Modelos Generalizados

Andrés Martínez

27 de noviembre de 2018

Introducción

El modelo de regresión lineal que trabaja con mínimos cuadrados ordinarios, muestra que es igual de consistente con la estimación por máxima verosimiltud cuando los errores de los regresores tienen una distribución normal $N \sim (0, \sigma^2)$.La definición se obtiene del libro Applied Econometrics with R de Kleiber and Zeileis (2008).

Las propiedades de una regresión lineal se puede resumir en los siguientes tres numerales:

1. EL predictor lineal $\nu_i = x_i^T \beta$ a través de $\mu_i = E(y_i|x_i)$ que depende de $k \times 1$
2. La distribución de la variable dependiente $y_i|x_i$ es $N(\mu_i, \sigma^2)$.
3. La respuesta esperada es igual a la linea del predictor $\mu_i = \nu_i$

Por otra parte el modelo de métodos generalizados trabaja con más familias de distribución para y y con más relaciones entre $E(y_i|x_i)$. La definición de la función de densidad con la cual se trabaja en un modelo de métodos generalizados es:

$$f(y, \theta, \phi) = exp\Big\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \Big\} \tag{1}$$

Donde θ es un parámetro canónico que pertenece a la familia con la que se está trabajando y ϕ es un parámetro de dispersión. La definición de 3 pertenece a la de las familias exponenciales, en donde algunas funciones de distribucion discretas y continuas se pueden describir de forma exponencial y se pueden generalizar a partir de 3. Esto es importante para un modelo GMM porque ellos pueden trabjar con diferentes funciones de distribución además de la normal cuando es necesario y es lo que lo diferencia del método de mínimos cuadrados de la regresión lineal múltiple y simple.

Supuestos del modelo

a Elpredictor lineal $\nu_i = x_i^T \beta$ con experanza condicional $\mu = E[y_i|x_i]$ depende de un vector $k \times 1$ de x_i observaciones y de paramétros β_i.

b La distribución de la variable dependiente $y_i|x_i$ es una familia exponencial lineal.

c La respuesta esperada y le predictor lineal están relacionados por una función monotónica $g(\mu_i) = \nu_i$ llamada la función link de GLM.

Objetivo de la función link.

% Table generated by Excel2LaTeX from sheet 'Hoja1'

Table 1: Función Link

Familia	Canonical	Nombre
Binomial	$log\{\mu/(1-\mu)\}$	Logit
Gaussiano	μ	Identidad
Poisson	$log\mu$	Log

%

Ejemplo Modelo Microeconométrico

En este modelo se usará un modelo probit que hace referencia a una distribución Bernoulli que es familia de una distrbución binomial y que se puede leer también a través de una regresión logit.

El valor esperado del modelo se define asi:

$$E(y_i|x_i) = p_i = F(x_i^T \beta), \quad i = 1, ..., n \tag{2}$$

Que se obtienen a partir de la probabilidad generada por la función de distribución.

En este ejemplo se considera la participación laboral de la mujer de una muestra de 872 observaciones de suiza.

Modelo:

$$participacion = \beta_0 + \beta_1 income + \beta_2 education + \beta_3 age + \beta_4 age^2 \tag{3}$$

La participación está ne función de el ingreso, la educación, y la edad.

```
##
## Call:
## glm(formula = participation ~ . + I(age^2), family = binomial(link = "probit"),
##     data = SwissLabor)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9191  -0.9695  -0.4792   1.0209   2.4803
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.74909    1.40695   2.665  0.00771 **
## income      -0.66694    0.13196  -5.054 4.33e-07 ***
## age          2.07530    0.40544   5.119 3.08e-07 ***
## education    0.01920    0.01793   1.071  0.28428
## youngkids   -0.71449    0.10039  -7.117 1.10e-12 ***
## oldkids     -0.14698    0.05089  -2.888  0.00387 **
## foreignyes   0.71437    0.12133   5.888 3.92e-09 ***
## I(age^2)    -0.29434    0.04995  -5.893 3.79e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1203.2  on 871  degrees of freedom
## Residual deviance: 1017.2  on 864  degrees of freedom
## AIC: 1033.2
##
## Number of Fisher Scoring iterations: 4
```

