

# ESTIMACIÓN DE UN MODELO PROBIT Y LOGIT PARA LA EVOLUCIÓN VIRAL DE LOS PACIENTES CON VIH. CASO: HOSPITAL DE INFECTOLOGÍA “JOSÉ DANIEL RODRÍGUEZ” DE GUAYAQUIL

Leonardo Calle Hurtado<sup>1</sup>, Manuel González<sup>2</sup>

<sup>1</sup> Ingeniero en Estadística e Informática 2004

<sup>2</sup> Director de tesis, Economista, Escuela Superior Politécnica del Litoral, 1998, Postgrado Chile, Universidad de Chile, 2000. Profesor de ESPOL, 2000.]

## RESUMEN

El presente estudio analiza la propensión de los pacientes con VIH del Hospital de Infectología “José Daniel Rodríguez” de la ciudad de Guayaquil, a tener una evolución favorable luego del tratamiento correspondiente, es de decir una mejora, la cual es verificada por la revisión clínica. Como tales propensiones son, observables con precisión, construyendo una variable que desempeña el papel de endógena (MEJORA) y que podía tomar los siguientes valores: 1 si el paciente mejoró su estado de salud, y 0 en caso contrario. Mediante esta variable endógena, se estudiará aquellos factores que permiten discriminar entre no mejorar (categoría 0) y mejorar (categoría 1), a través de los modelos de regresión binaria Probit y Logit.

Palabras Claves: VIH, Probit, Logit

## INTRODUCCIÓN

El SIDA es una enfermedad infecciosa causada por un virus denominado Virus de la Inmunodeficiencia Humana (VIH o HIV) o virus del SIDA. Una enfermedad infecciosa es aquella que es causada por un germen y que puede transmitirse de una persona a otra. Un virus es una pequeña partícula infecciosa que sólo puede vivir como parásito de las células del organismo, donde se multiplica. Fuera de las células sobrevive con dificultad y durante poco tiempo.

El VIH tiene la peculiaridad de atacar a los linfocitos que son los directores de nuestro sistema de defensa. Inicialmente el VIH subsiste en estado latente, es decir, “adormecido” dentro de los linfocitos. En algunos casos, al cabo de un tiempo, a menudo años y por causas aun no bien determinadas, el virus se activa es decir “despierta” y comienza a destruir los linfocitos. De esta forma el VIH debilita progresivamente el sistema inmune,

logrando que nuestro organismo no pueda luchar adecuadamente contra diversos gérmenes. Existen varias alternativas para extender la vida de los pacientes con VIH a un promedio de 20 años de vida, que puede ser menos o más de acuerdo al sistema de vida y los cuidados que lleve el enfermo.

En general, se tomaron en cuenta las principales características personales (edad, estado civil, inclinación sexual), y la bitácora dentro de la enfermedad.

Entre un 50 y un 90 por ciento de las personas que se han infectado por el virus del SIDA experimenta síntomas similares a un catarro o una gripe leve (cansancio, fiebre, pérdida de apetito) que remite a los pocos días, también puede sufrir diarrea, sudoraciones nocturnas o aumento de los ganglios linfáticos. La única forma de saber que se ha contraído la infección es con una prueba específica realizada por un profesional sanitario. Es por esto que surgió el Plan de Ayuda para pacientes con VIH, el cual es subvencionado por el gobierno y este es el grupo a considerar para el modelo.

**Cuadro I. Codificación de Variables**

| Tipo                | Variable     | Codif.   |
|---------------------|--------------|--|
| <b>DEMOGRÁFICAS</b> | Sexo         | 0 Femenino<br>1 Masculino  |
|                     | Edad         | En años  |
|                     | Estado Civil | 0 Soltero<br>1 Casado<br>2 Viudo<br>3 Unión Libre<br>4 Separado  |
|                     | Instrucción  | 0 Primaria<br>1 Secundaria<br>2 Superior<br>3 Ninguna  |
|                     | Inclin.Sex.  | 0 Heterosexual<br>1 Homosexual<br>2 Bisexual   |
| <b>CLÍNICAS</b>     | Tratamiento  | 1 Combivir, Stocrin<br>2 Combivir, Viracept<br>3 Videx, Zerit, Stocrin<br>4 Videx, Zerit, Viracept<br>5 Combivir, Crixivan<br>6 Combivir<br>7 Stocrin<br>8 Eпивir, AZT |
|                     | Tiempo       | En meses   |
|                     | Enferm.      | 0 Universal<br>1 Primaria<br>2 Secundaria  |

|  |           |                            |
|--|-----------|----------------------------|
|  | Dif./Peso | En kilos                   |
|  | Mejora    | 0 No mejoró<br>1 Sí mejoró |

## TIPOS DE ENFERMEDADES

### Enfermedad Tipo Universal (0)

Este grupo está compuesto por las enfermedades que son típicas de los pacientes con VIH, tales como:

AFECCIÓN DERMICA  
 AFECCIÓN DE TRANSMISIÓN SEXUAL (candidiasis, criptosporidiasis)  
 AFECCIÓN ESTOMACAL  
 AFECCIÓN INTESTINAL (enterocolitis)  
 AFECCIÓN OCULAR (retinitis)  
 AFECCIÓN RESPIRATORIA (tuberculosis, neumonía)  
 AFECCIÓN VIRAL (herpes, histoplasmosis, citomegalovirus).

