UNIVERSIDAD SIMÓN BOLÍVAR
DECANATO DE ESTUDIOS PROFESIONALES
COORDINACIÓN DE MATEMÁTICAS

ADAPTACIÓN DE MODELOS DE REGRESIÓN LASSO BAYESIANO A LOS DATOS DE BROTES DE ETA EN CHILE 2017

Por
Oropeza Oropeza, Keily Marian

PROYECTO DE GRADO
Presentado ante la ilustre Universidad Simón Bolívar
como requisito parcial para optar al título de
Licenciado en Matemáticas, opción Estadística y Matemáticas Computacionales

Sartenejas, Marzo de 2020

K. OROPEZA
2020

ADAPTACIÓN DE MODELOS DE REGRESIÓN LASSO BAYESIANO A LOS DATOS DE BROTES DE ETA EN CHILE
2017

USB
LICENCIATURA EN MATEMÁTICAS

UNIVERSIDAD SIMÓN BOLÍVAR
DECANATO DE ESTUDIOS PROFESIONALES
COORDINACIÓN DE MATEMÁTICAS

ADAPTACIÓN DE MODELOS DE REGRESIÓN LASSO BAYESIANO A LOS DATOS DE BROTES DE ETA EN CHILE 2017

Por
Oropeza Oropeza, Keily Marian

Realizado con la asesoría de:

Desireé Villalta

PROYECTO DE GRADO
Presentado ante la ilustre Universidad Simón Bolívar
como requisito parcial para optar al título de
Licenciado en Matemáticas, opción Estadística y Matemáticas Computacionales

Sartenejas, Marzo de 2020

Página reservada para el acta de evaluación

DEDICATORIA

AGRADECIMIENTOS

RESUMEN

Palabras claves: Enfermedades transmitidas por alimentos,Regresión, Métodos, Modelos lineales generalizados, Lasso, Ridge.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

LISTA DE ACRÓNIMOS

ETA Enfermedades Transmitidas por Alimentos

DEIS Departamento de Estadística e Información de Salud, en Chile

OPS Organización Panamericana de la Salud

MINSAL Ministerio de Salud, en Chile

LASSO Least Absolute Shrinkage and Selection Operator

OMS Organizacion Mundial de la Salud

ECM Error Cuadrático Medio

iid independientes e idénticamente distribuidos

FDA Administración de Alimentos y Medicamentos, en Estados Unidos

BCN Biblioteca del Congreso Nacional de Chile

OLS Minimos Cuadrados Ordinarios

CDC

SAIA

FAO

INTRODUCCIÓN

CAPÍTULO I

MARCO TEÓRICO

1.1. Modelos lineales

En esta parte se explica los aspectos fundamentales de los modelos lineas, algunos casos donde no se pueden usar y las condiciones que cumplen los modelos lineales generalizados.

Durante el proceso de modelaje de una variable respuesta Y de tamaño n x 1 dependiente de otra variable X que es la matriz de diseño de n x p, donde la fila i-ésima representa las observaciones del i-ésimo individuo en las p variables explicativas y la columna j-ésima representa las observaciones de la j-ésima variable en los n individuos; se utilizan modelos de regresión que tienen como fin definir dicha dependencia y a partir del modelo obtenido hacer predicciones cuando se tiene que X es conocida. Una forma de regresión es la lineal que establece la siguiente relación entre las variables:

$$Y = \beta X + \epsilon,$$

donde β es el parámetro a encontrar y ϵ es el vector n x 1 de errores aleatorios independientes e idénticamente distribuidos (iid), obtenidos de la aproximación por una recta usando el parámetro β (Allasia, 2016)).

Existen algunos tipos de variable que no se pueden trabajar con modelos lineales, sino que se trabajan como modelos lineales generalizados (MLG). Por ejemplo:Las variables de conteo de casos,las variables de conteo de casos expresados como proporciones y las variables establecidas como binaria.

En estos casos el interés recae en variables cuya distribución pertenece a la familia exponencial.

1.1.1. Familia exponencial

Se condidera una variable aleatoria Y cuya función de probabilidad o distribución, $p(Y|\theta)$ depende de un único parámetro θ. Dicha función de probabilidad pertenece a la familia exponencial con r parámetros si $p(Y|\theta)$ puede escribirse como:

$$p(Y|\theta) = a(Y) * exp\left\{\sum_{j=1}^{r} U_j(Y)\phi_j(\theta) + b(\theta)\right\},$$

donde $Y \in \mathcal{Y} \sqsubset \Re^n$ y $\mathcal{Y}$ no depende de θ (Migon,1999).

1.1.2. Modelos lineales generalizados

Este modelo es definido en termino de un conjunto de variables aleatorias independientes $Y_1, ..., Y_N$, cada una con una distribución perteneciente a la familia exponencial y que cumple las siguientes propiedades:

1. La distribución de cada Y_i está en forma canónica y depende de un único parámetro θ_i (los θ_i no todos tienen que ser iguales).