Grafica del modelo

Representación gráfrica de cada variable con respecto a la participación

```
## Warning: Ignoring unknown parameters: family
```

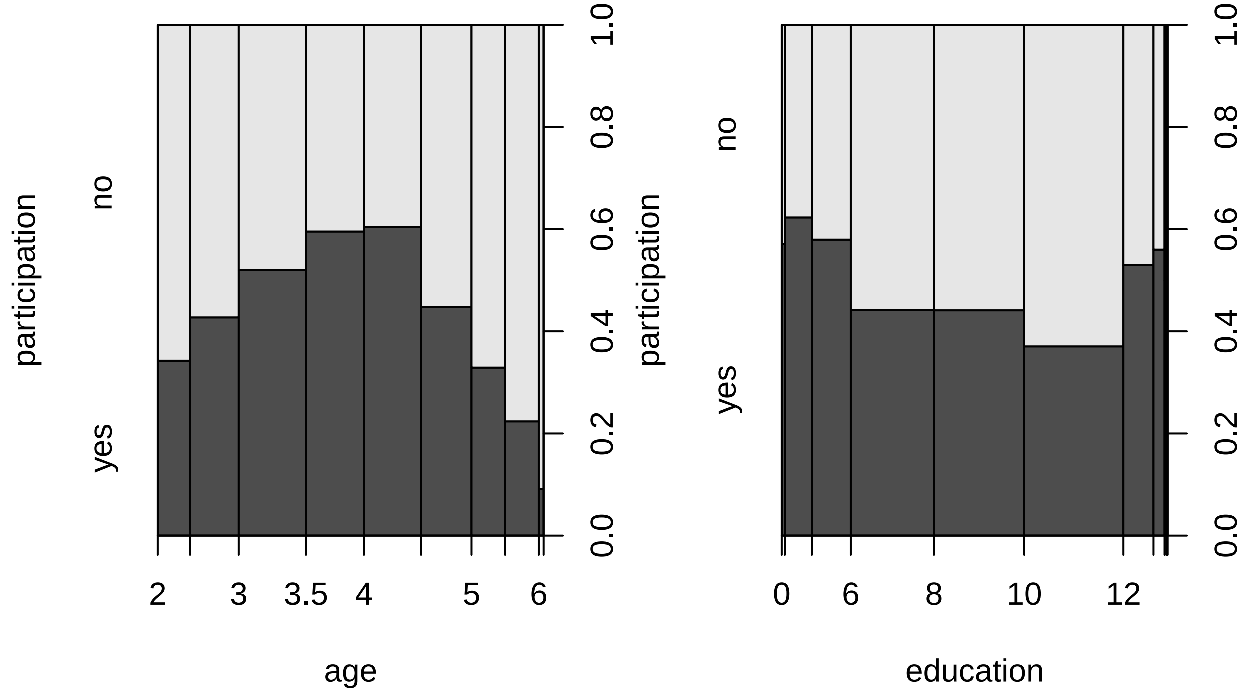

Figure 1: Spinogramas para variables dependientes binarias

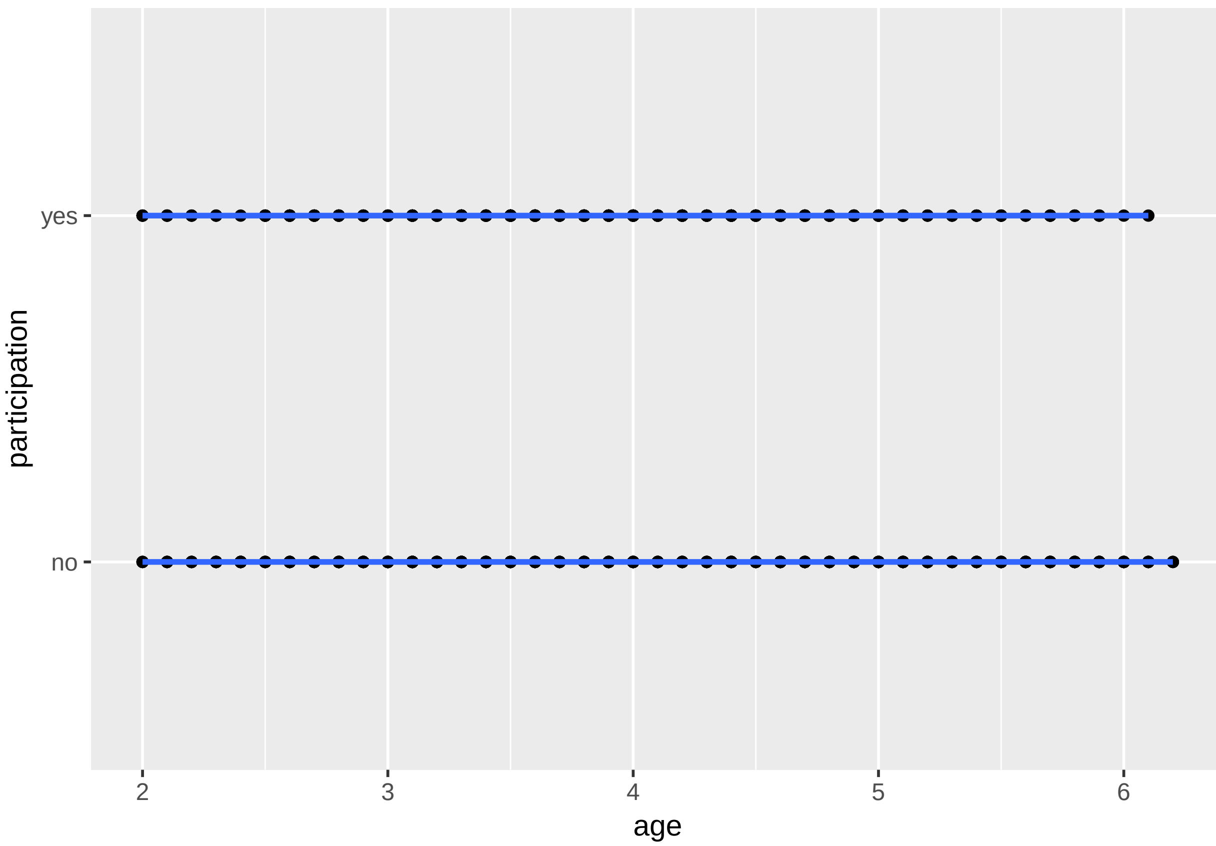

Figure 2: Edad Participación

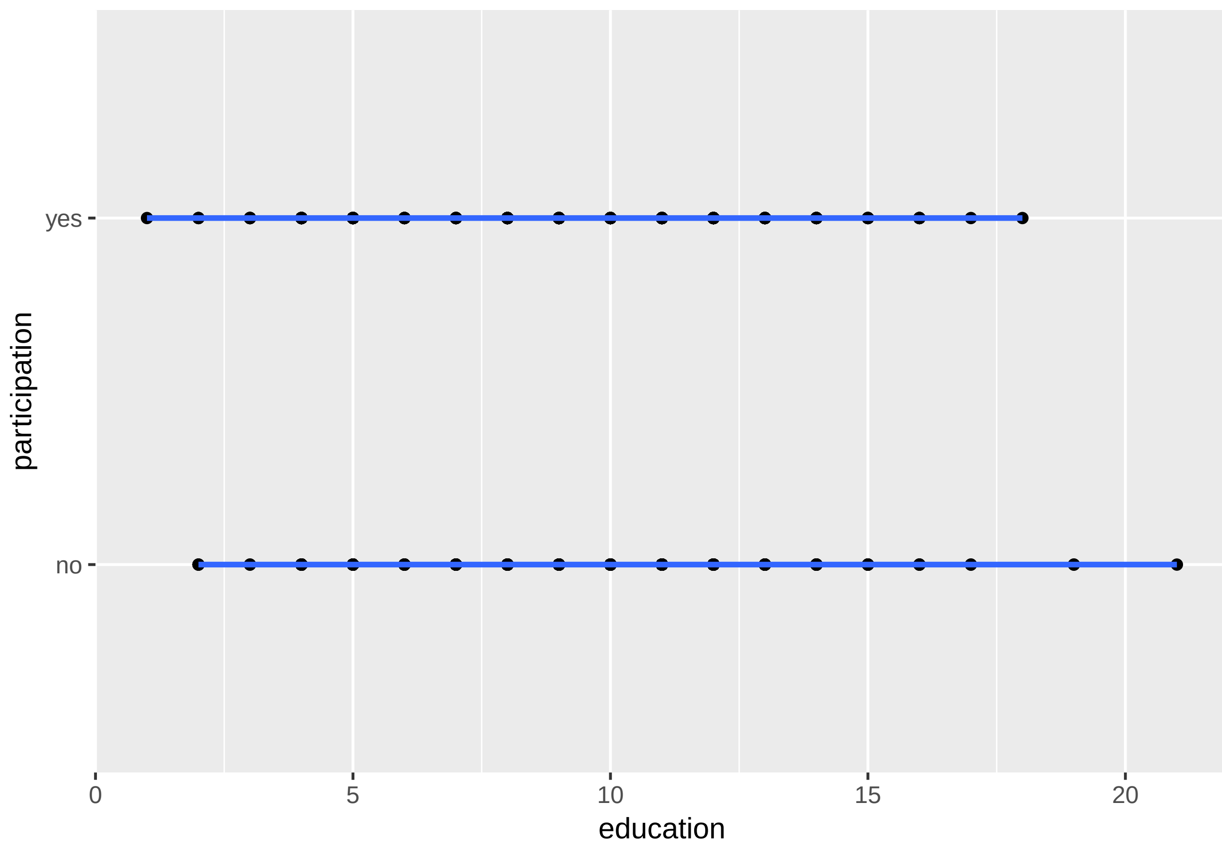

Figure 3: Educación Vs Participación

```
## Warning: Ignoring unknown parameters: family
```

Presentación de resultados usando el promedio para presentar efectos marginales para una regresión que tienen una variable dependiente probit.

```
fav <- mean(dnorm(predict(swiss_probit, type = "link")))
fav * coef(swiss_probit)
```

```
##  (Intercept)       income          age    education    youngkids
##  1.241929965 -0.220931858  0.687466185  0.006358743 -0.236682273
##      oldkids   foreignyes     I(age^2)
## -0.048690170  0.236644422 -0.097504844
```