### Enfermedad Tipo Primaria (1)

En este tipo de enfermedad se agrupan las afecciones o síntomas que se presentan con mayor frecuencia, entre las cuales tenemos:

SINDROME FEBRIL (fiebre, dolor de cabeza)  
 SINDROME DE CONSUMO (pérdida de peso, falta de apetito, diarrea).

### Enfermedad Tipo Secundaria (2)

Aquí se agrupan las afecciones que por lo general vienen dadas como consecuencia de la pérdida de defensas del paciente, tales como:

AFECCIÓN NEURONAL  
 AFECCIÓN HEPATICA  
 AFECCIÓN CARDIACA  
 AFECCIÓN URINARIA  
 ANEMIA

## CONTENIDO

### EL MODELO

El modelo está conformado por la variable dependiente MEJORA y las variables explicativas SEXO, EDAD, ESTADO CIVIL, NIVEL DE INSTRUCCIÓN, INCLINACIÓN SEXUAL, TRATAMIENTO, TIEMPO EN TRATAMIENTO, ENFERMEDAD y DIFERENCIA DE PESO.

En el estudio de variables de naturaleza discreta, es conveniente utilizar los modelos de respuesta cualitativa, estimados usualmente por máxima verosimilitud. En este estudio se considera a cada paciente que esté en el grupo de personas infectadas con VIH del Hospital de Infectología "José Daniel Rodríguez" quienes tiene dos alternativas después del tratamiento antirretroviral (mejora clínica ó no).

De manera que:

$$y_j = \begin{cases} 1 & \text{si hay mejora clínica} \\ 0 & \text{en caso contrario} \end{cases}$$

La variable dependiente, al tomar los valores de 1 y 0, indica si este evento ocurre o no, y

$$y_j^* = \begin{cases} CV < Q_{\text{INDETECTAB}} \\ \text{existencia de mejora} \\ CV > Q_{\text{en caso contrario}} \end{cases}$$

CV: Carga Viral

Se denota como  $P_j$  la probabilidad (condicional) de ocurrencia del evento. Así, lo que el modelo de respuesta binaria intenta representar es la probabilidad  $P_j$  condicionada a un set de información  $\Omega_j$ , el cual consiste de variables explicativas y predeterminadas

$$P_j \equiv Pr(y_j = 1 | \Omega_j) = E(y_j | \Omega_j) \quad (1)$$

Se tiene un vector  $X_j$  que denote un vector de  $k$  variables que pertenecen al set de información  $\Omega_j$ , incluyendo el término constante. Entonces, un modelo de regresión lineal especificaría  $E(y_j | \Omega_j)$  como  $X_j \beta$ ,

$$y_j^* = X_j \beta + u_j \quad (2)$$

donde  $u_j \sim N(0, \sigma^2)$

Las variables contenidas en el vector  $X_j$  explican a la variable MEJORA (Y), de manera que,

$$\begin{aligned} Pr(Y=1) &= F(X, \beta) \\ Pr(Y=0) &= 1 - F(X, \beta) \end{aligned} \quad (3)$$

donde el vector de parámetros  $\beta$  refleja el impacto que  $X$  tiene sobre la probabilidad. Dado que  $E(y_j | \Omega_j)$  es una probabilidad, ésta se restringe a la distribución normal acumulada, determinando el modelo de regresión *PROBIT*.

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} X^2\right) dX \quad (4)$$

Tal especificación resuelve los problemas de utilizar el modelo de probabilidad lineal, cumpliéndose,

$$\begin{aligned} \lim_{\beta'X \rightarrow +\infty} Pr(Y=1) &= 1 \\ \lim_{\beta'X \rightarrow -\infty} Pr(Y=0) &= 0 \end{aligned} \quad (5)$$

Los modelos de elección binaria se estiman habitualmente por máxima verosimilitud. En el caso de la variable MEJORA, se tiene:

$$\ell(y, \beta) = \sum_{i=1}^n \left\{ \begin{aligned} &y_j \ln \Phi \left( \frac{X' \beta_j}{\sigma} \right) + \\ &(1 - y_j) \ln \left( 1 - \Phi \left( \frac{X' \beta_j}{\sigma} \right) \right) \end{aligned} \right\} \quad (6)$$

siendo esta ecuación la función de log-verosimilitud a utilizarse en las estimaciones de los modelos *PROBIT*, donde  $\Phi(\bullet)$  es la distribución normal acumulada.

