410	0.0351771819	0.1715636008	0.1963427813	0.1655945046
1691	1692	1693	1694	1695
0.1187445298	0.1536563120	0.1476872157	0.1178417344	0.1476872157
1696	1697	1698	1699	1700
0.1476872157	0.1187445298	0.1127754335	0.1715636008	0.0948681447
1701	1702	1703	1704	1705

0.1238108307	0.0650226633	0.1366518186	0.2014090822	0.1357490232
1706	1707	1708	1709	1710
0.1596254083	0.0769608559	0.0769608559	0.0709917596	0.1008372410
1711	1712	1713	1714	1715
0.2073781785	0.0888990484	0.1059035418	0.1068063372	0.1963427813
1716	1717	1718	1719	1720
0.1297799269	0.1476872157	0.1357490232	0.2193163710	0.0411462782
1721	1722	1723	1724	1725
-0.0304828771	0.1715636008	0.0939653493	0.0172698931	0.1059035418
1726	1727	1728	1729	1730
0.1008372410	0.0650226633	0.0709917596	0.1247136260	0.0709917596
1731	1732	1733	1734	1735
0.2014090822	0.1366518186	0.0590535670	0.1008372410	0.1187445298
1736	1737	1738	1739	1740
0.1476872157	0.1844045888	0.2909455263	0.1655945046	0.1596254083
1741	1742	1743	1744	1745
0.0700889642	-0.0304828771	0.1247136260	0.0462125791	0.1297799269
1746	1747	1748	1749	1750
0.1127754335	0.1417181195	0.1008372410	-0.0185446845	0.0888990484
1751	1752	1753	1754	1755
0.0650226633	0.2849764300	0.1357490232	0.1357490232	0.1536563120
1756	1757	1758	1759	1760
0.0709917596	0.1247136260	0.1357490232	0.1417181195	0.1476872157
1761	1762	1763	1764	1765
0.0530844708	0.1844045888	0.0590535670	0.0948681447	0.1835017934
1766	1767	1768	1769	1770
0.1417181195	0.0530844708	-0.0304828771	0.1118726381	0.1357490232
1771	1772	1773	1774	1775
0.1655945046	0.0471153745	0.1655945046	0.1366518186	0.1596254083
1776	1777	1778	1779	1780
0.1476872157	0.1536563120	0.2551309487	0.0829299521	0.0471153745
1781	1782	1783	1784	1785
0.1297799269	0.1596254083	0.0879962530	0.0879962530	0.1187445298
1786	1787	1788	1789	1790
0.0709917596	0.1476872157	0.0888990484	-0.0254165762	0.1127754335
1791	1792	1793	1794	1795
0.0820271567	0.0590535670	0.0769608559	0.1357490232	0.1417181195
1796	1797	1798	1799	1800
0.1775326971	0.1068063372	0.1068063372	0.1238108307	0.1357490232
1801	1802	1803	1804	1805
0.0351771819	0.1536563120	0.1357490232	0.0888990484	0.0411462782
1806	1807	1808	1809	1810
0.1844045888	0.2014090822	0.1366518186	0.2014090822	0.1835017934

1811	1812	1813	1814	1815
0.1476872157	0.1178417344	0.2142500702	-0.0543592622	0.2133472748
1816	1817	1818	1819	1820
0.1187445298	0.2014090822	0.2133472748	0.1784354925	0.1844045888
1821	1822	1823	1824	1825
0.1306827223	0.1903736851	0.0709917596	0.1247136260	0.1596254083
1826	1827	1828	1829	1830
0.1426209149	0.1655945046	0.1068063372	0.1724663962	0.1426209149
1831	1832	1833	1834	1835
0.1306827223	0.0113007969	0.1306827223	0.1715636008	0.0462125791
1836	1837	1838	1839	1840
0.1715636008	0.0232389894	0.1417181195	0.1417181195	0.1655945046
1841	1842	1843	1844	1845
0.1068063372	0.1536563120	0.1187445298	0.0172698931	0.1536563120
1846	1847	1848	1849	1850
0.0879962530	-0.0125755882	0.1127754335	0.1008372410	0.0581507716
1851	1852	1853	1854	1855
0.1059035418	0.1068063372	0.1059035418	0.1059035418	0.0351771819
1856	1857	1858	1859	1860
0.0769608559	0.1238108307	0.0709917596	0.0530844708	0.1357490232
1861	1862	1863	1864	1865
0.1655945046	0.0471153745	0.1068063372	0.1247136260	0.1068063372
1866	1867	1868	1869	1870
0.0709917596	0.1187445298	0.1187445298	0.1127754335	0.0471153745
1871	1872	1873	1874	1875
0.1306827223	0.1127754335	0.1426209149	0.1426209149	0.1187445298
1876	1877	1878	1879	1880
0.1306827223	0.1426209149	0.1664973000	0.1485900111	0.1247136260
1881	1882	1883	1884	1885
0.1008372410	0.1366518186	0.1417181195	0.1306827223	0.0760580605
1886	1887	1888	1889	1890
0.2133472748	0.0172698931	0.1366518186	0.1426209149	0.1664973000
1891	1892	1893	1894	1895
0.1536563120	0.1655945046	0.0650226633	0.1366518186	0.1187445298
1896	1897	1898	1899	1900
0.1715636008	0.1187445298	0.1605282037	0.1366518186	0.1008372410
1901	1902	1903	1904	1905
0.2073781785	0.1068063372	0.0700889642	0.1178417344	0.2014090822
1906	1907	1908	1909	1910
0.1835017934	0.3924201630	0.0650226633	0.0939653493	0.1247136260
1911	1912	1913	1914	1915
0.2073781785	0.1536563120	0.2969146226	0.1426209149	0.1545591074
1916	1917	1918	1919	1920



0.2014090822	0.1247136260	0.1366518186	0.1835017934	0.1178417344
1921	1922	1923	1924	1925
0.1536563120	0.0530844708	0.1127754335	0.1008372410	0.1247136260
1926	1927	1928	1929	1930
0.0590535670	0.0471153745	0.7983187096	0.7983187096	0.1366518186
1931	1932	1933	1934	1935
0.1545591074	0.1485900111	0.1247136260	0.1059035418	0.1247136260
1936	1937	1938	1939	1940
0.2073781785	0.1894708897	0.2312545636	0.1187445298	0.1536563120
1941	1942	1943	1944	1945
0.0769608559	0.1835017934	0.0888990484	0.1426209149	0.1366518186
1946	1947	1948	1949	1950
0.1485900111	0.1775326971	0.1008372410	0.1835017934	0.0769608559
1951	1952	1953	1954	1955
0.1187445298	0.1187445298	0.1844045888	0.1536563120	0.1187445298
1956	1957	1958	1959	1960
0.1536563120	0.1775326971	0.1306827223	0.1426209149	0.2312545636
1961	1962	1963	1964	1965
0.1127754335	0.1485900111	0.1775326971	0.1306827223	0.0769608559
1966	1967	1968	1969	1970
0.1187445298	0.1835017934	0.1954399859	0.3267601040	0.1059035418
1971	1972	1973	1974	1975
0.0709917596	0.1357490232	0.0411462782	0.1545591074	0.1238108307
1976	1977	1978	1979	1980
0.0948681447	0.1596254083	0.1605282037	0.1485900111	0.1247136260
1981	1982	1983	1984	1985
0.1366518186	0.2252854673	0.1306827223	0.1008372410	0.1068063372
1986	1987	1988	1989	1990
0.0650226633	0.1187445298	0.1247136260	0.0530844708	0.1127754335
1991	1992	1993	1994	1995
0.0590535670	0.1068063372	0.1014341415	0.1276981742	0.0905629662
1996	1997	1998	1999	2000
0.1846956172	0.1118800600	0.0596504789	0.1429119433	0.0333864576
2001	2002	2003	2004	2005
-0.0209323208	0.0906897727	0.0972557923	0.1312796228	0.1127754335
2006	2007	2008	2009	2010
0.0098010896	0.0814302448	0.1342641710	0.1199383536	0.0811392278
2011	2012	2013	2014	2015
0.1420240144	0.0751701315	0.1163568822	0.1109847092	0.1327644751
2016	2017	2018	2019	2020
0.0805423159	0.1265043504	0.0915777131	0.1032248886	0.1363459351
2021	2022	2023	2024	2025
-0.0107848582	0.0787515802	0.0659105922	0.1624905718	0.1074032377

2026	2027	2028	2029	2030
0.1253105266	0.1536563120	0.1469709306	0.1139692573	0.1387932637
2031	2032	2033	2034	2035
0.1125366642	0.1566408602	0.1303768388	0.1050156129	0.1363459351
2036	2037	2038	2039	2040
0.1761598077	0.0954650452	0.1176104097	0.0990465166	0.1211321774
2041	2042	2043	2044	2045
0.0918239042	0.2145410986	0.1644006808	0.0972557923	0.1799203447
2046	2047	2048	2049	2050
0.1817110691	0.1195801996	0.1044186896	0.1579540686	0.0820271567
2051	2052	2053	2054	2055
0.0931967936	0.2599062211	0.1350327266	0.1005387907	0.0996434171
2056	2057	2058	2059	2060
0.1655945046	0.0912866846	0.1354579948	0.1721605014	0.1097908854
2061	2062	2063	2064	2065
0.1715636008	0.0965991881	0.1229229017	0.0817361283	0.1223259784
2066	2067	2068	2069	2070
0.0763639440	0.1253105266	0.0948681447	0.0883021365	0.1151630811
2071	2072	2073	2074	2075
0.1360549181	0.1345551994	0.1632068570	0.0930774204	0.0721855834
2076	2077	2078	2079	2080
0.0805423159	0.1235198022	0.1187445298	0.1267953788	0.1271535214
2081	2082	2083	2084	2085
0.1193414303	0.0703948477	0.0787515802	0.1262655811	0.0694994855
2086	2087	2088	2089	2090
0.1265640541	0.0136884331	0.0952262987	0.1217290779	0.1247136260
2091	2092	2093	2094	2095
0.0942712328	0.0868098624	0.1512686644	0.1211321774	0.0841237759
2096	2097	2098	2099	2100
0.1612370670	0.1513954937	-0.0030250319	0.0703948477	0.1870832420
2101	2102	2103	2104	2105
0.1287129096	0.1074032377	0.0434071023	0.1223259784	0.1715636008
2106	2107	2108	2109	2110
0.1638037802	0.1166479106	0.1057916018	0.0823330402	0.0894959603
2111	2112	2113	2114	2115
0.1336672705	0.1649976041	0.0978526928	0.1109847092	0.1441057671
2116	2117	2118	2119	2120
0.0811392278	0.0083759410	0.1239376372	0.0971364076	0.0892571910
2121	2122	2123	2124	2125
0.1159987510	0.0894959603	0.1211321774	0.1524624882	0.1047171626
2126	2127	2128	2129	2130
0.0381617301	0.0972557923	0.1439266787	0.0999344456	0.1655945046
2131	2132	2133	2134	2135

0.1127754335	0.1127754335	0.1596254083	0.1031592190	0.0834074794
2136	2137	2138	2139	2140
0.1315706513	0.1102684240	0.1414270911	0.0049138657	0.1072838759
2141	2142	2143	2144	2145
0.1210127927	0.0188815519	0.0099279074	0.1187445298	0.1420165697
2146	2147	2148	2149	2150
0.1664973000	0.1250643354	0.1396363667	0.1247136260	0.1664973000
2151	2152	2153	2154	2155
0.1715636008	0.1261387746	0.1693027825	0.1596254083	0.1649976041
2156	2157	2158	2159	2160
0.1835017934	0.1068063372	0.1388006969	0.0888990484	0.1307349814
2161	2162	2163	2164	2165
0.1589762487	0.1136111033	0.1366518186	0.1123575986	0.1260865155
2166	2167	2168	2169	2170
0.1040008547	0.1357564450	0.1550963270	0.0733793958	0.1151630811
2171	2172	2173	2174	2175
0.1890604766	0.1484706265	0.0198962987	0.1811215904	0.1546710475
2176	2177	2178	2179	2180
0.1403452301	0.1347342764	0.1618339675	0.1857700563	0.1379650271
2181	2182	2183	2184	2185
0.0235374397	0.2368655059	0.1366443854	-0.0283936911	0.0942115404
2186	2187	2188	2189	2190
0.1137901689	0.0496223954	0.0939130902	0.0654330650	0.1554544810
2191	2192	2193	2194	2195
0.1063885023	-0.0317363875	0.1106191334	0.1176700906	0.1703697770
2196	2197	2198	2199	2200
0.1022026971	0.0800647887	0.0788112725	0.1290636304	0.1523431036
2201	2202	2203	2204	2205
0.1239376372	0.1900081092	0.1127754335	0.1838002436	0.1895902743
2206	2207	2208	2209	2210
0.1398154323	0.1205352541	0.1229229017	0.1512089835	0.0953382387
2211	2212	2213	2214	2215
0.0516518890	0.1369502689	0.0961813531	0.1427328549	0.2053486848
2216	2217	2218	2219	2220
0.1353908806	0.1051349976	0.1059035418	0.1177297943	0.1142677076
2221	2222	2223	2224	2225
0.1196921510	0.0952859796	0.1614235772	0.1421359544	0.1414793502
2226	2227	2228	2229	2230
0.1423821456	0.0955247489	0.1799203447	0.0641198679	-0.0254165762
2231	2232	2233	2234	2235
0.1575959147	0.1927613327	0.1469112269	0.0749313622	0.0923611238
2236	2237	2238	2239	2240
0.1540218878	0.1762791924	0.2133472748	0.1429716242	0.1266834387

2241	2242	2243	2244	2245
0.0781546797	0.0847803688	0.0871680050	0.0948681447	0.1019116801
2246	2247	2248	2249	2250
0.1060303483	0.0945100021	0.0779756027	0.1005387907	0.1324734467
2251	2252	2253	2254	2255
0.0555914917	0.1393304832	0.1262655811	0.1645797691	0.1008372410
2256	2257	2258	2259	2260
0.1775923781	0.1398080105	0.1337866551	0.1568199258	0.0809601508
2261	2262	2263	2264	2265
0.0684847387	0.1431581344	0.1629681104	0.1310931240	0.0342818198
2266	2267	2268	2269	2270
0.0608442914	0.0966588690	0.1161181356	0.0877052246	0.1154540981
2271	2272	2273	2274	2275
0.1220872319	0.1022101304	0.1223185566	0.0603667641	0.1613564517
2276	2277	2278	2279	2280
0.1148049271	0.1303768388	0.1263252848	0.1596254083	0.1596254083
2281	2282	2283	2284	2285
0.0053317006	0.0521816754	0.0829299521	0.1068063372	0.2491618524
2286	2287	2288	2289	2290
0.0760580605	0.1187445298	0.0172698931	0.0948681447	0.1127754335
2291	2292	2293	2294	2295
0.0760580605	0.0709917596	0.0829299521	0.1187445298	0.0650226633
2296	2297	2298	2299	2300
0.1306827223	0.1536563120	0.1118726381	0.1247136260	0.0848997534
2301	2302	2303	2304	2305
0.0820942823	0.1248330107	0.1217290779	0.1223259784	0.1476872157
2306	2307	2308	2309	2310
0.1894708897	0.0650226633	0.0769608559	0.0888990484	0.1775326971
2311	2312	2313	2314	2315
0.1715636008	0.1775326971	0.1536563120	0.2014090822	0.0948681447
2316	2317	2318	2319	2320
0.1342641710	0.1578346840	0.0495030107	0.0411462782	0.1068063372
2321	2322	2323	2324	2325
0.1127754335	0.1769357966	0.1835017934	0.2014090822	0.0590535670
2326	2327	2328	2329	2330
0.1605282037	0.1068063372	0.0232389894	0.1187445298	0.0948681447
2331	2332	2333	2334	2335
0.1596254083	0.1844045888	0.1366518186	0.1068063372	0.1306827223
2336	2337	2338	2339	2340
0.1703697770	0.0829299521	0.1372487191	0.0829299521	0.1954399859
2341	2342	2343	2344	2345
0.1068063372	0.1059035418	0.1127754335	-0.0424210696	0.0590535670
2346	2347	2348	2349	2350

0.1664973000	0.1596254083	0.1366518186	0.0888990484	0.1426209149
2351	2352	2353	2354	2355
0.1485900111	0.0471153745	0.1008372410	0.1247136260	0.1127754335
2356	2357	2358	2359	2360
0.1545591074	0.1008372410	0.1068063372	0.1187445298	0.1417181195
2361	2362	2363	2364	2365
0.1417181195	0.0650226633	0.1008372410	0.1306827223	0.1187445298
2366	2367	2368	2369	2370
0.0939653493	0.0948681447	0.1775326971	0.1306827223	0.0999344456
2371	2372	2373	2374	2375
0.1417181195	0.1357490232	0.1008372410	0.1596254083	0.1008372410
2376	2377	2378	2379	2380
0.0888990484	0.1297799269	0.0590535670	0.1417181195	0.1596254083

```
# Porcentaje predicho correctamente
ppc <- data.frame(denymod2$model$deny, denymod2$fitted.values)
head(ppc)
```

	denymod2.model.deny	denymod2.fitted.values
1	0	0.03577409
2	0	0.10888809
3	0	0.12590745
4	0	0.09486814
5	0	0.11874453
6	0	0.04711537

```
# Cambiar de nombre
names(ppc) <- c("deny", "VA")
head(ppc)
```

	deny	VA
1	0	0.03577409
2	0	0.10888809
3	0	0.12590745
4	0	0.09486814
5	0	0.11874453
6	0	0.04711537

```
# Creando la variable y virgulilla
```

```
ppc$y.c <- ifelse(ppc$VA>0.5,1,0)
```

```
head(ppc)
```

	deny	VA	y.c
1	0	0.03577409	0
2	0	0.10888809	0
3	0	0.12590745	0
4	0	0.09486814	0
5	0	0.11874453	0
6	0	0.04711537	0

```
tail(ppc)
```

	deny	VA	y.c
2375	0	0.10083724	0
2376	0	0.08889905	0
2377	0	0.12977993	0
2378	0	0.05905357	0
2379	1	0.14171812	0
2380	1	0.15962541	0

```
# Creando la variable PPC
```

```
ppc$ppc <- ifelse(ppc$deny==ppc$y.c,1,0)
```

```
head(ppc)
```

	deny	VA	y.c	ppc
1	0	0.03577409	0	1
2	0	0.10888809	0	1
3	0	0.12590745	0	1
4	0	0.09486814	0	1
5	0	0.11874453	0	1
6	0	0.04711537	0	1

```
# Calculando el PPC
```

```
prop.table(table(ppc$ppc))*100
```