Presentación de resultados para una característica especial, en este caso para las mujeres extranjeras de suiza.

```
av <- colMeans(SwissLabor[, -c(1, 7)])
av <- data.frame(rbind(swiss = av, foreign = av),  foreign = factor(c("no", "yes")))
av <- predict(swiss_probit, newdata = av, type = "link")
av <- dnorm(av)
av["swiss"] * coef(swiss_probit)[-7]
```

```
##  (Intercept)       income          age    education    youngkids
##  1.495137092 -0.265975880  0.827628145  0.007655177 -0.284937521
##      oldkids     I(age^2)
## -0.058617218 -0.117384323
```

```r
av["foreign"] * coef(swiss_probit)[-7]
```

```
##  (Intercept)       income          age    education    youngkids
##  1.136517140 -0.202179551  0.629115268  0.005819024 -0.216593099
##      oldkids     I(age^2)
## -0.044557434 -0.089228804
```

Goodness of ???t and prediction

- Primer método

El R^2 es poco aceptado en un modelo no lineal, sin embargo, existe una forma de obtener un seudo R^2 con ayuda de la siguiente ecuación:

$$R^2 = 1 - \frac{l(\widehat{\beta})}{l\overline{y}} \tag{4}$$

Donde $l(\widehat{\beta})$ es la verosimilitud logarítmica del modelo ajustado, y $l\overline{y}$ es la verosimilitud logarítmica de que contiene solo el término constante.

```r
 swiss_probit0 <- update(swiss_probit, formula = . ~ 1)
1 - as.vector(logLik(swiss_probit)/logLik(swiss_probit0))
```

```
## [1] 0.1546416
```

La predicción del modelo se realiza

```r
table(true = SwissLabor$participation, pred = round(fitted(swiss_probit)))
```

```
##      pred
## true    0   1
##   no  337 134
##   yes 146 255
```

- Segundo Método

Visualizar las predicciones constrastando con los datos reales es otra forma de mdeir la eficiencia del modelo. La libreria ROC permite comparar los resultados del modelo con los datos reales.