Es importante observar que los parámetros  $\beta$  estimados, así como los de cualquier otro modelo de regresión no lineal, no son necesariamente los efectos marginales. Sin embargo, una aproximación es:

$$\frac{\partial E[y|X]}{\partial X} = \phi(X' \beta) \beta_i \quad (7)$$

siendo  $\phi(\bullet)$  la función de densidad normal estándar. La ecuación 7 se interpreta como el efecto marginal de la variable  $X$  sobre la probabilidad condicional.

En términos generales, los coeficientes de una regresión *PROBIT* son insesgados y eficientes en relación con el término de perturbación estocástica, y similares a los de las regresiones *LOGIT*.

$$\begin{aligned} \Pr(y_i = 1 | x_i, \beta) &= \\ 1 - e^{-x/\beta} / (1 + e^{-x/\beta}) & \quad (8) \\ = e^{x/\beta} / (1 + e^{x/\beta}) & \end{aligned}$$

El cual está basado en la función de distribución logística acumulada.

## RESULTADOS

Aunque cualitativamente los modelos Logit y Probit tienen resultados similares, las estimaciones de los parámetros de los dos modelos no son directamente comparables.

Los resultados de los coeficientes de la estimación de la regresión tienen

interpretación por medio del signo, mas no por la magnitud del mismo.

Luego de calcular el Modelo Logit, resultó que a excepción de las variables SOLTERO, TRAT1, TRAT3, ENF1 y DIFPESO, el resto de las variables tuvieron un efecto no significativo sobre la variable dependiente MEJORA, como se puede observar en la Tabla II.

Luego de estimar el modelo, se observa que:

Las variables SOLTERO, TRAT1 (Combivir, Stocrin), TRAT3 (Videx, Zerit y Stocrin), ENF1 (Tipo Universal) y DIFPESO son significativas al 9 %. Es decir que todas estas variables explicativas tienen un efecto significativo sobre la probabilidad de mejorar o no el estado de salud de los pacientes.

En cuanto a las interacciones, se puede señalar que el resto de tratamientos (TRAT 2, 3, 4,...,8) en conjunto con el tipo de enfermedad ENF2 ó ENF3 aumentan la probabilidad de MEJORA (1), el hecho que estas combinaciones no hayan sido significativas para el modelo se presume a que no eran representativos en relación al resto.

Y que también influyen positivamente las variables SOLTERO, TRAT1 y ENF1 en conjunto, para alcanzar una mejora en el estado de salud de los pacientes.

En cuanto al coeficiente de Mc Fadden se observa que el grado de determinación del modelo es similar en ambos casos (Probit y Logit), es decir que a pesar de que la medida de bondad de ajuste es baja (0.30), -es raro tener valores muy altos- la medida indica un aceptable grado de determinación de las variables explicativas hacia la probabilidad condicional.

**Tabla I. Resultado de Estimación de Modelo Probit sobre la Variable Dependiente MEJORA**

| Variable           | Coef.    | Error Std. | Estad. z | Prob.  |
|--------------------|----------|------------|----------|--------|
| SOLTERO            | -0.9437  | 0.4054     | -2.3278  | 0.0199 |
| TRAT1              | 0.8627   | 0.3731     | 2.3122   | 0.0208 |
| TRAT3              | 1.0895   | 0.5192     | 2.0982   | 0.0359 |
| ENF1               | 1.7990   | 0.7040     | 2.5554   | 0.0106 |
| DIFPESO            | 0.1077   | 0.0395     | 2.7271   | 0.0064 |
| TRAT1*ENF1         | -2.8073  | 1.2792     | -2.1944  | 0.0282 |
| SOLTERO*TRAT1*ENF1 | 1.9377   | 1.1815     | 1.6400   | 0.1010 |
| $R^2_{MF}$         | 0.298124 |            |          |        |

Debido a que la distribución de los datos era compleja, se decidió interpretar los resultados de la estimación del modelo Logit y no del Probit, debido a que este es habitualmente usado para estudios para estimaciones de muestras pequeñas y en donde las VI's son cualitativas y cuantitativas, como se señaló anteriormente.