	0	1
	11.84874	88.15126

### 3.3 Ejemplo 2: Determinantes del trabajo femenino (Wooldridge 2009)

- $y$ : **inlf**: (la fuerza de trabajo femenino), una variable binaria que indica, si una mujer casada participó en la fuerza de trabajo durante 1975:  $inlf = 1$  la mujer informa haber trabajado fuera de la casa, por un salario ese año, cero otro caso
- $x_1 = nwifwinc$  los ingreso del esposo en miles de dólares (-)
- $x_2 = educ$  años de educación (+)
- $x_3 = exper$  años de experiencia (+)
- $x_4 = exper^2$  años de experiencia al cuadrado (-)
- $x_5 = edad$  en años (-)
- $x_6 = kidslt6$  hijo < 6 años (-)
- $x_7 = kidsge6$  hijos entres 6 y 18 años (+)

```
data("mroz", package = "wooldridge")
str(mroz)
```

```
'data.frame': 753 obs. of 22 variables:
 $ inlf      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ hours     : int  1610 1656 1980 456 1568 2032 1440 1020 1458 1600 ...
 $ kidslt6   : int  1 0 1 0 1 0 0 0 0 0 ...
 $ kidsge6   : int  0 2 3 3 2 0 2 0 2 2 ...
 $ age       : int  32 30 35 34 31 54 37 54 48 39 ...
 $ educ      : int  12 12 12 12 14 12 16 12 12 12 ...
 $ wage      : num  3.35 1.39 4.55 1.1 4.59 ...
 $ repwage   : num  2.65 2.65 4.04 3.25 3.6 ...
 $ hushrs    : int  2708 2310 3072 1920 2000 1040 2670 4120 1995 2100 ...
 $ husage    : int  34 30 40 53 32 57 37 53 52 43 ...
 $ huseduc   : int  12 9 12 10 12 11 12 8 4 12 ...
 $ huswage   : num  4.03 8.44 3.58 3.54 10 ...
 $ faminc    : num  16310 21800 21040 7300 27300 ...
 $ mtr       : num  0.721 0.661 0.692 0.781 0.622 ...
 $ motheduc  : int  12 7 12 7 12 14 14 3 7 7 ...
```

```

$ fatheduc: int  7 7 7 7 14 7 7 3 7 7 ...
$ unem      : num  5 11 5 5 9.5 7.5 5 5 3 5 ...
$ city      : int  0 1 0 0 1 1 0 0 0 0 ...
$ exper     : int  14 5 15 6 7 33 11 35 24 21 ...
$ nwifeinc: num  10.9 19.5 12 6.8 20.1 ...
$ lwage     : num  1.2102 0.3285 1.5141 0.0921 1.5243 ...
$ expersq   : int  196 25 225 36 49 1089 121 1225 576 441 ...
- attr(*, "time.stamp")= chr "25 Jun 2011 23:03"

```

```
# Correr el modelo y crear el objeto
```

```
mroz.mpl <- lm(inlf~
               nwifeinc+
               educ+
               exper+
               I(exper^2)+
               age+
               kidslt6+
               kidsge6,
               data = mroz)
```

```
# usando stargazer
```

```
library(stargazer)
```

```
stargazer(mroz.mpl,
           type = "text",
           digits = 5)
```

```
=====
                        Dependent variable:
                        -----
                                inlf
                        -----
nwifeinc                    -0.00341**
                               (0.00145)

educ                        0.03800***
                               (0.00738)

exper                       0.03949***
                               (0.00567)
```



I(exper2)	-0.00060*** (0.00018)
age	-0.01609*** (0.00248)
kidslt6	-0.26181*** (0.03351)
kidsge6	0.01301 (0.01320)
Constant	0.58552*** (0.15418)

```
-----
Observations      753
R2                0.26422
Adjusted R2       0.25730
Residual Std. Error 0.42713 (df = 745)
F Statistic       38.21795*** (df = 7; 745)
=====
```

Note:                \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```
# Punto de inflexión
abs((coefficients(mroz.mpl)[4])/(coefficients(mroz.mpl)[5]*2))
```

```
exper
33.11387
```

```
# Probar la multicolinealidad aproximada
library(car)

mean(vif(mroz.mpl))
```

```
[1] 3.417547
```

```
# Normalidad de los errores

library(tseries)
jarque.bera.test(mroz.mpl$residuals)
```

### Jarque Bera Test

```
data: mroz.mpl$residuals
X-squared = 36.741, df = 2, p-value = 1.051e-08
```

$$H_0 : u \sim N(\mu, \sigma^2)$$

### 3.3.1 El porcentaje predicho correctamente

Es una medida de bondad de ajuste

```
PPC.DTF <- data.frame(mroz$inlf, mroz.mpl$fitted.values)
names(PPC.DTF) <- c("inlf", "VA.inlf")
PPC.DTF$ajuste <- ifelse(mroz.mpl$fitted.values>=0.5,1,0)
PPC.DTF$PPC <- ifelse(PPC.DTF$inlf==PPC.DTF$ajuste,1,0)
prop.table(table(PPC.DTF$PPC))*100
```

0	1
26.56042	73.43958

### 3.3.2 Interpretaciones ceteris paribus

- Para interpretar las estimaciones, hay que recordar que una variación en la variable independiente modifica la probabilidad de que  $inlf = 1$ . Por ejemplo, *educ* si las demás variables permanecen constantes, una año más de educación hace que la probabilidad de participación en la fuerza laboral aumente en 3.8%. Si consideramos de forma literal a esta ecuación, entonces 10 años más educación incrementarían la probabilidad de permanecer en la fuerza laboral en 38%.
- El coeficiente de **nwifeinc** significa que si  $\Delta nwifeinc = 10$  (un incremento de \$10,000), la probabilidad de que una mujer permanecer en la fuerza de trabajo disminuye en 3.4%. Como se puede ver, esta disminución es pequeña a pesar de aumentar el salario en 10,000 dólares.
- La experiencia ha sido introducida como una función cuadrática para que el efecto de la experiencia sea decreciente sobre la probabilidad de participar en la fuerza laboral. *Ceteris paribus*, la variación de la probabilidad se aproxima como  $0.039 - 2(0.0006)exper = 0.039 - 0.0012exper$ . El punto en el que la experiencia transcurrida no tiene efecto sobre

la probabilidad de participación en la fuerza laboral es:  $0.039/0.0012 = 32.5$ . Sólo 13 mujeres de las 753 en esta muestra tiene más de 32 años de experiencia.

```
mroz$dico.exper <- ifelse(mroz$exper>32.5,1,0)
table(mroz$dico.exper)
```

```
0    1
740 13
```

- A diferencia de la cantidad de hijos entre 6 y 18 años, la cantidad de hijos menores a 6 años tiene un impacto enorme sobre la probabilidad de participación en la fuerza de trabajo. A tal punto que, tener un hijo menor a seis años adicional, reduce la probabilidad de participación en la fuerza trabajo en 26.18%. En la muestra, menos de 20% de las mujeres tienen al menos un hijo pequeño.

```
mroz$dic.hijo <- ifelse(mroz$kidslt6>=1,1,0)
prop.table(table(mroz$dic.hijo))*100
```

```
      0      1
80.47809 19.52191
```

- Respecto al PPC el modelo predice en 73.44% a la variable **infl**.

### 3.3.3 Límites del MPL

- Las dos desventajas más importantes son que las probabilidades ajustadas pueden ser menores que cero o mayores que uno.

```
summary(mroz.mpl$fitted.values)
```

```
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.3451  0.4016  0.5880  0.5684  0.7592  1.1272
```

Se demuestra para este ejemplo que algunos valores ajustados son menores que cero y mayores que uno.

- y, el efecto parcial de cualquier variable explicativa (si aparece en la ecuación en su nivel) es constante

### 3.4 Ejemplo 3: Un modelo de probabilidad lineal para arrestos (Wooldridge 2009)

Sea **arr86** una variable binaria igual a uno si un hombre fue arrestado en 1986 e igual a cero si no fue así. La población es un grupo de hombres de California nacidos en 1960 o en 1961, que habían sido detenidos al menos una vez antes de 1986. Un modelo de probabilidad lineal para describir **arr86** es:

$$arr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 tottime + \beta_4 ptime86 + \beta_5 qemp86 + u$$

donde:

- *pcnv* = proporción de arrestos previos que condujeron a una condena (+)
- *avgsen* = sentencia promedio cumplida en condenas previas (en meses) (-)
- *tottime* = meses en prisión y desde los 18 años de edad anteriores a 1986 (-)
- *ptime86* = meses en prisión en 1986 (+ -)
- *qemp86* = cantidad de trimestres (0 a 4) que el hombre estuvo empleado legalmente en 1986.(-)

```
data("crime1", package = "wooldridge")
str(crime1)
```

```
'data.frame':  2725 obs. of  16 variables:
 $ narr86 : int  0 2 1 2 1 0 2 5 0 0 ...
 $ nfarr86: int  0 2 1 2 1 0 2 3 0 0 ...
 $ nparr86: int  0 0 0 1 0 0 1 5 0 0 ...
 $ pcnv   : num  0.38 0.44 0.33 0.25 0 ...
 $ avgsen : num  17.6 0 22.8 0 0 ...
 $ tottime: num  35.2 0 22.8 0 0 ...
 $ ptime86: int  12 0 0 5 0 0 0 0 9 0 ...
 $ qemp86 : num  0 1 0 2 2 4 0 0 0 3 ...
 $ inc86  : num  0 0.8 0 8.8 8.1 ...
 $ durat  : num  0 0 11 0 1 ...
 $ black  : int  0 0 1 0 0 0 1 0 1 0 ...
 $ hispan : int  0 1 0 1 0 0 0 0 0 1 ...
 $ born60 : int  1 0 1 1 0 1 1 1 1 1 ...
 $ pcnvsq : num  0.1444 0.1936 0.1089 0.0625 0 ...
 $ pt86sq : int  144 0 0 25 0 0 0 0 81 0 ...
 $ inc86sq: num  0 0.64 0 77.44 65.61 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

```
# Creando la variable y
table(crime1$narr86)
```

```

  0    1    2    3    4    5    6    7    9   10   12
1970 559 121  42  12  13   4   1   1   1   1

```

```

crime1$arr86 <- ifelse(crime1$narr86>0,1,0)

arr86.MPL <- lm(arr86~
                pcnv+
                avgsen+
                tottime+
                ptime86+
                qemp86,
                data = crime1)
stargazer(arr86.MPL,
           digits = 5,
           type = "text")

```

```

=====
                        Dependent variable:
                        -----
                                arr86
                        -----
pcnv                        -0.16244***
                             (0.02124)

avgsen                      0.00611
                             (0.00645)

tottime                    -0.00226
                             (0.00498)

ptime86                   -0.02197***
                             (0.00463)

qemp86                    -0.04283***
                             (0.00540)

```

```

Constant                                0.44062***
                                         (0.01723)

-----
Observations                            2,725
R2                                       0.04735
Adjusted R2                             0.04560
Residual Std. Error      0.43731 (df = 2719)
F Statistic                27.02966*** (df = 5; 2719)
=====
Note:                                *p<0.1; **p<0.05; ***p<0.01

```

### 3.4.1 Porcentaje predicho correctamente

```

PPC.arr <- data.frame(crime1$arr86, arr86.MPL$fitted.values)
names(PPC.arr)<- c("arr86", "valores_ajustados")
PPC.arr$ajuste <- ifelse(PPC.arr$valores_ajustados>=0.5,1,0)
PPC.arr$PPC <-ifelse(PPC.arr$arr86==PPC.arr$ajuste,1,0)
prop.table(table(PPC.arr$PPC))*100

```

```

      0      1
27.70642 72.29358

```

### 3.4.2 Interpretaciones

- $\hat{\beta}_0 = 0.44062$ , es la probabilidad de ser arrestado que se predice a un hombre, que no ha sido condenado, que no ha estado en prisión despues de los 18 años, que no ha estado en prisión en 1986 y que ha estado desempleado todo el año
- $\hat{\beta}_2; \hat{\beta}_3$  que pertenecen a las variables *avg*sen y *tot*time, respectivamente. No son estadísticamente significativas (no tiene asteriscos). El signo de *avg*sen es contra intuitivo, pues se esperaria que condenas más largas disminuyan la probabilidad de ser arrestado en 1986. Con respecto a *tot*time, según los datos haber tenido meses en prisión de los 18 y antes de 1986, disminuye la probabilidad de ser arrestado en 1986.
- **ptimes86** el aumento de probabilidad de ser condenado en 1986, disminuye la probabilidad de ser arrestado en promedio en 2.2%. Si un hombre esta en prisión no puede ser arrestado. Como **ptimes86** esta medida en mese, 6 meses más en prisión, reduce la probabilidad de ser detenido en  $0.0022 \times 6 \approx 0.132$ . En esta variable se puede observar otra vez, que el MPL no cierto en todos los rangos de variables independientes.

Por ejemplo, si un hombre está en prisión durante 12 meses de 1986, no puede ser detenido en 1986. **Ceteris paribus**, cuando  $ptime86 = 12$  la probabilidad predicha es de  $0.44 - 0.22 \times 12 = 0.177$  que es distinta de cero

- **qemp86** tener un empleo reduce la probabilidad de detención de manera significativa. **Ceteris paribus** un hombre que ha sido empleado durante 4 trimestres, la probabilidad de ser detenido se reduce en  $0.04283 \times 4 \approx 0.172$

### 3.5 Incorporando regresores binarios al MPL

En los modelos de variable dependiente binaria, se puede incluir variables independientes binarias. Este coeficiente mide la diferencia que se predice para la probabilidad en la relación con el grupo base. Así, incluimos regresores binarios en el MPL para **arr86**

```
arr86.MPL.bi <- lm(arr86~
  pcnv+
  avgsen+
  tottime+
  ptime86+
  qemp86+
  black+
  hispan,
  data = crime1)

stargazer(arr86.MPL.bi,
  digits = 5,
  type = "text")
```

```
=====
                        Dependent variable:
                        -----
                        arr86
                        -----
pcnv                    -0.15206***
                        (0.02107)

avgsen                  0.00462
                        (0.00639)

tottime                 -0.00256
```

```

(0.00493)

ptime86          -0.02370***
                  (0.00459)

qemp86           -0.03847***
                  (0.00540)

black            0.16976***
                  (0.02367)

hispan           0.09619***
                  (0.02071)

Constant         0.38043***
                  (0.01873)

-----
Observations      2,725
R2               0.06819
Adjusted R2       0.06579
Residual Std. Error 0.43265 (df = 2717)
F Statistic      28.40542*** (df = 7; 2717)
=====
Note:             *p<0.1; **p<0.05; ***p<0.01

```

```
# Limites del MPL
```

```
summary(arr86.MPL.bi$fitted.values)
```

```

      Min.   1st Qu.   Median     Mean  3rd Qu.    Max.
-0.05598  0.21483   0.26501   0.27706  0.36119   0.61273

```

### 3.5.1 Interpretaciones

- El coeficiente **black** significa que, *ceteris paribus* un hombre negro tiene una probabilidad del 17% mayor de ser detenido frente a un hombre blanco. Otra forma de expresar esto, es que la probabilidad de ser detenido es de 17 puntos porcentuales mayor para los negros que para los blancos.

De la misma manera que la versión del modelo de regresión lineal múltiple en el MPL se puede verificar el cumplimiento de los supuestos.



### 3.5.2 Supuestos

#### 3.5.2.1 Homocedasticidad

- La **homocedasticidad** o varianza constante, su incumplimiento se conoce como **heterocedasticidad** o varianza no constante
- Su incumplimiento tiene efecto sobre la eficiencia de los estimadores de MCO.
- Su incumplimiento, hace que las pruebas  $t$  o  $f$  se invaliden, pues el cálculo de la varianza supone homocedasticidad que no se cumple. Por lo tanto, la matriz de varianza covarianza esta mal calculada.
- Existen dos formas de la heterocedasticidad, conocida y desconocida. Es común la forma desconocida, por tal motivo calculamos **errores estándar heterocedástico robustos**

#### Hipótesis

$$H_0 : \sigma^2$$

$$H_a : \sigma_i^2$$

```
# Test de homcedasticidad
library(lmtest)

# El test de Breusch-Pagan
bptest(mroz.mpl)
```

studentized Breusch-Pagan test

```
data:  mroz.mpl
BP = 24.224, df = 7, p-value = 0.00104
```

```
# Test de Goldfeld-Quandt
gqtest(mroz.mpl)
```

# Goldfeld-Quandt test

```
data: mroz.mpl
GQ = 2.488e+27, df1 = 369, df2 = 368, p-value < 2.2e-16
alternative hypothesis: variance increases from segment 1 to 2
```

```
library(sandwich)
rob_es <- list(sqrt(diag(vcovHC(mroz.mpl, type = "HC1"))))

library(stargazer)

stargazer(mroz.mpl, mroz.mpl,
          se = rob_es,
          digits = 5,
          type = "text",
          column.labels = c("Heterocedástico-Robusto", "Homocedástico"))
```

=====		
	Dependent variable:	
	-----	
	inlf	
	Heterocedástico-Robusto	Homocedástico
	(1)	(2)
-----		
nwifeinc	-0.00341** (0.00152)	-0.00341** (0.00145)
educ	0.03800*** (0.00727)	0.03800*** (0.00738)
exper	0.03949*** (0.00581)	0.03949*** (0.00567)
I(exper2)	-0.00060*** (0.00019)	-0.00060*** (0.00018)
age	-0.01609*** (0.00240)	-0.01609*** (0.00248)
kidslt6	-0.26181***	-0.26181***

	(0.03178)	(0.03351)
kidsge6	0.01301 (0.01353)	0.01301 (0.01320)
Constant	0.58552*** (0.15226)	0.58552*** (0.15418)
-----		
Observations	753	753
R2	0.26422	0.26422
Adjusted R2	0.25730	0.25730
Residual Std. Error (df = 745)	0.42713	0.42713
F Statistic (df = 7; 745)	38.21795***	38.21795***
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

### 3.5.2.2 Multicolinealidad aproximada

```
library(regclass)
```

```
Cargando paquete requerido: bestglm
```

```
Cargando paquete requerido: leaps
```

```
Cargando paquete requerido: VGAM
```

```
Cargando paquete requerido: stats4
```

```
Cargando paquete requerido: splines
```

```
Adjuntando el paquete: 'VGAM'
```

```
The following object is masked from 'package:AER':
```

```
tobit
```

The following object is masked from 'package:lmtest':

```
lrtest
```

The following object is masked from 'package:car':

```
logit
```

Cargando paquete requerido: rpart

Cargando paquete requerido: randomForest

randomForest 4.7-1.2

Type rfNews() to see new features/changes/bug fixes.

