Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias Físico- Matemáticas

EL MODELO DE COX APLICADO A UNA BASE DE DATOS FINANCIEROS

Anayeli Cocoletzi Conde, Francisco Solano Tajonar Sanabria, Hugo Adán Cruz Suárez, Fernando Velasco Luna

VII CONGRESO INTERNACIONAL DE MATEMÁTICAS Y SUS APLICACIONES

Introducción

El modelo de Cox, es un modelo cuyo enfoque principal se dirige al análisis de supervivencia, la cual a su vez es una línea de investigación de la estadística inferencial que nace de manera paralela a la teoría de la confiabilidad, en este cartel se presentan los resultados más relevantes que surgieron al aplicar el modelo de Cox en una base de datos de una financiera.

Para el análisis de supervivencia existen diferentes áreas de oportunidad, por lo que la aplicación se hizo sobre el ámbito financiero, es decir, el conjunto de datos pertenecen a datos básicos que se reúnen cuando se otorga un crédito, dentro de la base de datos se pueden encontrar datos como el monto del crédito, la edad de la persona, el sexo de la misma y el monto del seguro si este fue solicitado. De manera que la aplicación consistió en observar como las diferentes características que se pueden asociar a una persona influyen en el compromiso que la misma tiene para pagar el monto de su deuda, y además saber si existe un punto de cambio en el monto del crédito donde este se vuelve impagable.

Para desarrollar la aplicación se usaron las herramientas de la estadística inferencial así como los principales resultados del análisis de supervivencia, por lo que a continuación se abordan algunos conceptos importantes y luego se explicara brevemente como se aplico el modelo de Cox.

Conceptos Básicos del Análisis de Supervivencia

El interés del análisis de supervivencia se centra en analizar el tiempo en que falla un sistema o un articulo a partir de algún momento seleccionado. Para llevar a cabo un buen análisis de tiempos de falla se necesita lo siguiente:
•Tener bien definido el evento de interés.
•Definir de forma apropiada el origen o inicio del estudio, es decir, indicar el comienzo del análisis.
•Definir la escala de tiempo [1].

Definición: *El análisis de supervivencia es el conjunto de técnicas que permiten analizar, estudiar y modelar la variable (en este caso representa evento-tiempo) de interés.*
T será una variable aleatoria continua, no negativa que representará a una población homogénea.

Funciones asociadas a T

Función de densidad de probabilidad : es la probabilidad no condicional de que ocurra el evento de falla en el tiempo t.
Función de Supervivencia: es la probabilidad de que el objeto sobreviva más del tiempo t.
Función de Riesgo: es la probabilidad de que un individuo de *t* unidades de tiempo experimente el evento de falla en el siguiente instante de tiempo [1].

Función de densidad de probabilidad

Definición 1. Sea una variable aleatoria continua. Entonces para ***t*** existe una función denotada por ***f(t)*** , denominada función de densidad de probabilidad, que satisface las siguientes propiedades [4]

a) $f(t) \geq 0$ para toda

b) $\int_{-\infty}^{+\infty} f(t)dt = 1$

c) Para cualquier $t_1, t_2 \in (-\infty, \infty)$ tal que $t_1 \leq t_2$, tenemos que
$P[t_1 \leq T \leq t_2] = \int_{t_2}^{t_1} f(x)dx$

Función de Distribución

Definición 2. La función de distribución para la variable aleatoria (v.a.) ***T*** es la probabilidad de que un individuo falle antes del tiempo ***t***, esta función se denota por ***F(t)*** [3] y está dada por:

$$F(t) = P\,[\textit{Un individuo falle antes o hasta el tiempo } t\,]$$

$$= P[T \leq t]$$

Propiedades,

a) La función $F(t)$ es no decreciente, es decir, si $t_1 \leq t_2$ entonces $F(t_1) \leq F(t_2)$.

b) $\lim_{t \to -\infty} F(t) = 0$ y $\lim_{t \to +\infty} F(t) = 1$

c) La función $F(t)$ es continua a la derecha.