Analisis residual

- Primer método

Obteniendo la diferencia de cada uno de los valores reales con respecto a la predicción $y_i - \widehat{\mu}_i$, estos valores son los residuales y se encuentran dentro del modelo.

```r
deviance(swiss_probit)
```

```
## [1] 1017.155
```

```r
sum(residuals(swiss_probit, type = "deviance")^2)
```

```
## [1] 1017.155
```

```r
sum(residuals(swiss_probit, type = "pearson")^2)
```

```
## [1] 866.5145
```

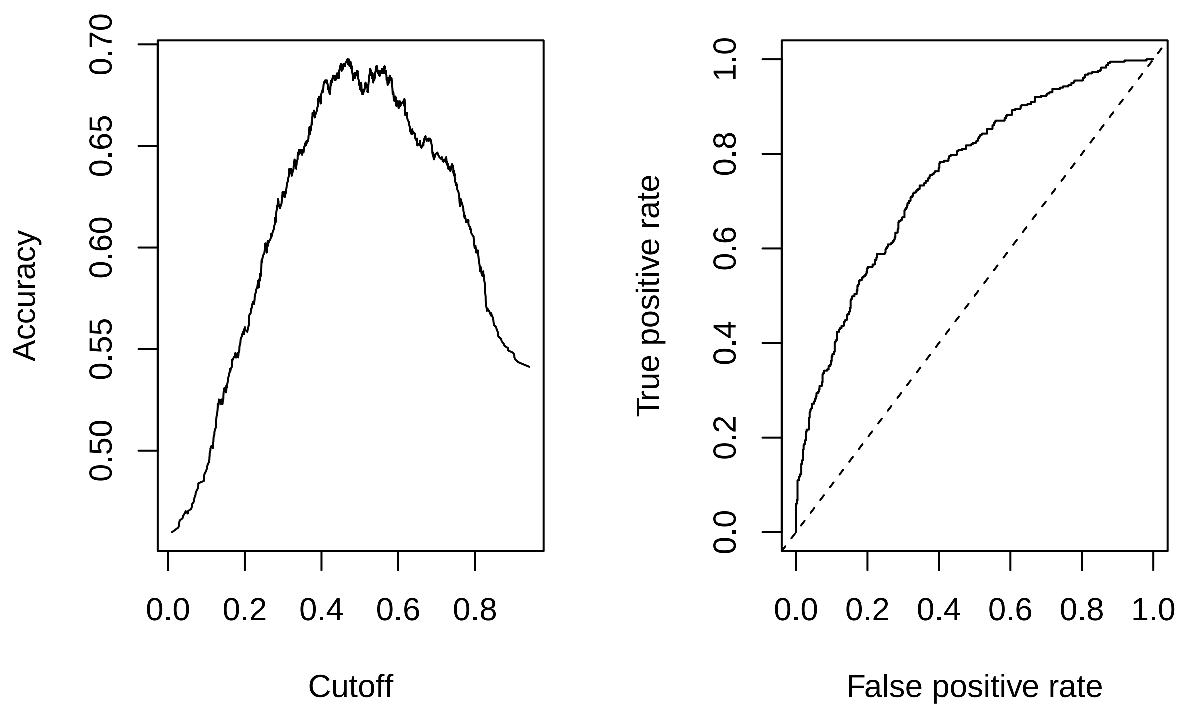

Figure 4: Comparación Gráfica

EL siguiente método es conocido como el método de covarianza o sandwich.

```
coeftest(swiss_probit, vcov = sandwich)
```

```
##
## z test of coefficients:
##
##               Estimate Std. Error z value  Pr(>|z|)
## (Intercept)  3.749091   1.327072  2.8251  0.004727 **
## income      -0.666941   0.127292 -5.2395 1.611e-07 ***
## age          2.075297   0.398580  5.2067 1.922e-07 ***
## education    0.019196   0.017935  1.0703  0.284479
## youngkids   -0.714487   0.106095 -6.7344 1.646e-11 ***
## oldkids     -0.146984   0.051609 -2.8480  0.004399 **
## foreignyes   0.714373   0.122437  5.8346 5.391e-09 ***
## I(age^2)    -0.294344   0.049527 -5.9430 2.798e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Referencias

Kleiber, Christian, and Achim Zeileis. 2008. *Applied Econometrics with R*. Springer Science & Business Media.