Important regclass change from 1.3:

All functions that had a . in the name now have an \_  
all.correlations -> all\_correlations, cor.demo -> cor\_demo, etc.

```
VIF(mroz.mpl)
```

nwifeinc	educ	exper	I(exper^2)	age	kidslt6	kidsge6
1.170686	1.166001	8.636160	8.770985	1.658272	1.270358	1.250370

```
mean(VIF(mroz.mpl))
```

```
[1] 3.417547
```

En promedio el factor de inflación de la varianza es menor que 10. Por lo tanto, no debe preocuparme la multicolinealidad aproximada

### Normalidad de los errores

La  $H_0$ : los errores siguen una distribución normal

```
library(tseries)
jarque.bera.test(mroz.mpl$residuals)
```

Jarque Bera Test

```
data: mroz.mpl$residuals
X-squared = 36.741, df = 2, p-value = 1.051e-08
```

## 3.6 Modelos Logit y Probit para la respuesta binaria

El MPL es un modelo simple, que tiene varias desventajas. Las dos más importantes, como vimos en los ejemplos anteriores, son que las probabilidades ajustadas pueden ser menores que cero o mayores que uno y el efecto parcial de cualquier variable explicativa es constante. Estas limitaciones del MPL se superan con **modelos de respuesta binaria** más sofisticados.

En un modelo de respuesta binaria, el interés principal yace en la **probabilidad de respuesta**

$$P(y = 1|\mathbf{x}) = P(y = 1|x_1, x_2, \dots, x_k)[6]$$

Donde:  $\mathbf{x}$  denota el conjunto total de variables explicativas. Por ejemplo,  $\mathbf{x}$  podría contener varias características individuales como la educación, edad, estado civil, etc., que afecta, por ejemplo, al estado del empleo, incluye una variable de binaria para la participación en reciente programa de empleo

### 3.6.1 Especificación del modelo logit y probit

En el MPL, se suponía que la probabilidad de respuesta es lineal al conjunto de parámetros,  $\beta_j$ . Para evitar las limitaciones del MPL, considere una clase de modelos de respuesta binaria de la forma:

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k) = G(\beta_0 + \mathbf{x})[7]$$

donde  $G(\cdot)$  es una función que asume valores estrictamente entre 0 y 1:  $0 < G(\cdot) < 1$  para todos los números reales  $z$ . Esto asegura que las probabilidades de respuesta estimada sean estrictamente entre cero y uno. Note que:  $\mathbf{x} = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$

Se han sugerido varias funciones no lineales para la función  $G(\cdot)$  a fin de asegurar que las probabilidades estén entre cero y uno. Las dos funciones que estudiaremos en esta clase, se usan en la mayoría de aplicaciones (junto con el MPL). En el **Modelo Logit**,  $G(\cdot)$  es la función logística:

$$G(z) = \frac{\exp(z)}{[1 + \exp(z)]} = \frac{e^z}{[1 + e^z]} = \Lambda(z)[8]$$

Que está entre cero y uno para todos los números reales  $z$ . Esta es la función de distribución acumulada (fda) para una variable aleatoria logística estándar.

En el **Modelo Probit**,  $G(\cdot)$  es la función de distribución acumulada normal estándar, que viene dada de la siguiente forma:

$$G(z) = \Phi(z) \equiv \int_{-\infty}^z \phi(v)dv[9]$$

Donde:

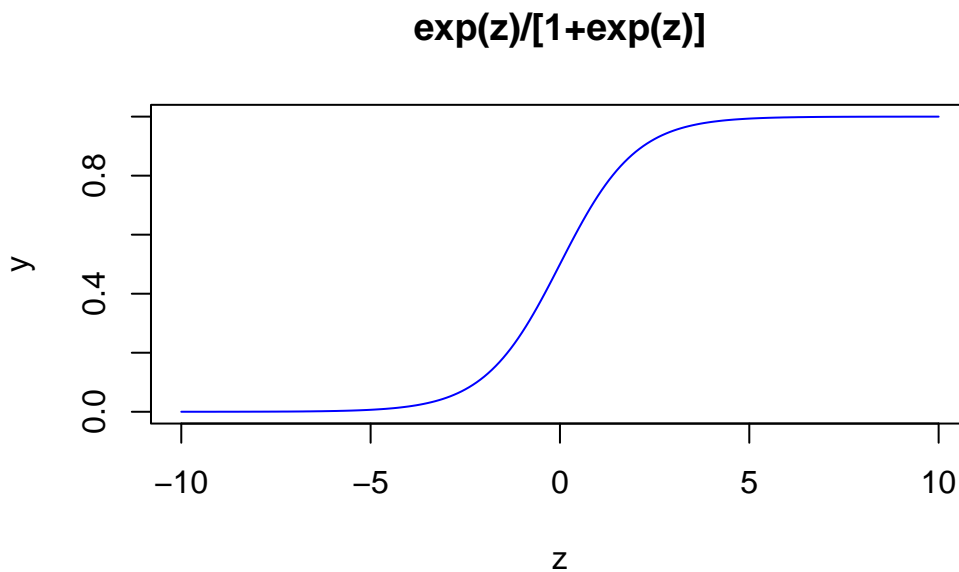
- $\phi(z)$  es la función de densidad normal estándar

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}[10]$$

Esta elección de  $G(\cdot)$  asegura que  $0 < P(y = 1|\mathbf{x}) < 1$ , para todos los valores de los parámetros y las  $x_j$

Las funciones  $G(\cdot)$  de **Logit** y **Probit** son crecientes. Cada una aumenta con más rapidez en  $z = 0$ ,  $G(z) \rightarrow 0$  a medida que  $z \rightarrow -\infty$  y,  $G(z) \rightarrow 1$  a medida que  $z \rightarrow \infty$

```
z<-seq(-10,10,0.1)
y<-exp(z)/(1+exp(z))
plot(y~z,
     main = "exp(z)/[1+exp(z)]",
     type = "l", col="blue")
```



Los modelos Logit y Probit pueden derivarse a partir de un **modelo de variable latente** subyacente.

Sea  $y^*$  una variable inobservable, o *latente* determinada por:

$$y^* = \beta_0 + \mathbf{x} + e, y = 1[y^* > 0] \quad [11]$$

Aquí se introduce la notación  $1[\cdot]$  para definir un resultado binario. La función  $1[\cdot]$  es la **función indicador**, que asume valor de uno si el evento dentro de los corchetes es verdadero y cero si es falso. Entonces tenemos:

$$\begin{aligned} y &= 1[y^* > 0] \quad [12] \\ y &= 0[y^* \leq 0] \end{aligned}$$

Se supone que  $e$  es independiente de  $\mathbf{x}$  y que  $e$  tiene una distribución logística estándar o normal estándar. En cualquier caso,  $e$  se distribuye simétricamente en torno a cero, lo que significa que  $1 - G(-Z) = G(z)$  para todos los números reales de  $z$ .

Los economistas tienden a favorecer el supuesto de normalidad para  $e$ , lo cual es la razón por la que en Econometría el **modelo Probit** es más popular que el **logit**. Además, varios problemas de especificación, que se tratarán después, se analizan fácilmente mediante Probit debido a las propiedades de la distribución normal.

Dado estos supuestos podemos calcular la probabilidad de respuesta para  $y$ :

$$P(y = 1|\mathbf{x}) = P(y^* > 0|\mathbf{x}) = P(e > -(\beta_0 + \mathbf{x})|\mathbf{x}) = 1 - G[-(\beta_0 + \mathbf{x})] = G(\beta_0 + \mathbf{x}) \quad [13]$$

uno de los objetivos de los modelos de respuesta binaria, es explicar los efectos de las  $x_j$  sobre la probabilidad de respuesta  $P(y = 1|\mathbf{x})$ . Cuidado, la fórmula de la variable latente tiende a dar la impresión de que lo que principalmente interesa son los efectos de cada  $x_j$  sobre  $y^*$ . Hay que aclarar que en los modelos Logit y Probit la dirección de efectos de  $x_j$  sobre  $E(y|\mathbf{x}) = P(y = 1|\mathbf{x}) = G(\beta_0 + \mathbf{x})$

Aclarando que:

$$E(y^*|\mathbf{x}) = \beta_0 + \mathbf{x} \quad [14]$$

Como la variable latente pocas veces tiene una unidad de medición definida, las magnitudes de cada  $\beta_j$  no son, útiles por sí mismas, a diferencia de las magnitudes calculadas por el **MPL**. Entonces para la mayoría de los propósitos, se requiere estimar el efecto de  $x_j$  sobre la probabilidad de éxito  $P(y = 1|\mathbf{x})$ , esto se complica por la naturaleza no lineal de  $G(\cdot)$ . Esto nos lleva a definir tres casos de efectos parciales:

### 3.6.2 Variables aproximadamente continuas:

Para hallar el efecto parcial de las variables aproximadamente continuas sobre la probabilidad de respuesta, se recurre al cálculo. Si  $x_j$  es una variable aproximadamente continua, su efecto parcial sobre  $p(x) = P(y = 1|\mathbf{x})$  se obtiene de la siguiente derivada parcial:

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = g(\beta_0 + \mathbf{x})\beta_j[15]$$

Donde:

$$g(z) \equiv \frac{dG}{dz}(z)[16]$$

Debido a que  $G$  es la fda de una variable aleatoria continua,  $g$  es la función de densidad de probabilidad.

En los casos de logit y probit,  $G(\cdot)$  es una fda estrictamente creciente y, por lo tanto,  $g(z) > 0 \forall z$ . Por lo tanto, el efecto parcial de  $x_j$  sobre  $p(\mathbf{x})$  depende de  $\mathbf{x}$  a través de la cantidad positiva  $g(\beta_0 + \mathbf{x})$ , lo que significa que el efecto parcial siempre tiene el mismo signo que  $\beta_j$ .

La ecuación de la derivada parcial muestra que los efectos *relativos* del cualquiera las variables explicativas continuas no depende de  $\mathbf{x}$ , la razón de los efectos parciales de  $x_j$  y  $x_h$  es  $\frac{\beta_j}{\beta_h}$ . El caso típico de que  $g$  sea una densidad simétrica en torno a cero, con una única moda en cero, el mayor efecto ocurre cuando  $\beta_0 + \mathbf{x} = 0$ . Por ejemplo:

#### 3.6.2.1 En el caso de Probit

$$g(z) = \phi(z) = \frac{e^{-z/2}}{\sqrt{2\pi}}$$
$$g(0) = \frac{e^{-0/2}}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \approx 0.40$$

```
1/sqrt(2*pi)
```

[1] 0.3989423



### 3.6.2.2 En el caso logit

$$g(z) = \frac{e^z}{[1 + e^z]^2}$$

Evaluable cuando  $z = 0$

$$g(0) = \frac{e^0}{[1 + e^0]^2}$$

$$g(0) = \frac{1}{[1 + 1]^2} = \frac{1}{4} = 0.25$$

### 3.6.3 Cuando la variable explicativa es binaria

Entonces el efecto parcial de cambiar  $x_1$  de cero a uno, manteniendo constante todas las demás variables, es así:

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \dots + \beta_k x_k) [17]$$

De nuevo, esto depende de todos los valores de las otras  $x_j$ . Por ejemplo, si  $y$  es un indicador de empleo y  $x_1$  es una variable binaria que indica la participación en un programa de capacitación laboral, entonces es el cambio en la probabilidad de empleo debido a este programa de capacitación; esto depende de las demás características que afectan la posibilidad de obtener el empleo, como la educación y la experiencia. Observe que saber el signo del  $\beta_1$  es suficiente para determinar si el programa tuvo un efecto positivo o negativo. Pero para hallar la **magnitud** del efecto, se tiene que estimar la cantidad usando la anterior ecuación [17].

### 3.6.4 Cuando la variable explicativa es discreta

Por ejemplo, el número de hijos. Si  $x_k$  denota esta variable, el efecto sobre la probabilidad de que  $x_k$  cambien de  $c_k$  a  $c_k + 1$  es simplemente:

$$G[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k (c_k + 1)] - G[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k c_k]$$

### 3.6.5 Estimación de máxima verosimilitud de los modelos Logit y Probit

Para los MPL se uso mínimos cuadrados ordinarios (MCO) o, si existe heterocedasticidad, mínimos cuadrados ponderados (MCP). Ahora bien, debido a naturaleza no lineal  $E(y|\mathbf{x})$ , MCO y MCP no son aplicables, por esta razón se usa la **estimación de máxima verosimilitud (EMV)**. Para estimar los modelos de variables dependientes limitadas, los métodos de máxima verosimilitud son indispensables. Como la EMV está basada en la distribución de  $y$  dada  $\mathbf{x}$ , la heterocedasticidad en  $Var(y|\mathbf{x})$  automáticamente se toma en cuenta.

Suponiendo que se tiene una muestra aleatoria  $n$ . Para obtener el estimador de máxima verosimilitud, condicional sobre las variables explicativas, es necesario la densidad de  $y_i$  dada  $\mathbf{x}_i$ . Esto se escribe como:

$$f(y|\mathbf{x}_i) = [G(\mathbf{x}_i)]^y [1 - G(\mathbf{x}_i)]^{1-y}, y = 0, 1 [17]$$

Para simplificar, se absorbe el intercepto en el vector  $\mathbf{x}_i$ . La **función log-verosimilitud** para cada observación  $i$  es una función de los parámetros y los datos  $(\mathbf{x}_i; y_i)$ , aplicando el logaritmo a la anterior ecuación tenemos:

$$l_i(\beta) = y_i \log[G(\mathbf{x}_i)] + (1 - y_i) \log[1 - G(\mathbf{x}_i)] [18]$$

Como  $G(\cdot)$  está estrictamente definida entre cero y uno para logit y probit,  $l_i(\beta)$  está bien definida para todos los valores  $\beta$

La log-verosimilitud para un tamaño de muestra  $n$  se obtiene al sumar todas las observación de la ecuación anterior:

$$\mathcal{L}_i(\beta) = \sum_{i=1}^n l_i(\beta) [19]$$

La EMV de  $\beta$ , denotada como  $\hat{\beta}$ , maximiza esta log-verosimilitud. Si  $G(\cdot)$  es la fda logit estándar, entonces  $\hat{\beta}$  será el estimador Logit; si  $G(\cdot)$  es la fda normal estándar, entonces  $\hat{\beta}$  será el estimador Probit.

### 3.6.6 Ejemplos de aplicación

Continuaremos con los ejemplos usados en el MPL, como son: la Participación en la fuerza laboral de las mujeres casadas, un modelo de probabilidad para arrestos y, sumaremos el ejemplo de la denegación de una hipoteca (Stock and Watson 2012)

### 3.6.6.1 Logit para los datos HMDA

```
library(AER)
data(HMDA)
```

```
HMDA |>
  str()
```

```
'data.frame':  2380 obs. of  14 variables:
 $ deny      : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ pirat     : num  0.221 0.265 0.372 0.32 0.36 ...
 $ hirat     : num  0.221 0.265 0.248 0.25 0.35 ...
 $ lvrat     : num  0.8 0.922 0.92 0.86 0.6 ...
 $ chist     : Factor w/ 6 levels "1","2","3","4",...: 5 2 1 1 1 1 1 2 2 2 ...
 $ mhist     : Factor w/ 4 levels "1","2","3","4": 2 2 2 2 1 1 2 2 2 1 ...
 $ phist     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ unemp     : num  3.9 3.2 3.2 4.3 3.2 ...
 $ selfemp   : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ insurance: Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 2 1 ...
 $ condomin  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 2 1 1 1 ...
 $ afam      : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ single    : Factor w/ 2 levels "no","yes": 1 2 1 1 1 1 2 1 1 2 ...
 $ hschool   : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
```

### 3.6.6.2 La participación den la fuerza laboral de las mujeres casadas

#### 3.6.6.2.1 Modelo Probit estimado con fda normal estándar

Para estimar modelos de variable dependiente limitada se usa el comando *glm()*

```
data("mroz", package = "wooldridge")

MPL.mroz <-lm(inlf~
              nwifeinc+
              educ+
              exper+
              expersq+
              age+
              kidslt6+
              kidsge6,
              data = mroz)
```

```

mroz.probit <- glm(inlf~
                  nwifeinc+
                  educ+
                  exper+
                  expersq+
                  age+
                  kidslt6+
                  kidsge6,
                  data = mroz,
                  family = binomial(link = "probit"))

library(stargazer)
stargazer(MPL.mroz, mroz.probit, type = "text")

```

=====		
	Dependent variable:	
	-----	
	inlf	
	OLS	probit
	(1)	(2)
-----		
nwifeinc	-0.003** (0.001)	-0.012** (0.005)
educ	0.038*** (0.007)	0.131*** (0.025)
exper	0.039*** (0.006)	0.123*** (0.019)
expersq	-0.001*** (0.0002)	-0.002*** (0.001)
age	-0.016*** (0.002)	-0.053*** (0.008)
kidslt6	-0.262*** (0.034)	-0.868*** (0.118)

kidsge6	0.013 (0.013)	0.036 (0.044)
Constant	0.586*** (0.154)	0.270 (0.508)
-----		
Observations	753	753
R2	0.264	
Adjusted R2	0.257	
Log Likelihood		-401.302
Akaike Inf. Crit.		818.604
Residual Std. Error	0.427 (df = 745)	
F Statistic	38.218*** (df = 7; 745)	
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

### 3.6.6.2.1.1 ¿Cómo funciona la mecánica del modelo Probit?

```
mroz.probit.simple <-glm(inlf~
                        kidslt6,
                        data = mroz,
                        family = binomial(link = "probit"))

mroz.probit.simple1 <-glm(inlf~
                        nwifeinc,
                        data = mroz,
                        family = binomial(link = "probit"))

stargazer::stargazer(mroz.probit.simple1, type = "text")
```

```
=====
                        Dependent variable:
                        -----
                        inlf
                        -----
nwifeinc                -0.013***
                        (0.004)

Constant                0.432***
```

(0.094)

```
-----  
Observations          753  
Log Likelihood        -509.662  
Akaike Inf. Crit.     1,023.324  
=====
```

Note:                \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```
mean(mroz$kidslt6)
```

```
[1] 0.2377158
```

```
# Efectos en cambios puntuales  
  
prediccion<-predict(mroz.probit.simple,  
                    newdata=data.frame("kidslt6"=c(1,2)),  
                    type = "response")  
  
diff(prediccion)*100
```

```
                    2  
-18.7173
```

### 3.6.6.3 Presentación del modelo estimado

#### Modelo probit simple

$$P(\widehat{infl = 1} | kidslt6) = \Phi(0.299 - 0.539kidslt6)$$

```
pnorm(coef(mroz.probit.simple)[1]+coef(mroz.probit.simple)[2]*2)-pnorm(coef(mroz.probit.simp
```

```
(Intercept)  
-0.187173
```

#### Ecuación estimada de probit

```
coef(mroz.probit)
```

(Intercept)	nwifeinc	educ	exper	expersq	age
0.270073573	-0.012023637	0.130903969	0.123347168	-0.001887067	-0.052852442
kidslt6	kidsge6				
-0.868324680	0.036005611				

$$P(inlf = 1|nwifeinc, ..., kidsge6) = \Phi(0.27 - 0.012nwifeinc + 0.131educ + 0.12exper - 0.0019exper^2 - 0.053age - 0.87kidslt6 + 0.036kidsge6)$$

¿Cuál es la probabilidad de salir de que María salga a trabajar, dado que tiene su esposo un ingreso mensual 300USD, tiene 4 años de educación, nunca ha trabajado, tiene 29 años y un niño de 3 años?