Función de Riesgo

Definición 3. La función de riesgo para una v.a. ***T*** se define como sigue: [1]

$$h(t) = \frac{1}{\Delta t} \lim_{\Delta t \to 0} P\,[\textit{Un individuo falle en } (t, t+\Delta t) \textit{ dado que ya ha sobrevivido hasta } t\,]$$

Observación 1: P[t ≤ T < t +Δ t |T ≥ t] es la probabilidad de que un individuo de ***t*** años experimente el evento de falla en las próximas ***Δt*** unidades de tiempo [2].
La función de riesgo es particularmente útil para determinar apropiadamente la distribución de falla. Utilizando información cuantitativa acerca del mecanismo de fracaso, para describir la forma en la cual la probabilidad de experimentar el evento cambia con el tiempo.
Esta tiene varias formas generales para porcentajes de riesgo distintos.
La única condición para la función de riesgo es la no negatividad es decir, ***h(t) ≥ 0.***

Modelo de Cox

En el área de supervivencia existen varios modelos que permiten conocer la relación que puede existir entre la tasa de falla de un objeto con respecto a un conjunto de variables que pueden ser medidas dentro de un periodo de observación. Uno de los modelos más conocidos por su fácil aplicación es el Modelo de Cox, el cual en su formulación más simple considera un vector de covariables constante, es decir cada una de sus entradas permanece igual durante el intervalo de tiempo que dure la observación. De manera que se tiene la siguiente definición.

Definición: Sea *Z un vector de covariables constante, el modelo de riesgos proporcionales supone que la función de riesgo esta definida por,*

$$h(t) = h_0(t)\exp(\beta_1 x_1 + \cdots, \beta_k x_k),$$

donde h(t) es la tasa de riesgo basal de cualquier individuo y la exponencial es la función que relaciona a la función de riesgo con el vector de covariables que anteriormente ya fue definido.

Una vez definido el modelo de Cox clásico, lo que continua es definir el modelo de Cox con un punto de cambio el cual considera un incremento o decremento en la í-esima covariable, que el caso de nuestra aplicación será el monto del crédito que fue solicitado, de manera que la función de riesgo se define de la siguiente manera.

Definición: *Sea T una variable aleatoria que representa el tiempo de vida y además sigue el modelo de Cox con un punto de cambio, se define a la función de riesgo asociada a T como,*

$$\lambda(t; Z) = \lambda_0 \exp\left[\{\beta + \theta I(t \leq \tau)\} Z\right],$$

Donde teta representa el cambio en la i-ésima covariable, tao representa el punto de cambio en la función de riesgo y como anteriormente **h(t)** *representa la función de riesgo basal.*

Después de definir la función de riesgo para el modelo propuesto lo que nos interesa ahora es conocer la función de supervivencia así como la función de densidad de probabilidad que son necesarias para hacer las estimaciones correspondientes a los parámetros desconocidos.

Teorema: *Sea T una variable aleatoria que representa un tiempo de falla y que además tiene una función de riesgo como se ha dado en la definición anterior, entonces la función de supervivencia asociada a T, esta dada por,*

$$\begin{aligned} S(t, Z) = &\exp\Big[-t\lambda_0 \exp[(\beta+\theta)Z]\Big] I(t \leq \tau) \\ &+ \exp\Big[-\tau\lambda_0 \exp[(\beta+\theta)Z] - (t-\tau)\lambda_0 \exp[\beta Z]\Big] I(t > \tau), \end{aligned}$$

para τ y λ_0 conocidas.

Teorema: *Sea T una variable aleatoria que representa un tiempo de falla y que además tiene una función de riesgo como se ha dado en la definición anterior, entonces la función de densidad de probabilidad asociada a T, esta dada por,*

$$\begin{aligned} f(t; Z) = &\lambda_0 \exp\{(\beta+\theta)Z\} \exp\Big(-t\lambda_0 \exp\{(\beta+\theta)Z\}\Big) I(t \leq \tau) + \\ &\lambda_0 \exp\{\beta Z\} \exp\Big(-\tau\lambda_0 \exp\{(\beta+\theta)Z\} - (t-\tau)\lambda_0 \exp\{\beta Z\}\Big) I(t > \tau), \end{aligned}$$

donde τ y λ_0 se consideran conocidas.

Después de definir las respectivas funciones asociadas a T, se utilizo el método de máxima verosimilitud para estimar los parámetros, además de realizar una regresión para observar la significancia de las características con respecto a el compromiso de pagar el crédito.
Los resultados que se obtuvieron arrojaron que no existe un punto de cambio donde el compromiso de pagar se vuelva mínimo.

Bibliografía

[1] Cox, D.C. Oakes, D. (1998). Analysis of Survival Data, Chapman & Hall.

[2] Lee, Elisa T. (2003). Statistical Methods for Survival Data Analysis, John Wiley & Sons, Inc,. Hoboken.

[3] Meyer, Paul L. (1970). Probabilidad y Estadística, Addison – Wesley Iberoamericana.

[4] Melody S. (2006),Survival Analysis with Change Point Hazart Functions, Revista Harvard University Biostatistics.

[5] Zimmermann, G. (2017), From Basic Survival Analytic Theory to a Non – Standart Application, Springer Spektrum.