**Tabla II. Resultado de Estimación del Modelo Logit sobre la Variable Dependiente MEJORA**

| Variable           | Coef.   | Error Std. | Estad. z | Prob.  |
|--------------------|---------|------------|----------|--------|
| SOLTERO            | -1.7094 | 0.7462     | -2.2906  | 0.0220 |
| TRAT1              | 1.6567  | 0.7193     | 2.3033   | 0.0213 |
| TRAT3              | 1.8216  | 0.9181     | 1.9839   | 0.0473 |
| ENF1               | 3.1558  | 1.2683     | 2.4881   | 0.0128 |
| DIFPESO            | 0.2061  | 0.0772     | 2.6701   | 0.0076 |
| TRAT1*ENF1         | -5.0910 | 2.2100     | -2.3035  | 0.0212 |
| SOLTERO*TRAT1*ENF1 | 3.3769  | 1.9764     | 1.7086   | 0.0875 |
| $R^2_{MF}$         | 0.3097  |            |          |        |

Aunque cualitativamente los modelos Logit y Probit tienen resultados similares, las estimaciones de los parámetros de los dos modelos no son directamente comparables.

## CONCLUSIONES

En este trabajo se ha estudiado el comportamiento del estado de salud de 62 pacientes con VIH luego del tratamiento antirretroviral, en su modalidad de mejora clínica, a través del modelo logit binomial. Debido a que la distribución de los datos era compleja, se decidió estimar el modelo Logit y no el Probit, debido a que este es usualmente usado para estudios para estimaciones de muestras pequeñas y en donde las VI's son cualitativas y cuantitativas, como se señaló anteriormente.

Como resultado de las estimaciones del modelo se encontró que:

I. Las características demográficas no representaban ser influyentes en la propensión a mejorar, con excepción de la variable SOLTERO que ejerce un efecto negativo en la variable MEJORA; no así las variables clínicas (TRAT1, TRAT3, ENF1, DIFPESO) que sí influían en el comportamiento individual de mejora clínica de los pacientes.

II. El bajo número de variables significativas da una noción de que la muestra, en general, es relativamente heterogénea. Es decir, no existe uniformidad en las características de los pacientes respecto al comportamiento del estado de salud frente a los tratamientos.

III. La variable DIFPESO, demostró que esta característica tiene la influencia en la propensión de mejora clínica de los pacientes, es decir que éste indicador hace que el incremento de peso tras el tratamiento aumente las posibilidades de mejorar de los pacientes observados.

IV. En general, la variable cualitativa ESTADO CIVIL indicó que las personas “no solteras” tenían mayor propensión a mejorar su estado de salud luego del tratamiento correspondiente, esto se puede presumir a que estos grupos de pacientes tengan más cuidados con la enfermedad, y cuenten con el apoyo familiar necesario para tener una alimentación adecuada.

V. Además se puede observar en la Tabla II, que las variables TRAT1 (Combivir, Stocrin) y TRAT3 (Videx, Zerit y Stocrin), tienen un nivel de significancia similar en el modelo, es decir que son los tratamientos más influyentes sobre el comportamiento de mejora de la enfermedad.

VI. También es necesario señalar que los tratamientos 2,3,4,5,6,7 y 8 asociadas con el tipo de enfermedad ENF2 ó ENF3 aumenta la probabilidad de MEJORA (1).

VII. Y que también influyen positivamente las variables SOLTERO, TRAT1 y ENF1 en conjunto, para alcanzar un cambio positivo en el estado de salud de los pacientes; es decir que el Tratamiento 1 (Combivir y Stocrin) al ser usado para casos con la Enfermedad Tipo Universal, aumentan la probabilidad de mejora del paciente de estado civil “Soltero”, resultado importante ya que éste es el grupo con mayor participación en esta variable (Ver Anexo).

4. Cuenca-Pardo J., Álvarez-Díaz C., Lepe Caamaño A., (2002), Efectividad de la Unidad de Quemados del Hospital de Traumatología “Dr. Victorio de la Fuente Narváez”. Estudio comparativo con unidades prototipo, México D.F., México.

5. Instituto Nacional de Estadísticas y Censos (INEC), (2003), “Censo Hospitalario del Ecuador”, Quito, Ecuador.

6. Martínez García M.A., Soler Cataluña J.J., Román Sánchez P., Cabero Salt L., Giménez Ibáñez I. y Gastaldo Palop T., (2003), Valor predictivo de la clínica para la identificación de los pacientes con síndrome de apneas-hipopneas durante el sueño susceptibles de tratamiento con presión positiva continua de la vía aérea (CPAP). Unidad de Neumología. Hospital General de Requena, Valencia, España.

7. Juan Manuel Malacara, (2003), Menopausia: Nuevas evidencias, Nuevos enigmas. Instituto de Investigaciones Médicas. Universidad de Guanajuato, Guanajuato, México.

## REFERENCIAS

1. Mendenhall W., Wackerly D., Scheaffer R., (1994), Estadística Matemática con Aplicaciones, Editorial Iberoamérica, 2da. Edición, México D.F., México.

2. Greene W. H., (1997), Econometric Analysis, 3ra. Edición, Prentice-Hall, Inc., U.S.A.

3. Iraurgi Castillo I., (2000), Cuestiones metodológicas en la evaluación de resultados terapéuticos, Bilbao, España.