```
coef(mroz.probit)
```

(Intercept)	nwifeinc	educ	exper	expersq	age
0.270073573	-0.012023637	0.130903969	0.123347168	-0.001887067	-0.052852442
kidslt6	kidsge6				
-0.868324680	0.036005611				

```
# Pregunta inicial
prediccion<-predict(mroz.probit,
                    newdata=data.frame("nwifeinc"=(300*12)/1000,
                                         "educ"=4,
                                         "exper"=0,
                                         "expersq"=0,
                                         "age"=29,
                                         "kidslt6"=1,
                                         "kidsge6"=0),
                    type = "response")

predict(mroz.probit,
        newdata=data.frame("nwifeinc"=(300*12)/1000,
                             "educ"=4,
                             "exper"=0,
                             "expersq"=0,
                             "age"=29+3,
```

```

                                "kidslt6"=1,
                                "kidsge6"=1),
                                type = "response")

```

1  
0.03809838

```

#Diferencia
cambio <- predict(mroz.probit,
                  newdata=data.frame("nwifeinc"=(300*12)/1000,
                                      "educ"=4,
                                      "exper"=0,
                                      "expersq"=0,
                                      "age"=c(29,29+3),
                                      "kidslt6"=1,
                                      "kidsge6"=c(0,1)),
                  type = "response")

diff(cambio)*100

```

2  
-1.130756

```

# En el MPL

cambio.MPL <- predict(MPL.mroz,
                     newdata=data.frame("nwifeinc"=(300*12)/1000,
                                         "educ"=4,
                                         "exper"=0,
                                         "expersq"=0,
                                         "age"=c(29,29+3),
                                         "kidslt6"=1,
                                         "kidsge6"=c(0,1)),
                     type = "response")

diff(cambio.MPL)*100

```

2  
-3.526018



#### 3.6.6.4 Modelo Logit estimado con FDA logística estándar

```
mroz.logit <- glm(inlf~
  nwifeinc+
  educ+
  exper+
  expersq+
  age+
  kidslt6+
  kidsge6,
  data = mroz,
  family = binomial(link = "logit"))

stargazer(mroz.logit, type = "text")
```

```
=====
                        Dependent variable:
                        -----
                                inlf
                        -----
nwifeinc                -0.021**
                        (0.008)

educ                    0.221***
                        (0.043)

exper                   0.206***
                        (0.032)

expersq                 -0.003***
                        (0.001)

age                     -0.088***
                        (0.015)

kidslt6                 -1.443***
                        (0.204)

kidsge6                  0.060
                        (0.075)
```

Constant	0.425 (0.860)
----------	------------------

```
-----
Observations          753
Log Likelihood        -401.765
Akaike Inf. Crit.     819.530
=====
```

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 3.6.6.4.0.1 ¿Cómo funciona la mecánica del modelo Logit?

```
mroz.logit.simple <-glm(inlf~
                        kidslt6,
                        data = mroz,
                        family = binomial(link = "logit"))

stargazer(mroz.probit.simple, type = "text")
```

```
=====
                        Dependent variable:
                        -----
                        inlf
                        -----
kidslt6                -0.539***
                        (0.094)

Constant              0.299***
                        (0.051)

-----
Observations          753
Log Likelihood        -497.367
Akaike Inf. Crit.     998.734
=====
Note: *p<0.1; **p<0.05; ***p<0.01
```

```
# Efectos en cambios puntuales
```

```
prediccion<-predict(mroz.logit.simple,
                    newdata=data.frame("kidslt6"=c(1,2)),
                    type = "response")
diff(prediccion)*100
```

```
      2
-18.29326
```

```
lambda <- function(z) 1/(1+exp(-z))
```

```
lambda(coef(mroz.logit.simple)[1]+coef(mroz.logit.simple)[2]*2)*100-lambda(coef(mroz.logit.s
```

```
(Intercept)
-18.29326
```

```
# Verificando que las probabilidades ajustadas se encuentren entre 0 y 1
```

```
summary(mroz.probit$fitted.values)
```

```
      Min.   1st Qu.   Median     Mean  3rd Qu.     Max.
0.002475 0.370959 0.609546 0.570109 0.794345 0.979904
```

```
summary(mroz.logit$fitted.values)
```

```
      Min.   1st Qu.   Median     Mean  3rd Qu.     Max.
0.008672 0.366410 0.610925 0.568393 0.796721 0.968541
```

## Presentación del modelo estimado

$$P(inlf = 1|nwifeinc, ..., kidsge6) = \Lambda(0.425 - 0.021nwifeinc + \dots + 0.060kidsge6)$$

### 3.6.6.5 Comparación de los modelos MPL, Logit y Probit

Para esta comparación se va usar los errores heterocedástico robustos.

```

library(sandwich)

eher<-list(sqrt(diag(vcovHC(MPL.mroz, type = "HC1"))),
           sqrt(diag(vcovHC(mroz.logit, type = "HC1"))),
           sqrt(diag(vcovHC(mroz.probit, type = "HC1"))))

stargazer(MPL.mroz, mroz.logit, mroz.probit,
           se = eher,
           digits = 3,
           type = "text",
           title = "Tabla 2: Estimaciones MPL, logit y probit de la participación en la fuerza",
           df=F,
           header = F)

```

Tabla 2: Estimaciones MPL, logit y probit de la participación en la fuerza laboral

Dependent variable:			
	inlf		
	OLS	logistic	probit
	(1)	(2)	(3)
nwifeinc	-0.003** (0.002)	-0.021** (0.009)	-0.012** (0.006)
educ	0.038*** (0.007)	0.221*** (0.045)	0.131*** (0.026)
exper	0.039*** (0.006)	0.206*** (0.032)	0.123*** (0.019)
expersq	-0.001*** (0.0002)	-0.003*** (0.001)	-0.002*** (0.001)
age	-0.016*** (0.002)	-0.088*** (0.015)	-0.053*** (0.008)
kidslt6	-0.262*** (0.032)	-1.443*** (0.204)	-0.868*** (0.117)
kidsge6	0.013	0.060	0.036

	(0.014)	(0.080)	(0.047)
Constant	0.586*** (0.152)	0.425 (0.864)	0.270 (0.507)
-----			
Observations	753	753	753
R2	0.264		
Adjusted R2	0.257		
Log Likelihood		-401.765	-401.302
Akaike Inf. Crit.		819.530	818.604
Residual Std. Error	0.427		
F Statistic	38.218***		
=====			
Note:	*p<0.1; **p<0.05; ***p<0.01		

Como podemos ver en la tabla 2 los signos y la significancia es la misma para todas las variables en los tres modelos. Por ejemplo, la variable *educ* y *exper* son estadísticamente significativas en los tres modelos y ambas tienen un signo positivo respecto a la probabilidad de la participación en la fuerza laboral de las mujeres. En un primer momento no es posible comparar las estimaciones logit y probit con las del MPL. Para hacerlas comparables se debe usar el **efecto parcial promedio (EPP)**. Wooldridge (2010, 585) sugiere factores escalares que se deben pre-multiplicar por los coeficientes de logit y probit para hacerlos comparables con el MPL. Para Probit es 0.301 y para logit es 0.179.

#### Usando factores escalares logit y probit para comparar con coeficientes MPL

El ejemplo de la variable *educ*. Si multiplico el coeficiente de *educ* en logit por su factor se obtiene:  $0.179(0.221) \approx 0.040$  y coeficiente probit *educ* es de alrededor de  $0.301(0.131) \approx 0.039$ . Como se puede observar, ambos coeficientes son muy cercanos a la estimación de MPL que es de 0.038. También la variable discreta *kidslt6*, los coeficientes escalados logit y probit son similares al coeficiente del MPL de  $-0.262$ . Estos son  $0.179(-1.443) \approx -0.258$  (logit) y,  $0.301(-0.868) \approx -0.261$  (probit)

La mayor diferencia entre el modelo MPL y los modelos logit y probit es que el MPL supone efectos *constantes* para *educ*, *exper*, *kidslt6*, etc., mientras que los modelos logit y probit implican magnitudes decrecientes de los efectos parciales

#### 3.6.6.6 Curva decreciente

En esta sección vamos a observar como los modelos no lineales logit y probit muestran que no es lo mismo tener niño pequeño, dos o tres, etc, para reducir la probabilidad de salir a trabajar

```

mpl.simple <- lm(inlf~
                kidslt6,
                mroz)
plot(x = mroz$kidslt6,
     y= mroz$inlf,
     main = "Modelo probit para los determinates del trabajo femenino",
     xlab = "Niños menore a seis años",
     ylab = "Infl, si una mujer casada sale a trabajar por un salario",
     pch=20,
     ylim = c(-0.4, 1.4),
     xlim = c(-0.2,8))
grid()
# Añadir las lineas horizontales y el texto
abline(h=1, lty=2, col="darkred")
abline(h=0, lty=2, col="darkred")
text(2.5, 0.9, cex = 0.8, "Sale a trabajar")
text(2.5, -0.1, cex = 0.8, "No sale a trabajar")

# añadiendo la linea de regresión probit
x <- seq(0,7,1)
y <- predict(mroz.probit.simple,
             list(kidslt6=x),
             type = "response")
lines(x,y,lwd=1.5, col="steelblue")

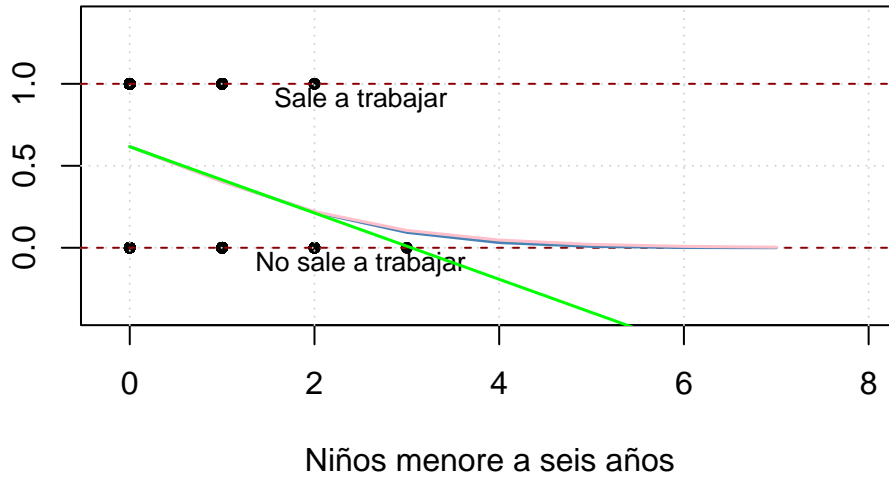
# añadiendo la linea de regresión logit
t <- predict(mroz.logit.simple,
             list(kidslt6=x),
             type = "response")
lines(x,t,lwd=1.5, col="pink")

# añadiendo la linea de regresión MPL
m <- predict(mpl.simple,
             list(kidslt6=x),
             type = "response")
lines(x,m,lwd=1.5, col="green")

```

fl, si una mujer casada sale a trabajar por un sa

### Modelo probit para los determinates del trabajo femenino



### 3.6.7 Interpretaciones de las estimaciones Logit y Probit

Las estimaciones de coeficientes, sus errores estándar y el valor de la función de log-verosimilitud se pueden obtener mediante todos los paquetes de software (R) que realicen logit y probit, y se deben reportar en cualquier aplicación. Los coeficientes dan los signos de los efectos parciales de cada  $x_j$  sobre la probabilidad de respuesta y la significancia estadística de  $x_j$  está determinada por si se puede rechazar  $H_0 : \beta_j = 0$  a un nivel de significancia ( $\alpha$ ).

Como vimos anteriormente para el MPL se puede calcular el **porcentaje predicho correctamente**

Existen varias medidas de bondad de ajuste como **pseudo R-cuadradas**. MacFadden (1974) sugiere la medida  $1 - \frac{\mathcal{L}_{nr}}{\mathcal{L}_o}$ , donde  $\mathcal{L}_{nr}$  es la función de log-verosimilitud para el modelo estimado y  $\mathcal{L}_o$  es la función de probabilidad de log en el modelo con sólo un intercepto. ¿Por qué esta medida es lógica? Recordar que las log-verosimilitud son negativas y, por tanto  $\frac{\mathcal{L}_{nr}}{\mathcal{L}_o} = \frac{|\mathcal{L}_{nr}|}{|\mathcal{L}_o|}$ . Además,  $|\mathcal{L}_{nr}| \leq |\mathcal{L}_o|$ . Si las covarianzas no tiene poder explicativo, entonces  $\frac{\mathcal{L}_{nr}}{\mathcal{L}_o} = 1$ , la **pseudo R-cuadrada** será igual a cero, como la R-cuadrada usual es cero en una regresión lineal cuando las covariadas no tienen poder explicativo.

Por lo general,  $|\mathcal{L}_{nr}| < |\mathcal{L}_o|$ , en cuyo caso  $1 - \frac{\mathcal{L}_{nr}}{\mathcal{L}_o} > 0$ . Supongamos que  $\mathcal{L}_{nr} \rightarrow 0$ , la pseudo-Rcuadrada tiene a uno. Pero en los modelos logit y probit no pueden llegar a cero  $\mathcal{L}_{nr}$  ya que eso requeriría que las probabilidades estimadas cuando  $y_i = 1$  fueran iguales a la unidad y que las probabilidades estimadas cuando  $y_i = 0$  fueran todas iguales a cero.

### 3.6.8 Cálculo de la pseudo- $R^2$ de MacFadden

Segun Stock y Watson (2011), las llamadas pseudo- $R^2$  se usan para medir la calidad del ajuste, estas medidas comparan el valor de la probabilidad máxima log-verosimilitud con todos los regresores, con la probabilidad de un modelo sin regresores (modelo nulo) **regresión en una constante**

Por ejemplo, considere una regresión Probit. El **pseudo- $R^2$**  esta dado por:

$$pseudo - R^2 = 1 - \frac{\ln(f_{full}^{max})}{\ln(f_{null}^{max})} [20]$$

Donde:  $f_j^{max} \in [0, 1]$  denota la probabilidad máxima para el modelo  $j$

El razonamiento detrás de esto, es que, la probabilidad maximizada aumenta a medida que se agregan regresores adicionales al modelo, de manera similar a la disminución en  $SRC$  cuando se agregan regresores en un modelo de regresión lineal. Si el modelo completo tiene una probabilidad maximizada similar a la del modelo nulo, el modelo completo no mejora realmente sobre un modelo que usa solo la información en la variable dependiente, por lo que  $pseudo - R^2 \approx 0$ . Si el modelo completo se ajusta muy bien a los datos, la probabilidad maximizada debe estar cerca de 1, tal que  $\ln(f_{full}^{max}) \approx 0$  y  $pseudo - R^2 \approx 1$

En Rstudio para los modelos estimados con  $glm()$  podemos utilizar las entradas de desviación residual (**desviance**) y la desviación nula (**null.desviance**). Estos han sido calculados de la siguiente forma:

$$desviance = -2 \times [\ln(f_{saturated}^{max}) - \ln(f_{full}^{max})] \text{ null.desviance} = -2 \times [\ln(f_{saturated}^{max}) - \ln(f_{null}^{max})]$$

Donde:  $f_{saturated}^{max}$  es la probabilidad maximizada para un modelo que asume que cada observación tiene su propio parámetro (hay  $n + 1$  parámetros a estimar que conducen a un ajuste perfecto). Para los modelos con una variable dependiente binaria, se tiene que:

$$pseduo - R^2 = 1 - \frac{desviance}{null.desviance} = 1 - \frac{\ln(f_{full}^{max})}{\ln(f_{null}^{max})} [21]$$

**Cálculo del  $pseudo - R^2$  para los modelos Logit y Probit** del ejemplo, La participación en la fuerza laboral de las mujeres casadas

```
# Probit

pseudo.R2.P <- 1-(mroz.probit$deviance/mroz.probit$null.deviance)

pseudo.R2.P*100
```



## # Logit

Si usamos la interpretación usual del  $R^2$  de la regresión lineal, diremos que según los **pseudo-R<sup>2</sup>** de logit y probit, aproximadamente la variación de la probabilidad de la participación en la fuerza laboral de las mujeres casadas esta explicada por las variables regresoras en aproximadamente un 22%.

En cualquier caso, la bondad de ajuste suele ser menos importante que intentar obtener estimaciones convincentes de los efectos **ceteris paribus** de las variables explicativas.

### 3.6.9 Efecto parcial promedio y el efecto parcial en el promedio

Parte importante de estos modelos es estimar los efectos de las  $x_j$  sobre las probabilidades de respuesta,  $P(y = 1|\mathbf{x})$ . Si  $x_j$  es aproximadamente continua teníamos:

$$\Delta \hat{P}(y = 1|\mathbf{x}) \approx [g(\hat{\beta}_0 + \mathbf{x}\hat{\beta})\hat{\beta}_j]\Delta x_j[22]$$

Entonces, para pequeños cambios en  $x_j$ . Así que, para  $\Delta x_j = 1$  el cambio en la probabilidad de éxito es aproximadamente  $g(\hat{\beta}_0 + \mathbf{x}\hat{\beta})\hat{\beta}_j$ . En comparación con el MPL, el costo de usar modelos probit y logit es que los efectos parciales en la ecuación anterior son más difíciles de resumir debido a que el factor de escala  $g(\hat{\beta}_0 + \mathbf{x}\hat{\beta})$ , depende de  $\mathbf{x}$ . Una posibilidad es insertar valores interesante para las  $x_j$  (medias, medianas, mínimos, máximos, cuartiles, etc.) y, ver como cambia  $g(\hat{\beta}_0 + \mathbf{x}\hat{\beta})$ . Pero, a pesar de ser un proceso atractivo es tedioso y puede dar como resultado demasiada información aun si el número de variables explicativas es moderado.

Como resumen rápido para obtener magnitudes de efectos parciales, es útil tener un factor escalar único que se pueda multiplicar con cada  $\hat{\beta}_j$  (o al menos aquellos coeficiente de variables aproximadamente continuas). Un método que suele usarse en paquetes econométricos es reemplazar cada variable explicativas con su promedio muestral. En otras palabras, el factor de ajuste es:

$$g(\hat{\beta}_0 + \bar{\mathbf{x}}\hat{\beta}) = g(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \dots + \hat{\beta}_k\bar{x}_k)[23]$$

Donde:  $g(\cdot)$  es la densidad normal estándar ( $\phi$ ) para el caso probit y,  $g(z) = \frac{\exp(z)}{[1+\exp(z)]^2}$  para logit. Cuando a la ecuación anterior se multiplica por  $\hat{\beta}_j$  obtenemos el efecto de  $x_j$  para la persona promedio en la muestra. Por lo tanto, si multiplico el coeficiente  $\beta_j$  por la ecuación [23], se obtiene el **efecto parcial en el promedio (EPeP)**.

### 3.6.9.1 Ejemplo con los determinantes del trabajo femenino

```
data("mroz", package = "wooldridge")

Epep.probit <- glm(inlf~nwifeinc,
                  mroz,
                  family = binomial(link = "probit"))
stargazer::stargazer(Epep.probit, type = "text")
```

```
=====
                        Dependent variable:
-----
                        inlf
-----
nwifeinc                -0.013***
                        (0.004)

Constant                0.432***
                        (0.094)

-----
Observations              753
Log Likelihood            -509.662
Akaike Inf. Crit.        1,023.324
=====
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

Una vez que tengo el modelo, lo uso para ejemplificar el **EPeP**,

```
dnorm(coef(Epep.probit)[1]+coef(Epep.probit)[2]*mean(mroz$nwifeinc))
```

```
(Intercept)
0.3929979
```

```
phi <- function(z) (1/sqrt(2*pi))*exp(-z^2/2)

phi(coef(Epep.probit)[1]+coef(Epep.probit)[2]*mean(mroz$nwifeinc))
```

```
(Intercept)
0.3929979
```

```
# Efecto de aumentar el salario en una unidad son 1000 USD
```

```
dnorm(coef(Epep.probit)[1]+coef(Epep.probit)[2]*mean(mroz$nwifeinc))*coef(Epep.probit)[2]*1000
```

```
(Intercept)
-0.5052942
```

El mismo ejemplo para **logit**

```
Epep.logit <- glm(inlf~nwifeinc,
                  mroz,
                  family = binomial(link = "logit"))
stargazer::stargazer(Epep.logit, type = "text")
```

```
=====
                        Dependent variable:
                        -----
                                inlf
                        -----
nwifeinc                        -0.021***
                                (0.007)

Constant                        0.695***
                                (0.152)

                        -----
Observations                     753
Log Likelihood                  -509.654
Akaike Inf. Crit.               1,023.309
=====
Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
# fda logística estándar
```

```
lambda.minus <- function(z) exp(z)/(1+exp(z))^2
```

```
lambda.minus(coef(Epep.logit)[1]+coef(Epep.logit)[2]*mean(mroz$nwifeinc))*coef(Epep.logit)[2]
```

(Intercept)  
-50.91089

Existen dos problemas con el uso del **EPeP**. Primero, si algunas de las variables explicativas son discretas, sus promedios no representan a nadie en la muestra. Por ejemplo, si  $x_1 = \text{mujeres}$  y 47.5% de la muestra son mujeres ¿qué sentido tiene insertar  $\bar{x}_1 = 0.475$  para representar a la persona “promedio”? Segundo, si una variable explicativa continua aparece como función no lineal, por ejemplo, como un log-natural o cuadrática, no es claro si se quiere promediar la función no lineal o insertar el promedio en la función no lineal. Por ejemplo, ¿Se debe usar  $\log(\bar{\text{ventas}})$  o  $\log(\text{ventas})$  para representar el tamaño promedio de la empresa?. Los paquetes econométrico se quedan en el primero, el paquete está programado para calcular los promedios de los regresores incluidos en la estimación probit o logit.

Un método diferente para calcular un factor escalar elude la cuestión de qué valores a insertar para las variables explicativas. En lugar de ello, el segundo factor escalar resulta al promediar los efectos parciales individuales a través de la muestra, lo que genera en algunas veces llamado **efecto parcial promedio (EPP)**. Por ejemplo, para una variable aproximadamente continua el **EPP** es:

$$n^{-1} \sum_{i=1}^n [g(\hat{\beta}_0 + \mathbf{x}_i \hat{\beta}) \hat{\beta}_j] = n^{-1} \sum_{i=1}^n [g(\hat{\beta}_0 + \mathbf{x}_i \hat{\beta})] \hat{\beta}_j [24]$$

El término que se multiplica a  $\hat{\beta}_j$  actúa como un factor escalar:

$$n^{-1} \sum_{i=1}^n [g(\hat{\beta}_0 + \mathbf{x}_i \hat{\beta})] [25]$$

Los factores escalares que sirven para obtener el *EPP* y *EPeP* que fueron detallados anteriormente de la aproximación del cálculo, ninguna es lógica para variables explicativas discretas. Es su lugar, se debe estimar directamente el cambio de probabilidad. Para un cambio  $x_k$  de  $c_k$  a  $c_k + 1$ , es análogo al efecto parcial en el promedio:

$$G[\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_{k-1} \bar{x}_{k-1} + \hat{\beta}_k (c_k + 1)] - G[\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_{k-1} \bar{x}_{k-1} + \hat{\beta}_k c_k] [26]$$

El efecto parcial promedio es:

$$n^{-1} \sum_{i=1}^n (G[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k (c_k + 1)] - G[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k c_k]) [27]$$

La función anterior se puede interpretar de forma particular cuando  $x_k$  es binaria. Para cada unidad  $i$ , se estima la diferencia predicha en la probabilidad de que  $y_i = 1$  cuando  $x_k = 1$  y  $x_k = 0$ , de la siguiente forma:

$$n^{-1} \sum_{i=1}^n G[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k] - G[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{k-1} x_{ik-1}] [28]$$

Para finalizar la aplicación de MPL, Logit y Probit. Es importante tener un tipo de efecto marginal que sea interpretable para los modelos no lineales (logit y probit), estos se obtienen de la siguiente manera usando el ejemplo de:

### 3.6.9.1.1 Efecto parcial promedio ejemplo

```
# Para probit
```

```
mean(dnorm(coef(Epep.probit)[1]+coef(Epep.probit)[2]*mroz$nwifeinc))*coef(Epep.probit)[2]*10
```

```
nwifeinc
-0.4998448
```

```
# Para logit
```

```
mean(lambda.minus(coef(Epep.logit)[1]+coef(Epep.logit)[2]*mroz$nwifeinc))*coef(Epep.logit)[2]
```

```
nwifeinc
-0.5021655
```

[Participación en la fuerza laboral de las mujeres casadas]

```
library(mfx)
```

```
# Probando lo hecho a mano
```

```
probitmfx(inlf~
          nwifeinc,
          data = mroz)
```

Call:

```
probitmfx(formula = inlf ~ nwifeinc, data = mroz)
```

Marginal Effects:

	dF/dx	Std. Err.	z	P> z
nwifeinc	-0.0050529	0.0015912	-3.1755	0.001496 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
# Efectos marginales Probit
marginales.probit<-probitmfx(inlf~
  nwifeinc+
  educ+
  exper+
  expersq+
  age+
  kidslt6+
  kidsge6,
  data = mroz)
```

```
marginales.probit$mfkest[1:7]*100
```

```
[1] -0.46961881  5.11284287  4.81768957 -0.07370502 -2.06430891
[6] -33.91499645  1.40630594
```

```
# Efectos marginales Logit
marginales.logit<-logitmfx(inlf~
  nwifeinc+
  educ+
  exper+
  expersq+
  age+
  kidslt6+
  kidsge6,
  data = mroz)
```

```
marginales.logit$mfkest[1:7]*100
```

```
[1] -0.51900534  5.37773087  5.00569282 -0.07669166 -2.14030205
[6] -35.09498193  1.46162143
```

```
# Comparación entre logit y probit
```

```
variables<-c("nwifeinc", "educ", "exper", "expersq", "age", "kidslt6", "kidsge6")
```

```
comparacion <-data.frame(variables,marginales.logit$mfcest[1:7]*100, marginales.probit$mfcest[1:7]*100)
names(comparacion)<-c("Betas", "Logit", "Probit")
```

```
comparacion
```

	Betas	Logit	Probit
1 nwifeinc	-0.51900534	-0.46961881	
2 educ	5.37773087	5.11284287	
3 exper	5.00569282	4.81768957	
4 expersq	-0.07669166	-0.07370502	
5 age	-2.14030205	-2.06430891	
6 kidslt6	-35.09498193	-33.91499645	
7 kidsge6	1.46162143	1.40630594	

```
100000/12
```

```
[1] 8333.333
```

Una vez, establecidos los valores de los betas interpretables, podemos pasar a mirar la exactitud del estimaciones de los dos modelos no lineales.

### 3.6.10 Porcentaje predicho correctamente y la matriz de confusión

En lugar de solo calcular el PPC, se presentará la matriz de confusión que permite mostrar cuantas veces el modelo predijo correctamente los valores de  $y$

```
library(vcd)

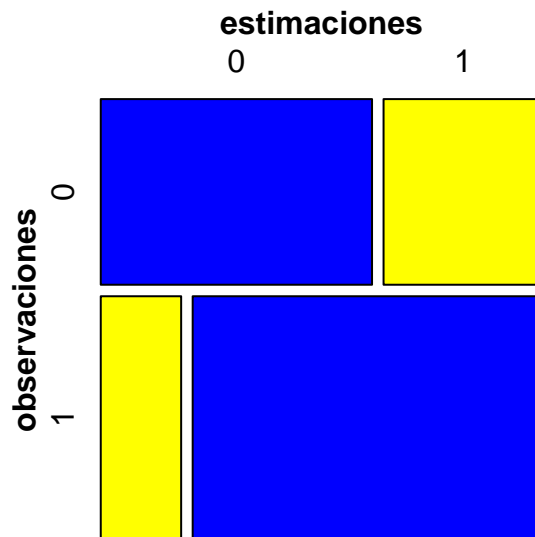
#Probit
predi.probit <- ifelse(mroz.probit$fitted.values>0.5,1,0)

MCProbit <- table(mroz$inlf, predi.probit,
                  dnn = c("observaciones", "estimaciones"))
PPC.probit <- data.frame(mroz$inlf, predi.probit, ifelse(mroz$inlf==predi.probit,1,0))
names(PPC.probit)=c("inlf", "Valores ajustados", "PPC")

prop.table(table(PPC.probit$PPC))[2]*100
```

1  
73.43958

```
mosaic(MCProbit,  
  shade = T,  
  colorize = T,  
  gp = gpar(fill = matrix(c("blue", "yellow", "yellow", "blue"),  
    2,2)))
```



```
library(gmodels)  
  
CrossTable(mroz$inlf, predi.probit,  
  digits = 2,  
  format = "SPSS",  
  prop.c = F,  
  prop.chisq = F,  
  prop.t = F,  
  dnn=c("Observado", "estimado"))
```

Cell Contents

|-----|



	Count
	Row Percent
-----	-----

Total Observations in Table: 753

	estimado		
Observado	0	1	Row Total
-----	-----	-----	-----
0	205	120	325
	63.08%	36.92%	43.16%
-----	-----	-----	-----
1	80	348	428
	18.69%	81.31%	56.84%
-----	-----	-----	-----
Column Total	285	468	753
-----	-----	-----	-----

**1. Sensitividad:** % de positivos (1) que sob clasificados como positivos (1). para el modelo probit seria ( $\frac{347}{428} = 81.07\%$ )

**2. Especificidad:** % negativos(0) que son clasificado como negativops(0). . en nuestro ejemplo\_: ( $\frac{207}{325} = 63.98\%$ )

**Falsos positivos:** % de negativos (o) clasificados como positivos (1). En nuestro ejemplo: ( $\frac{120}{325} = 36.92\%$ )

**Falsos Negativos:** % de positivos clasificados (1) como negativos (0). en nuestro ejemplo. ( $\frac{80}{428} = 18.69\%$ )

```
predi.logit <- ifelse(mroz.logit$fitted.values>=0.5,1,0)

m.confu.logit<- table(mroz$inlf, predi.logit,
                      dnn = c("observaciones", "estimaciones"))
PPC.logit <- data.frame(mroz$inlf, predi.logit, ifelse(mroz$inlf==predi.logit,1,0))
names(PPC.logit)=c("inlf", "Valores ajustados", "PPC")

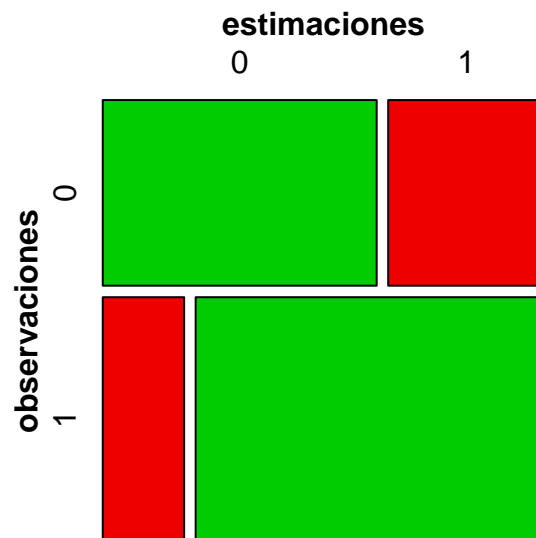
prop.table(table(PPC.logit$PPC))[2]*100
```

1  
73.57238

```

mosaic(m.confu.logit, shade = T, colorize = T,
       gp = gpar(fill = matrix(c("green3", "red2", "red2", "green3"),
                                2,2)))

```



```

CrossTable(mroz$inlf, predi.logit,
           digits = 2,
           format = "SPSS",
           prop.c = F,
           prop.chisq = F,
           prop.t = F,
           dnn=c("Observado", "estimado"))

```

Cell Contents	
	Count
	Row Percent

Total Observations in Table: 753

| estimado

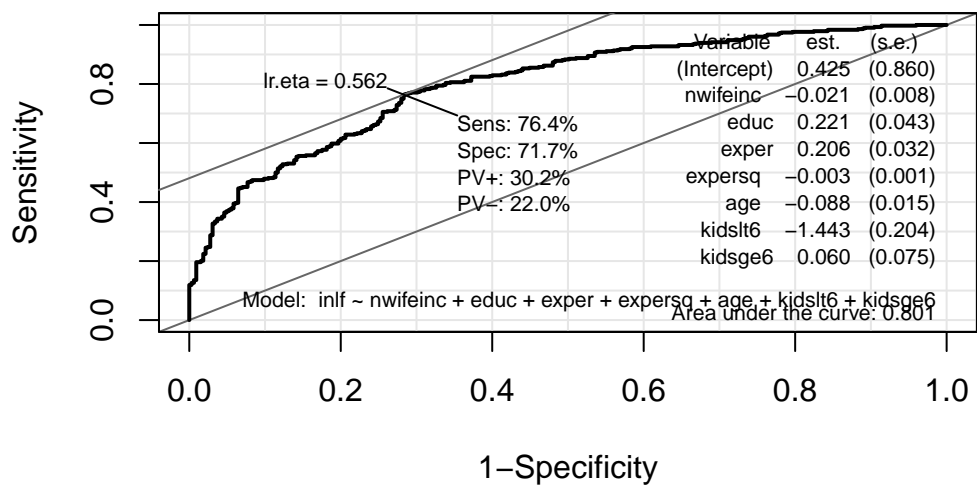
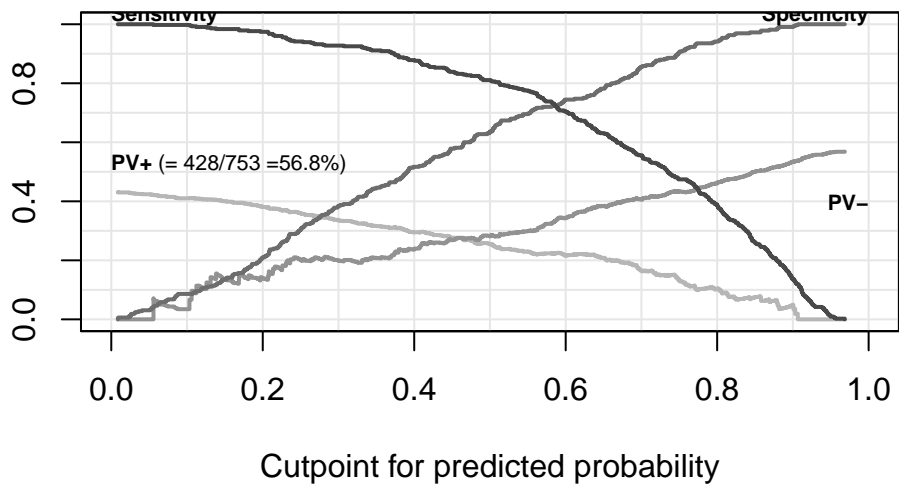
Observado	0	1	Row Total
0	207 63.69%	118 36.31%	325 43.16%
1	81 18.93%	347 81.07%	428 56.84%
Column Total	288	465	753

Los modelos logit y probit son capaces de clasificar correctamente el 73.5 de las observaciones cuando se emplean los datos de trabajo femenino.

### 3.6.11 Capacidad discriminante del modelo

```
library("Epi")

ROC(form=inlf~nwifeinc+
     educ+
     exper+
     expersq+
     age+
     kidslt6+
     kidsge6,
     data = mroz)
```



**Sensibilidad:** la probabilidad de que el modelo prediga un resultado positivo (1) para una observación cuando en realidad el resultado es positivo (1)

**Especificidad:** La probabilidad de que el modelo prediga un resultado negativo para una observación cuando en realidad el resultado es negativo.

## 4 Ejercicio 17.2 del libro de Wooldridge

Sea  $grad$  una variable binaria para si un atleta colegial en una universidad grande se graduará en cinco años. Sean  $hsGPA$  y  $SAT$  el promedio de calificaciones de bachillerato y las puntuaciones del  $SAT$  de admisión a la universidad, respectivamente. Sea  $study$  el número de horas por semana que pasa un estudiante en un aula de estudio. Suponga que, usando los datos sobre 420 atletas colegiales se obtiene el siguiente modelo logit:

$$\widehat{P}(grad = 1 | hsGPA, SAT, study) = \Lambda(-1.17 + 0.24hsGPA + 0.00058SAT + 0.073study)$$

$$\Lambda = \frac{\exp(z)}{[1 + \exp(z)]}$$

Si mantiene  $hsGPA = 3.0$  y el  $SAT = 1200$ , calcule la diferencia estimada en la probabilidad de graduación para alguien que pasa 10 horas a la semana en el aula de estudio y alguien que pasa 5 horas por semana.

```
lambda <-(-1.17+0.24*3+0.00058*1200+0.073*10)
lambda_5 <- (-1.17+0.24*3+0.00058*1200+0.073*5)

diferencia <-(exp(lambda)/(1+exp(lambda)))-(exp(lambda_5)/(1+exp(lambda_5)))
diferencia*100
```

```
[1] 7.814493
```

## 5 Tareas

1. Tarea: realizar todos los cálculos para el modelo de los arrestos, igual como se hizo en clase para los dos modelos, es decir, con y sin variables binarias.

## 6 Modelo Tobit

### 6.1 Motivación

Otro tipo de variable dependiente limitada es una de respuesta de solución de esquina. La variable dependiente es cero para una fracción no trivial *representativa*, pero también existen valores de una distribución **aproximadamente continua** a través de valores positivos. Por ejemplo, el salario, habrá algunos individuos que ganen cero dólares por hora y otros que ganen valores aproximadamente continuos. Otro ejemplo, la cantidad que el individuo gasta en alcohol cada mes. Esta variable asume un amplio rango de valores en personas mayores a los 18 años.

### 6.2 Especificación matemática

Sea  $y$  una variable que asume datos aproximadamente continuos en valores estrictamente positivos, pero que asume cero con probabilidad positiva. En este caso se podría usar un modelo lineal para  $y$ . De hecho, un modelo lineal podría ser una buena aproximación a  $E(y|x_1, x_2, \dots, x_k)$  en especial para  $x_j$  cerca de los valores promedio. Nuevamente, obtendríamos valores ajustado negativos, lo que generaría predicciones negativas para  $y$ , es decir, problemas análogos a los del MPL. También el supuesto de que una variable explicativa que aparece en la forma de nivel tiene un efecto parcial constante sobre  $E(y|\mathbf{X})$  puede ser engañoso. Probablemente la  $Var(y|\mathbf{X})$  sería no constante o heterocedástica, debido a que la distribución de  $y$  se acumula en cero; esta claro que  $y$  no puede tener una distribución normal condicional. Por lo tanto, al igual que en el MPL la inferencia sólo se justifica asintóticamente.

Es importante tener un modelo que implique valores predichos no negativos para  $y$  y, que tenga efectos parciales sensatos sobre un amplio rango de las variables independientes. Además, algunas veces es necesario estimar las características de la distribución de  $y$  dadas  $x_1, \dots, x_k$  más allá de la expectativa condicional. El **Modelo Tobit** es idóneo, el cual expresa la respuesta observada,  $y$ , en términos de una variable latente subyacente:

$$\begin{aligned}y^* &= \beta_0 + \mathbf{x} + u, u|\mathbf{x} \sim Normal(0, \sigma^2)[1] \\ y &= \max(0, y^*)[2]\end{aligned}$$

La variable latente  $y^*$  satisface los supuesto del modelo lineal clásico, en particular tiene:

- Distribución normal
- Homocedástica con una media condicional lineal

La ecuación [2] implica que la variable observable,  $y = y^* \Leftrightarrow y^* \geq 0$ , caso contrario  $y = 0 \Leftrightarrow y^* < 0$ . Debido a que  $y^*$  se distribuye como una normal,  $y$  tiene una distribución continua a través de valores estrictamente positivos. En particular, la densidad de  $y|\mathbf{X}$  es la misma de  $y^*|\mathbf{X}$  para valores positivos, Además:

$$P(y = 0|\mathbf{x}) = P(y^* < 0|\mathbf{x}) = P(u < -\mathbf{x}|\mathbf{x}) = P(u|\sigma < -\mathbf{x} / |\mathbf{x}) = \Phi(-\mathbf{x} / \sigma) = 1 - \Phi(\mathbf{x} / \sigma)[3]$$

Notar que  $u/\sigma \sim N(0,1)$  y es independiente de  $\mathbf{x}$ ; se ha absorbido el intercepto en  $\mathbf{x}$  por simplicidad notacional. Por lo tanto, si  $\mathbf{x}_i, y_i$  se extraen aleatoriamente de la población, la densidad de  $y_i|\mathbf{x}_i$  es:

$$(2\pi\sigma^2)^{-1/2} \exp[-(y_i - \mathbf{x}_i)^2 / (2\sigma^2)] = (1/\sigma)\phi[(y_i - \mathbf{x}_i) / \sigma], y_i > 0[4]$$

$$P(y_i = 0|\mathbf{x}_i) = 1 - \Phi(\mathbf{x}_i / \sigma)[5]$$

Donde  $\phi$  es la función de densidad normal estándar

De las ecuaciones [4] y [5] se obtiene la función log-verosimilitud para cada observación  $i$ :

$$l_i(\beta, \sigma) = 1(y_i = 0)\log[1 - \Phi(\mathbf{x}_i / \sigma)] + 1(y_i > 0)\log((1/\sigma)\phi[(y_i - \mathbf{x}_i) / \sigma])[6]$$

Notese que, esto depende de  $\sigma$ , la desviación estándar de  $u$ , así como de las  $\beta_j$ . La log-verosimilitud para una  $n$  aleatoria se obtiene al sumar [6], a través de todas  $i$ . Las estimaciones de máxima verosimilitud de  $\beta$  y  $\sigma$  se obtienen al maximizar la log-verosimilitud; esto requiere métodos numéricos, pero el software lo realiza.

Para restricciones de exclusión múltiples es fácil usar la prueba de **Wald** o la razón de verosimilitudes.

### 6.3 Interpretaciones de las estimaciones Tobit.

Los resultados de Tobit y MCO son casi siempre similares, lo que hace tentador interpretar las  $\hat{\beta}_j$  de Tobit como si fueran estimaciones de MCO, pero se advierte que no es así de fácil.

La [1] muestra que las  $\beta_j$  miden efectos parciales de las  $x_j$  sobre el  $E(y^*|\mathbf{x})$ . La variable que se busca explicar es  $y$ , pues es el resultado observado (las horas trabajadas o la cantidad donaciones)



De la ecuación [5] podemos obtener  $P(y = 0|\mathbf{x})$ , de ahí podemos estimar  $P(y > 0|\mathbf{x})$  ¿Qué pasa si se quiere estimar el valor esperado de  $y$  en función  $\mathbf{x}$ ? En los modelos Tobit, existe dos expectativas. La primera  $E(y|y > 0, \mathbf{x})$ , que recibe el nombre de **expectativa condicional**. La segunda es  $E(y|\mathbf{x})$ , conocida como la **expectativa no condicional**. La expectativa  $E(y|y > 0, \mathbf{x})$  nos dice que, para los valores dados de  $\mathbf{x}$ , el valor esperado de  $y$  para la subpoblación donde  $y > 0$ , esta expectativa se halla con facilidades a partir  $E(y|\mathbf{x})$ :

$$E(y|\mathbf{x}) = P(y > 0|\mathbf{x}) \cdot E(y|y > 0, \mathbf{x}) = \Phi(\mathbf{x} / \sigma) \cdot E(y|y > 0, \mathbf{x}) [7]$$

Para obtener  $E(y|y > 0, \mathbf{x})$  se puede usar un resultado para las variables aleatorias con distribución normal: si  $z \sim Normal(0, 1)$  entonces  $E(z|z > c) = \phi(c)/[1 - \Phi(c)] \forall c$  constante. Pero  $E(y|y > 0, \mathbf{x}) = \mathbf{x} + E(u|u > -\mathbf{x}) = \mathbf{x} + \sigma E[(u/\sigma)|(u/\sigma) > -\mathbf{x} / \sigma] = \mathbf{x} + \sigma \phi(\mathbf{x} / \sigma) / \Phi(\mathbf{x} / \sigma)$

Debido a que:  $\phi(-c) = \phi(c)$ ,  $1 - \Phi(c) = \Phi(c)$ ,  $u/\sigma \sim Normal(0, 1)$  independiente de  $\mathbf{x}$ . Podemos reescribir a [7] así:

$$E(y|y > 0, \mathbf{x}) = \mathbf{x} + \sigma \lambda(\mathbf{x} / \sigma)$$

Donde:  $\lambda(c) = \frac{\phi(c)}{\Phi(c)}$  recibe el nombre de la **razón inversa de Mills**

## 6.4 Ejemplo

Estimación Tobit y MCO de las horas anuales trabajadas

La variables

- Dependiente ( $Y$ ) son las horas anuales trabajadas por las mujeres
- Variables explicativas o regresoras:
  - *nwifeinc*: Salario de esposo en miles de dólares
  - *educ*: años de educación
  - *exper*: años de experiencia
  - *exper*<sup>2</sup>: años de experiencia al cuadrado
  - *age*: edad de las mujeres
  - *kidslt6*: Niños menores a seis años
  - *kidsge6*: Niños entre 6 y 18 años

### 6.4.1 Verificar que la variable dependiente sea de solución de esquina

activación de paquetes

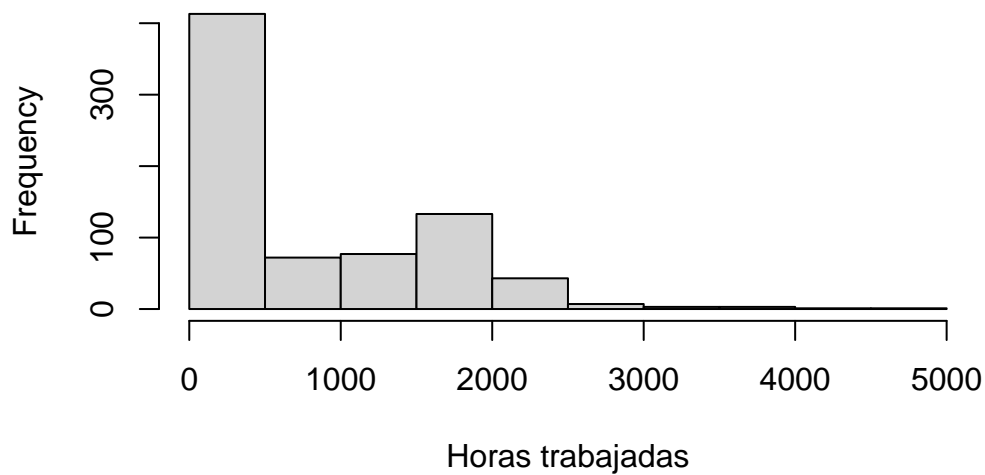
```
pacman::p_load(wooldridge,  
               dplyr,  
               ggplot2,  
               stargazer,  
               censReg,  
               AER)
```

```
datos <- data("mroz")
```

```
# Histograma
```

```
hist(mroz$hours, main = "Histograma de las horas trabajadas", xlab = "Horas trabajadas")
```

**Histograma de las horas trabajadas**

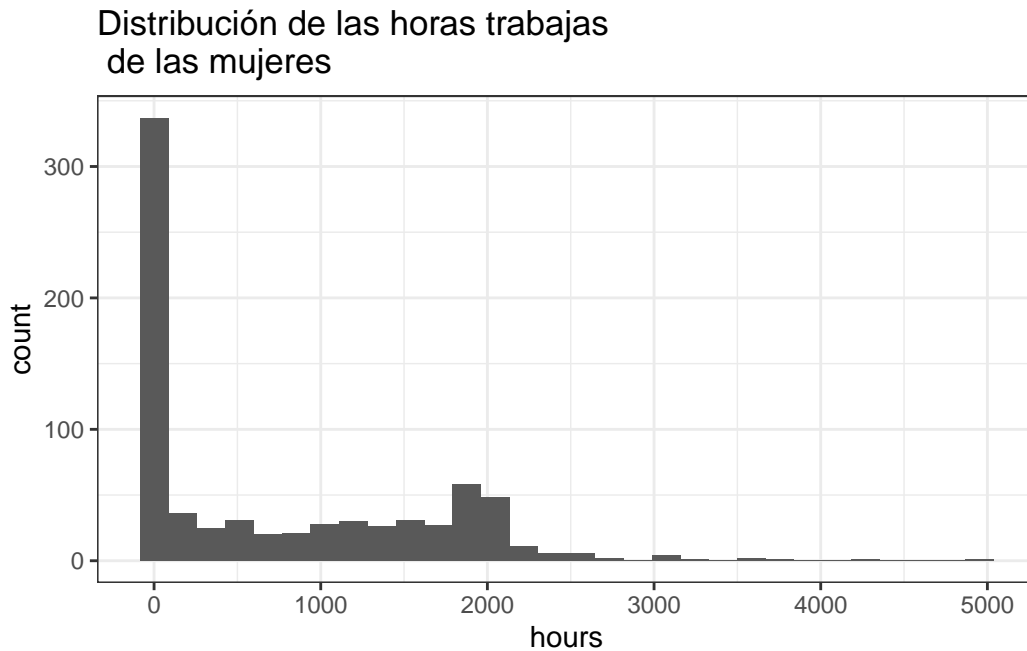


```
library(ggplot2)  
  
ggplot(data = mroz, aes(x=hours))+  
  geom_histogram(binwidth=10)+
```

```
theme_bw()+
labs(title = "Distribución de las horas trabajas\n de las mujeres")
```

Warning in geom\_histogram(bindwidth = 10): Ignoring unknown parameters:  
`bindwidth`

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



En el anterior histograma se puede observar que las mayor cantidad de obervaciones se encuentran en cero, así tenemos indicios de que la variable  $y$  es de solución de esquina, pues a además, la horas trabajadas anuales se amplían hasta 5000 al año. Para saber la proporcion de ceros que tiene la variable dependiente dicotomizamos dicha variables

```
# Transformación de la variable en binaria
mroz$dico.hours <- ifelse(mroz$hours==0,0,1)

# Calculando el porcentaje de ceros en Y
prop.table(table(mroz$dico.hours))*100
```

0	1
43.16069	56.83931

## 6.4.2 Usando MCO

Vamos a determinar las horas trabajadas al año por la mujeres, usando MCO

```
model.mco<-lm(hours~
              nwifeinc+
              educ+
              exper+
              expersq+
              age+
              kidslt6+
              kidsge6,
              data=mroz)

library(stargazer)

stargazer(model.mco,
          type = "text",
          dep.var.labels = "Horas al año trabajadas")
```

```
=====
                        Dependent variable:
-----
                        Horas al año trabajadas
-----
nwifeinc                -3.447
                        (2.544)

educ                    28.761**
                        (12.955)

exper                   65.673***
                        (9.963)

expersq                 -0.700**
                        (0.325)

age                     -30.512***
                        (4.364)

kidslt6                 -442.090***
```

```

(58.847)

kidsge6          -32.779
                  (23.176)

Constant          1,330.482***
                  (270.785)

-----
Observations      753
R2                0.266
Adjusted R2       0.259
Residual Std. Error 750.179 (df = 745)
F Statistic       38.495*** (df = 7; 745)
=====
Note:             *p<0.1; **p<0.05; ***p<0.01

```

# Si la variable "y" ajustada tiene valores menores que cero, significa que no es un buen ajuste

```
summary(model.mco$fitted.values)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-719.8	417.5	737.7	740.6	1093.1	1614.7

```
summary(mroz$hours)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	288.0	740.6	1516.0	4950.0

Como me arroja valores negativos ajustados, significa que el MCO no esta ajustando de forma adecuada a la variable *y* de solución de esquina. Además, el efecto es constante. Esto me sugiere que se debe usar un modelo Tobit para ajustar una variable de solución de esquina.

### 6.4.3 Modelo Tobit

Por lo antes mencionado, ajustamos a la variable *hours* con un modelo **Tobit**

Verificar el cumplimiento de la variable *hours* para usar un modelo Tobit.

```
# Porcentaje de ceros en la variable hours
```

```
mroz$dico <- ifelse(mroz$hours==0,0,1)
prop.table(table(mroz$dico))*100
```

```
      0      1
43.16069 56.83931
```

En este ejemplo aproximadamente el 43% de los datos de *hours* son cero y el resto datos aproximadamente continuos

Procedemos a ajustar un modelo Tobit

```
modelo.tobit <- censReg(hours~
  nwifeinc+
  educ+
  exper+
  expersq+
  age+
  kidslt6+
  kidsge6,
  data=mroz,
  left = 0)
```

```
modelo.tobit2 <- tobit(hours~
  nwifeinc+
  educ+
  exper+
  expersq+
  age+
  kidslt6+
  kidsge6,
  data=mroz)
```

Una vez, ejecutadas las dos regresiones (MCO y Tobit) las ponemos a comparación, de tal forma que, se replique la **Tabla 17.2B** del libro de Wooldridge (2010)

```
stargazer(model.mco,
  modelo.tobit,
  modelo.tobit2 ,
```

```

digits = 2,
type = "text",
df=F,
title = "Estimación Tobit y MCO de las horas anuales trabajas",
dep.var.labels = "Variable dependiente: horas anuales trabajadas",
header = F,
column.labels = c("MCO", "Tobit", "Tobit2"),
model.names = F)

```

# Estimación Tobit y MCO de las horas anuales trabajas

=====			
	Dependent variable:		
	-----		
	Variable dependiente: horas anuales trabajadas		
	MCO	Tobit	Tobit2
	(1)	(2)	(3)
-----			
nwifeinc	-3.45 (2.54)	-8.81** (4.46)	-8.81** (4.46)
educ	28.76** (12.95)	80.65*** (21.58)	80.65*** (21.58)
exper	65.67*** (9.96)	131.56*** (17.28)	131.56*** (17.28)
expersq	-0.70** (0.32)	-1.86*** (0.54)	-1.86*** (0.54)
age	-30.51*** (4.36)	-54.41*** (7.42)	-54.41*** (7.42)
kidslt6	-442.09*** (58.85)	-894.02*** (111.88)	-894.02*** (111.88)
kidsge6	-32.78 (23.18)	-16.22 (38.64)	-16.22 (38.64)
logSigma		7.02*** (0.04)	

Constant	1,330.48*** (270.78)	965.31** (446.44)	965.31** (446.44)
-----			
Observations	753	753	753
R2	0.27		
Adjusted R2	0.26		
Log Likelihood		-3,819.09	-3,819.09
Akaike Inf. Crit.		7,656.19	
Bayesian Inf. Crit.		7,697.81	
Residual Std. Error	750.18		
F Statistic	38.50***		
Wald Test			253.86***
=====			
Note:	*p<0.1; **p<0.05; ***p<0.01		

Si quiero hacer comparables las estimaciones Tobit con MCO se debe multiplicar por el factor de ajuste. El factor escalar  $\mathbf{EPP} \ n^{-1} \sum_{i=1}^n \Phi(\mathbf{x}_i' / \hat{\sigma})$  resulta que es aproximadamente de 0.589. Por ejemplo, *educ* por 0.589 se obtiene  $0.589(80.65) \approx 47.50$ , por lo tanto, si una mujer aumenta un año a su educación, en promedio se sumara 47.5 horas de trabajo, esto es mayor al MCO, que es de 28.76. Se podría usar otro escalar a partir de los valores promedio de todas las variables explicativas, entonces se calcula el *EPA*  $\Phi(\bar{\mathbf{x}}_i' / \hat{\sigma})$ , es aproximadamente 0.645

### 6.4.3.1 Efecto marginal

A continuación, uso el comando *margEff()* para encontrar los efectos marginales de la estimación Tobit

```
summary(margEff(modelo.tobit))
```

	Marg. Eff.	Std. Error	t value	Pr(> t )	
nwifeinc	-5.32644	2.69073	-1.9796	0.0481217	*
educ	48.73409	12.96341	3.7594	0.0001837	***
exper	79.50423	10.30497	7.7151	3.886e-14	***
expersq	-1.12651	0.32326	-3.4848	0.0005213	***
age	-32.87692	4.45770	-7.3753	4.383e-13	***
kidslt6	-540.25683	66.62393	-8.1091	2.220e-15	***
kidsge6	-9.80053	23.36134	-0.4195	0.6749580	
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					



## Interpretaciones

- Si el salario del esposo aumenta en 1000 dólares al año, las horas de trabajo de la mujer disminuyen en 5.32 horas. El mayor efecto, que es altamente significativo, sigue siendo el aumento de niños pequeños, pues en promedio, si se aumenta un infante menor a seis años las horas de trabajo decrecen en 540 horas al año.

### 6.4.4 No linealidad del modelo Tobit

```
mod.stobit <- censReg(hours~kidslt6 ,  
                      left = 0,  
                      data=mroz)  
mod.Smco <- lm(hours~kidslt6,  
               data=mroz)  
  
stargazer(mod.stobit, type = "text")
```

```
=====
                        Dependent variable:
                        -----
                                hours
                        -----
kidslt6                      -774.751***
                              (116.086)

logSigma                      7.191***
                              (0.038)

Constant                     489.522***
                              (59.105)

-----
Observations                  753
Log Likelihood                -3,930.753
Akaike Inf. Crit.             7,867.505
Bayesian Inf. Crit.           7,881.377
=====
Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
# Añadiendo la curva de regresión Tobit

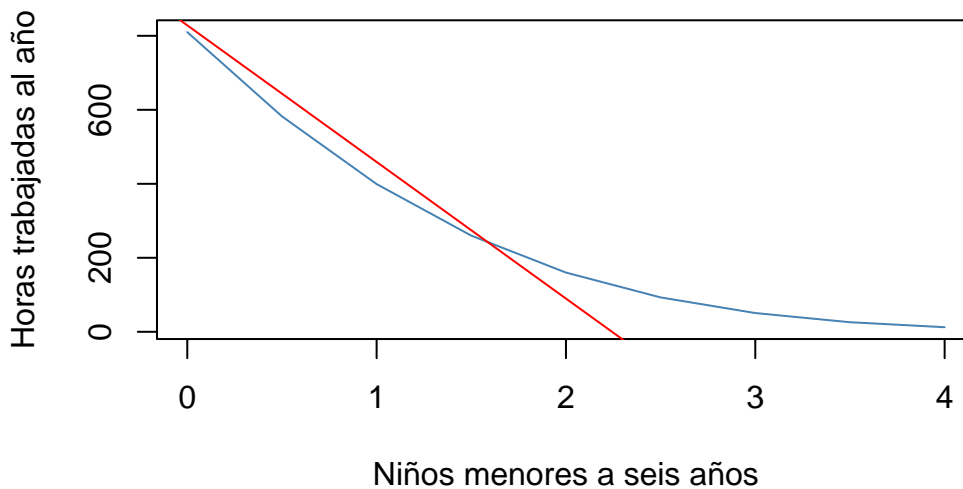
x <- seq(0,4, 0.5)

y <- pnorm((coef(mod.stobit)[1]+coef(mod.stobit)[2]*x)/exp(coef(mod.stobit)[3]))*(coef(mod.stobit)[4]+coef(mod.stobit)[5]*x)

plot(x=x,
     y=y,
     xlab = "Niños menores a seis años",
     ylab="Horas trabajadas al año",
     col="steelblue",
     main="Modelo Tobit para la relación niños pequeños y horas trabajadas de una mujer",
     type="l")

abline(mod.Smco, col="red")
```

## Tobit para la relación niños pequeños y horas trabajadas d



En el gráfico anterior podemos observar que añadir un niño menor a seis años hace que se reduzca las horas dedicadas al trabajo. Sin embargo, en la línea azul del modelo Tobit la reducción de horas es decreciente a medida que se tiene más hijos pequeños, en el MCO (línea de color rojo), la reducción de las horas trabajadas es la misma con el apareamiento de menor a seis años. Es decir, no importa si pasas de cero hijo a uno hijo o de 3 a 4, la reducción de las horas trabajadas es la misma. No está tomando en cuenta el aprendizaje de la madre con cada nuevo hijo.

#### 6.4.4.1 Evaluación informal del modelo Tobit [Problemas de especificación]

```
probit.tobit <-glm(dico.hours~
  nwifeinc+
  educ+
  exper+
  expersq+
  age+
  kidslt6+
  kidsge6,
  data=mroz,
  family = binomial(link = "probit"))

stargazer(probit.tobit,
  type = "text",
  title = "Estimación Tobit y MCO de las horas anuales trabajadas",
  dep.var.labels = "Variable dependiente: horas anuales trabajadas 0 y 1")
```

Estimación Tobit y MCO de las horas anuales trabajadas

```
=====
                        Dependent variable:
-----
Variable dependiente: horas anuales trabajadas 0 y 1
-----
```

nwifeinc	-0.012** (0.005)
educ	0.131*** (0.025)
exper	0.123*** (0.019)
expersq	-0.002*** (0.001)
age	-0.053*** (0.008)
kidslt6	-0.868***

```

(0.118)

kidsge6          0.036
                  (0.044)

Constant         0.270
                  (0.508)

-----
Observations      753
Log Likelihood    -401.302
Akaike Inf. Crit. 818.604
=====
Note:              *p<0.1; **p<0.05; ***p<0.01

```

Luego procedemos a usar los coeficientes para la comparación entre el modelo Probit y el Tobit, el objetivo es evaluar la validez del modelo Tobit.

```

z <- coef(probit.tobit)
m <- coef(modelo.tobit)

comparacion <- data.frame(z, (m[1:8])/exp(m[9]), z-(m[1:8])/exp(m[9]))
names(comparacion) <- c("probit", "beta/sigma", "diferencia")
comparacion

```

	probit	beta/sigma	diferencia
(Intercept)	0.270073573	0.860326776	-0.590253203
nwifeinc	-0.012023637	-0.007855680	-0.004167957
educ	0.130903969	0.071875266	0.059028703
exper	0.123347168	0.117256469	0.006090698
expersq	-0.001887067	-0.001661427	-0.000225640
age	-0.052852442	-0.048488379	-0.004364063
kidslt6	-0.868324680	-0.796795431	-0.071529248
kidsge6	0.036005611	-0.014454263	0.050459873

Tobit de *nwifeinc* entre  $\hat{\sigma} = 1122.02$ , se obtuvo  $-8.81/1122.02 = -0.0079$ ; el coeficiente probit de *nwifeinc* es de cerca de  $-0.012$ , lo cual es diferente, pero no de forma drástica. En *kidslt6*, el coeficiente estimado entre  $\hat{\sigma}$  es de cerca de  $-0.797$ , en comparación con la estimación probit de  $-0.868$ . De nuevo, ésta no es una diferencia enorme, pero indica que tener niños pequeños tiene un efecto mayor sobre la decisión inicial de participar en la fuerza laboral que sobre cuántas horas elige trabajar una mujer una vez que está en dicha fuerza. (Tobit promedia de

forma efectiva estos dos efectos.) No se sabe si los efectos son estadísticamente diferentes, pero son del mismo orden de magnitud.

Por lo tanto, se podría decir que el modelo Tobit es adecuado, para ajustar a la variable *hours*

¿Qué sucede si se concluye que el modelo Tobit es inadecuado? Existen modelos, que suelen conocerse como modelos de **dos partes** o **de obstáculos**, que se pueden usar cuando Tobit es inadecuado

# 7 Modelo Poisson

## 7.1 Introducción y motivación

Una tercera clase de variable dependiente no negativa es una **variable de conteo**, que puede asumir valores enteros no negativos:  $[0, 1, 2, \dots]$ , específicamente los que nos interesa son los casos en los que  $y$  asume pocos valores, incluido el cero. Ejemplos:

- El número de medallas que puede obtener un deportista en una olimpiada,
- El número de hijos que tiene una mujer
- El número de publicaciones al año de un científico

Al igual que las respuestas, binaria y Tobit, un modelo lineal para  $E(y|x_1, x_2, \dots, x_k)$ , quizá no proporciona el mejor ajuste a lo largo de todos los valores de las variables explicativas. Sin embargo, es informativo comenzar con un modelo lineal.

Como en un modelo **Tobit** no se puede obtener el logaritmo de una variable de conteo que asume valores de cero. Un método útil es modelar el valor esperado como una función exponencial:

$$E(y|x_1, x_2, \dots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad [1]$$

```
pacman::p_load(wooldridge,
               stargazer,
               tidyverse)
```

### 7.1.1 Recordatorio

```
data("wage1", package = "wooldridge")

#Modelo lineal

salario.lm<-lm(wage~educ,
```

```

wage1)

library(stargazer)

stargazer(salario.lm,
          type = "text")

```

```

=====
                        Dependent variable:
                        -----
                                wage
-----
educ                                0.541***
                                   (0.053)

Constant                            -0.905
                                   (0.685)

-----
Observations                        526
R2                                  0.165
Adjusted R2                         0.163
Residual Std. Error      3.378 (df = 524)
F Statistic               103.363*** (df = 1; 524)
=====
Note:          *p<0.1; **p<0.05; ***p<0.01

```

```

# Gráfica de la relación salario y la educación

plot(wage~educ,
     wage1,
     pch=20,
     col="steelblue",
     ylab = "Salario en USD por hora",
     xlab="años de educación",
     main="La relación entre el salario y la educación")

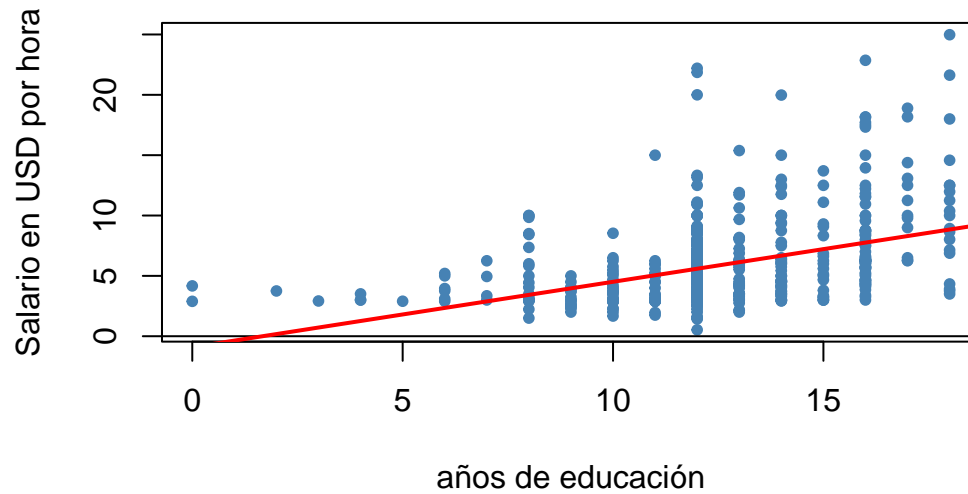
abline(0,0)

abline(salario.lm,

```

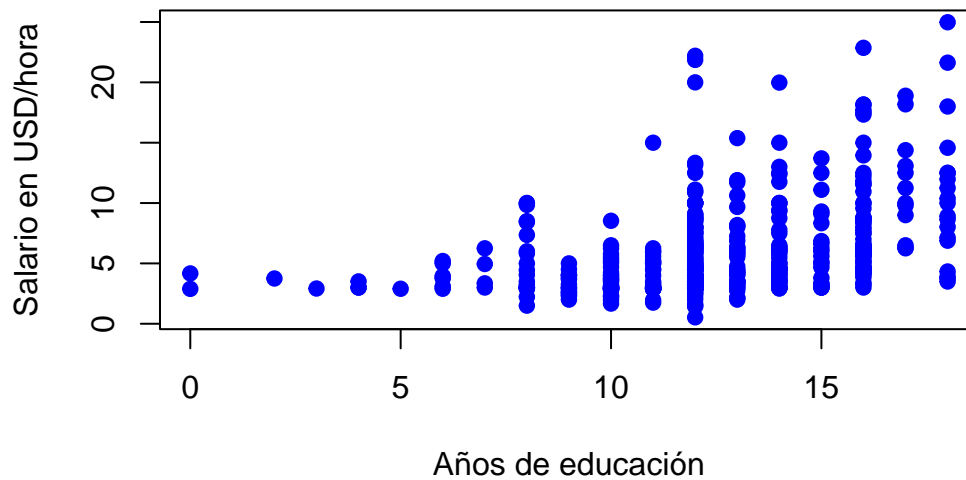
```
lw=2,  
col="red")
```

## La relación entre el salario y la educación



```
plot(y=wage1$wage,  
     x=wage1$educ,  
     col="blue",  
     pch=19,  
     xlab="Años de educación",  
     ylab="Salario en USD/hora")
```





### Interpretaciones

- Un aumento de un año de educación, esta asociado en promedio a un incremento en el salario de 54 centavos por cada trabajada.

Es decir que la forma funcional al parecer, es la siguiente:

$$wage = \exp(\beta_0 + \beta_1 educ + u) \quad [2]$$

La ecuación [2] no es lineal en los parámetros, para usar el modelo de regresión se usa un cambio usando la función logarítmica, tenemos:

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

```
log.lin<-lm(log(wage)~educ,
            wage1)
stargazer(log.lin,
          type = "text")
```

```
=====
Dependent variable:
```

```

-----
                                log(wage)
-----
educ                                0.083***
                                (0.008)

Constant                            0.584***
                                (0.097)

-----
Observations                        526
R2                                  0.186
Adjusted R2                        0.184
Residual Std. Error      0.480 (df = 524)
F Statistic                119.582*** (df = 1; 524)
=====
Note:                *p<0.1; **p<0.05; ***p<0.01

```

## Intepretación

```
exp(coef(log.lin)[1])
```

(Intercept)

1.792789

- $e^{0.584} = 1.79$  Si no hay cambios en la educación, se predice un ingreso promedio por hora trabajada de 1.79 USD
- Un aumento de un año en la educación esta asociado a un incremento de 8.3% en el salario por hora trabajada

```
# Gráfica de la relación salario y la educación
```

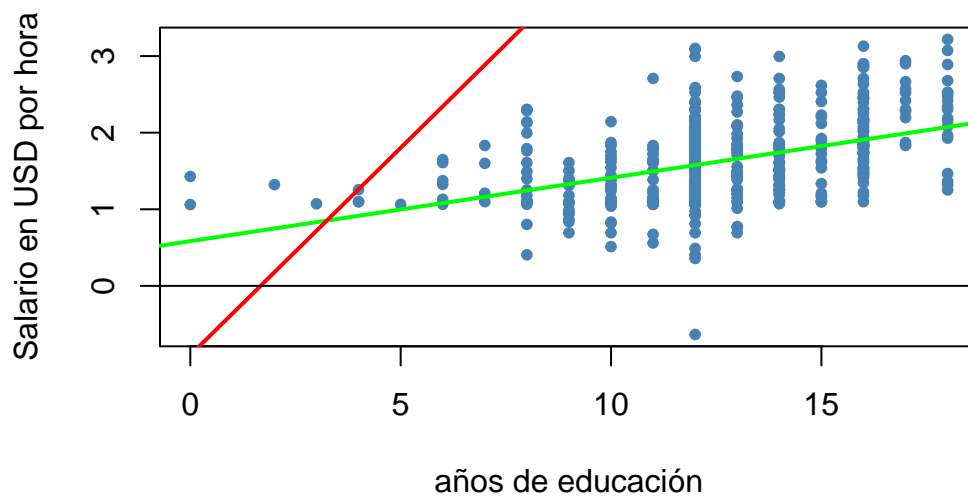
```
plot(log(wage)~educ,
     wage1,
     pch=20,
     col="steelblue",
     ylab = "Salario en USD por hora",
     xlab="años de educación",
     main="La relación entre el salario y la educación")
```

```
abline(log.lin,
       lw=2,
       col="green")

abline(salario.lm,
       col="red",
       lwd=2)

abline(0,0)
```

## La relación entre el salario y la educación



### 7.1.2 Otro ejemplo

```
data("CASchools", package = "AER")

CASchools$Notas<-(CASchools$read+CASchools$math)/2

lineal.model<-lm(Notas~income,
                 CASchools)
lineal.log<-lm(Notas~log(income),
               CASchools)
```

```

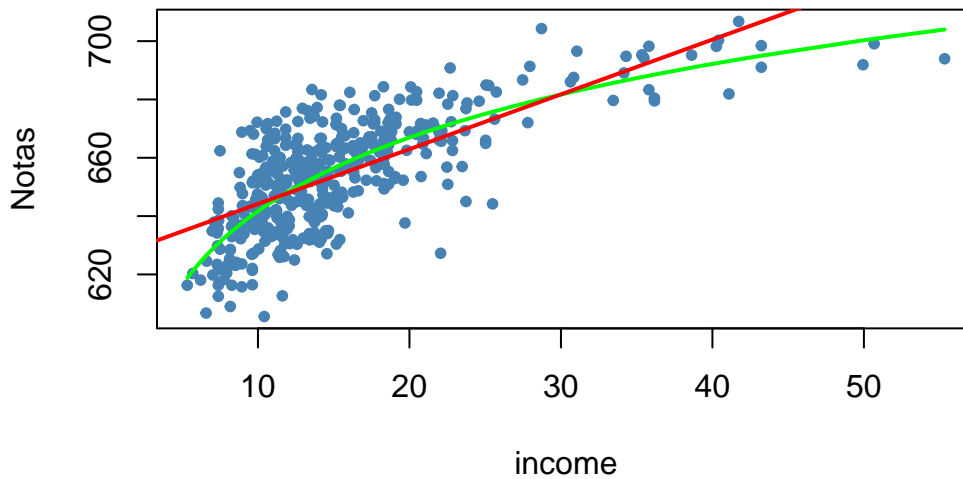
plot(Notas~income,
     CASchools,
     col="steelblue",
     pch=20,
     main="Linea de regresión Notas-ingreso")

order_id2<-order(CASchools$income)

lines(CASchools$income[order_id2],
      fitted(lineal.log)[order_id2],
      col="green",
      lwd=2)
abline(lineal.model,
      col="red",
      lwd=2)

```

### Linea de regresión Notas-ingreso



Volviendo a la ecuación [1], debido a que  $\exp(\cdot)$  siempre es positivo. [1] asegura que los valores predichos para  $y$  también sean positivos. Aunque [1] es más complicada que un modelo lineal, básicamente ya se sabe como interpretar los coeficientes, al obtener el logaritmo de la ecuación [1]

$$\log[E(y|x_1, \dots, x_k)] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad [2]$$

es decir, que el logaritmo del valor esperado es lineal. Por lo tanto, mediante las propiedades de la aproximación de la función logaritmo tenemos:

$$\% \Delta E(y|\mathbf{x}) \approx (100\beta_j)\Delta x_j [3]$$

Es decir,  $100\beta_j$  es el cambio porcentual en  $E(y|\mathbf{x})$ , dado un incremento de una unidad en  $x_j$ . A veces, es necesaria una estimación más precisa y es fácil de encontrar una, al observar los cambios discretos en el valor esperado. Manteniendo todas las variables explicativas fijas, excepto  $x_j$  y, sea  $x_k^0$  el valor inicial y  $x_k^1$  el valor siguiente. Entonces, el cambio proporcional en el valor esperado es:

$$[\exp(\beta_o + \mathbf{x}_{k-1} \beta_{k-1} + \beta_k x_k^1) / \exp(\beta_o + \mathbf{x}_{k-1} \beta_{k-1} + \beta_k x_k^0)] - 1 = \exp(\beta_k \Delta x_k) - 1 [4]$$

Donde:  $\mathbf{x}_{k-1} \beta_{k-1}$  es una abreviatura de  $\beta_1 x_1 + \dots + \beta_{k-1} x_{k-1}$  y,  $\Delta x_k = x_k^1 - x_k^0$ . Cuando  $\Delta x_k = 1$ , la variable  $x_k$  es binaria que se cambia de cero a uno, entonces el cambio es  $\exp(\beta_k) - 1$ . Dada  $\hat{\beta}_k$ , se obtiene  $\exp(\hat{\beta}_k) - 1$  y se multiplica por el 100 para transformar el cambio proporcional en un cambio porcentual.

Si por ejemplo  $x_j = \log(z_j)$  para alguna variable  $z_j > 0$ , entonces su coeficiente  $\beta_j$  se interpreta como una elasticidad respecto a  $z_j$

Debido a que [1] es no lineal en sus parámetros, no se puede usar métodos de regresión lineal. Entonces usamos la **estimación máxima verosimilitud (EMV)** y también el método relacionado a la **estimación de cuasi máxima verosimilitud (ECMV)**

A lo largo de los cursos de econometría se ha presentado la normalidad como el supuesto de distribución estándar para regresión lineal. Este supuesto no puede usarse en una variable de conteo (pues la distribución normal es para variables continuas que asuman todos los valores) que asume sólo pocos valores, la distribución será muy distinta a la normal. En su lugar, la distribución nominal para los datos de conteo es la **distribución Poisson**

Como nos interesa el efecto de las variables explicativas sobre  $y$ , se debe observar la distribución de Poisson condicional a  $\mathbf{x}$ . La distribución Poisson está determinada por completo por su media, así sólo se necesita especificar  $E(y|\mathbf{x})$ , esta tiene la misma forma de [1] que se abrevia  $\exp(\mathbf{x})$ . Entonces, la probabilidad de que  $y$  será igual al valor  $h$ , condicional sobre  $\mathbf{x}$ , es:

$$P(y = h|\mathbf{x}) = \exp[-\exp(\mathbf{x})][\exp(\mathbf{x})]^h / h!, h = 0, 1, \dots [5]$$

Donde  $h!$  denota el factorial. Esta distribución, que es la base del **modelo de regresión de Poisson**, permite hallar las probabilidades condicionales para cualquier valor de variables explicativas. Por ejemplo,  $P(y = 0|\mathbf{x}) = \exp[-\exp(\mathbf{x})]$ . Una vez que se tienen las estimaciones de  $\beta_j$ , se pueden insertar en las probabilidades para diferentes valores  $\mathbf{x}$ .

Dada una muestra aleatoria  $[(\mathbf{x}_i, y_i) : i = 1, 2, \dots, n]$ , se puede construir la función **log-verosimilitud**:

$$\mathcal{L}(\beta) = \sum_{i=1}^n l_i(\beta) = \sum_{i=1}^n [y_i \mathbf{x}_i - \exp(\mathbf{x}_i)] \quad [6]$$

Se desecha el término  $\log(y_i!)$ . Esta función se maximiza usando EMV, aunque la EMV de Poisson no es cerrada.

Igual que los modelo logit, probit y Tobit, no se pueden comparar directamente las magnitudes de las estimaciones del Poisson de una función exponencial con las estimaciones de MCO. se hace comparables de la siguiente forma:

### 7.1.2.1 Variables explicativas continuas

Se aplica el efecto parcial de  $x_j$  respecto a  $E(y|x_1, x_2, \dots, x_k)$ :

$$\frac{\partial E(y|x_1, x_2, \dots, x_k)}{\partial x_j} = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \times \beta_j [7]$$

Es interesante el factor escalar **EPP**:

$$n^{-1} \sum_{i=1}^n \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k) = n^{-1} \sum_{i=1}^n \hat{y}_i [8]$$

es simplemente el promedio muestral  $\bar{y}$  de  $y_i$  donde se definen los valores ajustados como  $\hat{y}_i = \exp(\hat{\beta}_0 + \mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ . Es decir, para la regresión Poisson con una función media exponencial, el promedio de los valores ajustados es el mismo que el promedio de los resultados originales de  $y_i$ , tal como el caso de regresión lineal. Esto simplifica el escalar de las estimaciones Poisson  $\hat{\beta}_j$ , para hacerlas comparables a las estimaciones MCO,  $\hat{\gamma}_j$  para una variable explicativa continua, se puede comparar con  $\hat{\gamma}_j$  con  $\bar{y} \cdot \hat{\beta}_j$

Aunque el análisis de EMV de Poisson es un primer paso para los datos de conteo, suele ser muy restrictivo. Todas las probabilidades y los momentos mayores de la distribución Poisson se determinan por completo por la media. Por ejemplo, la varianza es igual a la media:

$$\text{Var}(y|\mathbf{x}) = E(y|\mathbf{x}) \quad [9]$$

Esto es restrictivo y se viola en muchas aplicaciones. Por fortuna, la distribución de Poisson tiene una propiedad de robustez muy buena, es decir, que se mantenga o no la distribución de Poisson, se obtienen estimadores asintóticamente normales y consistentes con las  $\beta_j$

Cuando se EMV de Poisson, pero no se supone que la distribución de Poisson sea correcta, este análisis recibe el nombre de **Estimación de cuasi máxima verosimilitud (ECMV)**. LA ECMV de Poisson es muy útil debido a que esta programada en muchos paquetes econométricos. Sin embargo, a menos que el supuesto de varianza de Poisson [9] se mantenga, se deben ajustar los errores estándar, de la siguiente forma:

El ajuste a los errores estándar está disponible cuando se supone que la varianza es proporcional a la media:

$$Var(y|\mathbf{x}) = \sigma^2 E(y|\mathbf{x}) \quad [10]$$

Donde:  $\sigma^2$  es un parámetro desconocido.

- Cuando  $\sigma^2 = 1$  se obtiene el supuesto de varianza de Poisson [9]
- Si  $\sigma^2 > 1$  la varianza es mayor que la media para toda  $\mathbf{x}$ , esto se llama **sobredispersión** común en regresiones de conteo.
  - Si  $\sigma^2 < 1$  la varianza es menor que la media para toda  $\mathbf{x}$ , esto se llama **subdispersión** es poco común.

Bajo [10] es fácil ajustar los errores estándar de la EMV de Poisson. Si  $\hat{\beta}_j$  denota la ECMV de Poisson y se definen los residuales como  $\hat{u}_i = y_i - \hat{y}_i$ , donde  $\hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})$ . Un estimador consistente de  $\sigma^2 = (n - k - 1)^{-1} \sum_{i=1}^n \frac{\hat{u}_i^2}{\hat{y}_i}$ , donde la división entre  $\hat{y}_i$  es el ajuste apropiado de heterocedasticidad y  $n - k - 1 = gl$  dadas las  $n$  observaciones y  $k + 1$  estimadores  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ . Si  $\sigma = \sqrt{\sigma^2}$ , se multiplican los errores estándar Poisson usuales por  $\hat{\sigma}$ . Si  $\hat{\sigma}$  es notablemente mayor que uno, los errores estándar corregidos pueden ser mucho mayores que los errores estándar nominales, generalmente son incorrectos, de la EMV de Poisson.

Bajo el supuesto de distribución de Poisson, se puede usar el estadístico de la razón de verosimilitudes para probar las restricciones de exclusión, que siempre, tienen la forma de  $RV = 2(l_{nr} - L_r)$ . Si se tiene  $q$  restricciones de exclusión, el estadístico se distribuye aproximadamente con  $\chi_q^2$  bajo la hipótesis nula. Bajo el supuesto menos restrictivo de [10], un simple ajuste está disponible si se divide  $RV = 2(l_{nr} - L_r)$  entre  $\sigma^2$  donde  $\sigma^2$  se obtiene del modelo no restringido.

## 7.2 Ejemplo [Regresión de Poisson para número de arrestos]

La base de datos **crime1** contiene información sobre arrestos durante 1986 y otros datos, sobre 2725 hombres nacidos en California en 1960 o 1961. Cada hombre de la muestra fue arrestado al menos una vez antes 1986.

Las variables:

- **narr86**: indica el número de veces que un hombre fue arrestado durante 1986: esta variable es cero para la mayoría de los hombres de la muestra (72.29%) y varía desde 0 hasta 12. (El porcentaje de hombres detenidos una sola vez durante 1986 es 20.51%)

```
pacman::p_load(wooldridge,
               tidyverse)

data("crime1")

crime1 %>%
str()
```

```
'data.frame': 2725 obs. of 16 variables:
 $ narr86 : int 0 2 1 2 1 0 2 5 0 0 ...
 $ nfarr86: int 0 2 1 2 1 0 2 3 0 0 ...
 $ nparr86: int 0 0 0 1 0 0 1 5 0 0 ...
 $ pcnv : num 0.38 0.44 0.33 0.25 0 ...
 $ avgsen : num 17.6 0 22.8 0 0 ...
 $ tottime: num 35.2 0 22.8 0 0 ...
 $ ptime86: int 12 0 0 5 0 0 0 0 9 0 ...
 $ qemp86 : num 0 1 0 2 2 4 0 0 0 3 ...
 $ inc86 : num 0 0.8 0 8.8 8.1 ...
 $ durat : num 0 0 11 0 1 ...
 $ black : int 0 0 1 0 0 0 1 0 1 0 ...
 $ hispan : int 0 1 0 1 0 0 0 0 0 1 ...
 $ born60 : int 1 0 1 1 0 1 1 1 1 1 ...
 $ pcnvsq : num 0.1444 0.1936 0.1089 0.0625 0 ...
 $ pt86sq : int 144 0 0 25 0 0 0 0 81 0 ...
 $ inc86sq: num 0 0.64 0 77.44 65.61 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

```
# Tabla de porcentaje
crime1 %>%
  with(table(narr86)) %>%
  prop.table() %>%
  round(digits = 2) %>%
  print()
```

```
narr86
  0    1    2    3    4    5    6    7    9   10   12
0.72 0.21 0.04 0.02 0.00 0.00 0.00 0.00 0.00 0.00 0.00
```

- **pcnv:** Es la proporción (no el porcentaje) de detenciones anteriores a 1986 que condujeron a una condena (?)



```
crime1 %>%
  with(table(pcnv)) %>%
  round(digits = 2) %>%
  print()
```

```
pcnv
      0 0.0799999982118607 0.09000000035762787 0.100000001490116
1260      2      1      1
0.109999999403954 0.129999995231628 0.140000000596046 0.170000001788139
      1      3      6      17
0.1800000007152557 0.2000000002980232 0.219999998807907 0.230000004172325
      2      24      4      4
      0.25 0.259999990463257 0.2700000010728836 0.28999999165535
      53      1      2      21
0.3000000011920929 0.3100000002384186 0.319999992847443 0.3300000013113022
      4      2      1      139
0.3600000014305115 0.379999995231628 0.389999985694885 0.400000005960464
      7      20      1      40
0.409999996423721 0.419999986886978 0.4300000007152557 0.439999997615814
      1      5      14      10
0.449999988079071 0.469999998807907      0.5 0.529999971389771
      4      3      313      1
0.5400000021457672 0.5600000002384186 0.569999992847443 0.6000000023841858
      2      8      11      34
0.620000004768372 0.629999995231628 0.639999985694885 0.6700000016689301
      1      6      2      85
0.699999988079071 0.709999978542328 0.7300000019073486      0.75
      1      2      1      20
0.8000000011920929 0.829999983310699      1
      8      3      574
```

- **avgsen** es la duración promedio de las condenas anteriores cumplidas (cero para la mayoría de casos)

```
crime1 %>%
  with(table(avgsen)) %>%
  print()
```

```
avgsen
      0 0.3000000011920929 0.8000000011920929 0.899999976158142
2591      1      2      2
```

1.10000002384186	1.39999997615814	2.20000004768372	2.29999995231628
5	1	1	1
2.59999990463257	2.900000009536743	3.5	4
2	1	1	2
4.30000019073486	4.80000019073486	4.900000009536743	5.5
1	1	1	2
5.59999990463257	6	6.09999990463257	6.19999980926514
4	1	2	1
6.30000019073486	6.69999980926514	6.900000009536743	7.09999990463257
1	1	1	3
7.19999980926514	7.59999990463257	7.80000019073486	7.900000009536743
1	1	2	1
8.10000038146973	8.19999980926514	8.30000019073486	8.39999961853027
1	1	1	1
8.60000038146973	8.89999961853027	9	9.10000038146973
2	1	2	1
9.30000019073486	9.5	9.60000038146973	9.69999980926514
1	1	1	2
9.80000019073486	9.89999961853027	10	10.5
1	1	1	1
10.6000003814697	10.6999998092651	10.8999996185303	11
1	1	2	1
11.1000003814697	11.3000001907349	11.3999996185303	11.6000003814697
1	1	1	2
11.6999998092651	11.8000001907349	11.8999996185303	12.1000003814697
1	2	3	1
12.1999998092651	12.3999996185303	12.5	12.6999998092651
2	1	2	1
12.8999996185303	13.3000001907349	13.3999996185303	13.6999998092651
1	1	1	1
14.1999998092651	14.3000001907349	14.3999996185303	14.8000001907349
1	1	1	1
15	15.6999998092651	16	16.1000003814697
1	1	1	1
16.2000007629395	16.5	16.6000003814697	17
1	1	1	1
17.1000003814697	17.6000003814697	17.7000007629395	18.3999996185303
1	1	1	3
18.5	18.7000007629395	18.8999996185303	19.2999992370605
1	1	2	2
20.2999992370605	20.6000003814697	21.7000007629395	21.7999992370605
1	1	1	1
22	22.7999992370605	23.5	23.8999996185303

1	1	1	1
24.3999996185303	24.6000003814697	30.3999996185303	31.2999992370605
1	2	1	1
31.7000007629395	31.8999996185303	35.4000015258789	36.0999984741211
1	1	1	1
39	40.5	47.0999984741211	59.2000007629395
1	1	1	1

- **tottime**: tiempo en prisión desde los 18 años (meses)

```
crime1 %>%
  with(table(tottime)) %>%
  print()
```

```
tottime
0 0.300000011920929 0.800000011920929 0.899999976158142
2591 1 2 2
1.10000002384186 1.39999997615814 2.20000004768372 2.29999995231628
5 1 1 1
2.59999990463257 2.90000009536743 4 4.80000019073486
2 1 2 1
5.5 5.59999990463257 6 6.19999980926514
1 2 1 1
6.69999980926514 7 7.09999990463257 7.19999980926514
1 1 1 1
7.59999990463257 7.80000019073486 8.10000038146973 8.19999980926514
1 1 1 1
8.30000019073486 8.60000038146973 8.89999961853027 9
1 2 1 1
9.10000038146973 9.30000019073486 9.5 9.60000038146973
1 1 1 1
9.80000019073486 9.89999961853027 10.5 10.8999996185303
1 1 1 1
11 11.1000003814697 11.1999998092651 11.3000001907349
2 1 1 1
11.3999996185303 11.6000003814697 11.6999998092651 11.8000001907349
1 1 1 1
11.8999996185303 12.1999998092651 12.3999996185303 12.5
3 4 1 2
12.8999996185303 13.3000001907349 13.3999996185303 13.6999998092651
1 1 1 1
14.1999998092651 14.3000001907349 14.3999996185303 14.6999998092651
```

2	1	1	1
15.6999998092651	16.1000003814697	16.5	16.6000003814697
1	1	1	1
16.7999992370605	17	17.1000003814697	18
1	1	1	1
18.3999996185303	18.5	18.8999996185303	19.2999992370605
2	1	2	2
19.3999996185303	20	20.2999992370605	20.7000007629395
1	1	1	1
21.2000007629395	21.2999992370605	21.3999996185303	21.7000007629395
1	1	1	1
21.7999992370605	22	22.3999996185303	22.7999992370605
1	1	1	1
23.2000007629395	23.5	23.7000007629395	23.8999996185303
1	1	1	1
24.2000007629395	24.3999996185303	24.6000003814697	25.2000007629395
1	1	1	1
25.7999992370605	29.1000003814697	29.6000003814697	30
1	1	1	1
30.3999996185303	31.2000007629395	31.2999992370605	31.8999996185303
1	1	1	1
32	32.4000015258789	35.2000007629395	35.4000015258789
1	1	1	3
36.0999984741211	36.7999992370605	37.4000015258789	38.0999984741211
1	1	1	1
39	40.5	41.2000007629395	43.5999984741211
1	1	1	1
47.0999984741211	49.2000007629395	59.2000007629395	63.4000015258789
1	1	1	1

- **ptime86**: es el tiempo en meses que se ha pasado en prisión durante 1986

```
crime1 %>%
  with(table(ptime86)) %>%
  print()
```

```
ptime86
  0    1    2    3    4    5    6    7    8    9   10   11   12
2594  7   15   8   7   5   10   7   2   4   4   4   58
```

- **qemp86**: es la cantidad de trimestres que la persona tuvo empleo en 1986 (de cero a cuatro)