2. Un conjunto de parámetros
$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}$$
y una variable explicativa X
$$X = \begin{bmatrix} X_1^t \\ \vdots \\ X_N^t \end{bmatrix}.$$

3. Una función monótona, llamada función de vínculo g, tal que $g(\mu_i) = X_i^t\beta$ donde $\mu_i = E[Y_i]$.

1.2. Regresión por mínimos cuadrados ordinarios

En esta sección abordaremos el problema de la estimación del vector de parámetros desconocido del modelo, β.

La técnica de mínimos cuadrados ordinarios (OLS) es el método mas común usado para estimar los parámetros desconocidos en un modelo de regresión lineal para minimizar algún residual de cuadrados.

Para este método escogeremos como el estimador de β, el resultado obtenido de resolver un problema de programación cuadrática sin restricciones:

$$\hat{\beta}_{OLS} = argmin_{\beta} \left\{ \frac{1}{n} \|(Y - X\beta)\|_2^2 \right\}. \tag{1.1}$$

Esta técnica presenta un buen comportamiento en el caso de que todos los parámetros del modelo sean significativos. Por tanto, existen dos razones por las que el método de mínimos cuadrados podría no ser adecuado para estimar modelos con variables no relevantes.

1. Precisión de la predicción: Las estimaciones de los parámetros por mínimos cuadrados tienen bajo sesgo pero gran varianza. Por lo que la precisión de la predicción a veces se puede mejorar mediante la reducción o ajuste a cero de algunos coeficientes.
2. Interpretación: En el caso donde se tiene un gran número de predictores o variables explicativas suele ser de interés determinar un subconjunto más pequeño que muestre los efectos más fuertes.

De aquí que se busque otros métodos que permitan resolver las limitantes del método de mínimos cuadrados ordinarios.

1.3. Seleccion de variables

En esta parte se explica los motivos por los cuales se da la necesidad de seleccionar variables en un modelo de regresión.

En muchas situaciones cuando se estudia variables dependientes se dispone de un conjunto grande de posibles variables explicativas, por tal razón surge la pregunta sobre si todas las variables deben entrar en el modelo de regresión, y en caso contrario, saber que variables deben entrar y cuales no.

Los métodos de selección de variables se encargan del problema de construir y seleccionar el modelo. En general se tiene dos situaciones al resolver este problema, la primera se da cuando se desea incluir cada vez mas variables en un modelo de regresión, dando como resultado que el ajuste a los datos mejore, aumentando la cantidad de parámetros a estimar pero disminuyendo su precisión individual (desarrollando mayor varianza) y por lo tanto en la función de regresión estimada, se produce un sobreajuste. Por el contrario, si se desea incluyen menos variables de las necesarias en el modelo, las varianzas se reducen pero los sesgos aumentarían obteniéndose una mala descripción de los datos.

Por otra parte, la segunda situación que se suele presentar aveces se tiene cuando algunas variables predictoras pueden perjudicar la confiabilidad del modelo, esto ocurre especialmente si están correlacionadas con otras. De esta manera, el objetivo de los métodos de selección de variables es buscar un modelo que resuelva estas situaciones, es decir, que se ajuste bien a los datos y que a la vez sea posible tenga un equilibrio entre bondad de ajuste y sencillez.

Cuando vemos el caso del método de mínimos cuadrados ordinarios, como se mostró anteriormente tiene un par de razones por las cuales podría no ser adecuado.

Por estas razones se tiene que al estimar los parámetros de regresión por el método de mínimos cuadrados ordinarios, puede que ocurra que alguna de estas estimaciones sean casi cero y por tanto la variable correspondiente a dicho coeficiente tendría muy poca influencia en el modelo, sin embargo, es poco usual que estas estimaciones sean exactamente el valor cero, por tanto, este método no nos sirve para seleccionar variables. De este modo, se necesita de otros métodos para lograr tal objetivo.

Uno de los métodos que se desarrollo para resolver este inconveniente al seleccionar variables en el OLS, fue el métodos de mínimos cuadrados penalizados, el cual se basan en los mínimos cuadrados ordinarios pero añadiendo una penalización en la función objetivo, para forzar que alguna componente del vector de parámetros β sea cero y de esta manera conseguir estimación de los parámetros y selección de variables conjuntamente.

1.4. Métodos de mínimos cuadrados penalizados

En esta parte se explica acerca de los métodos de regresión basados en mínimos cuadrados penalizados, en particular del método LASSO desde el punto de vista de la estadística clásica y la estadística bayesiana, ademas del método Ridge.

Algunos métodos pueden resultar inestables o directamente son inaplicables cuando el número de variables p es similar o incluso superior al número de observaciones n. Una alternativa para estos son los métodos de regresión penalizada. La idea principal es la penalización, para así lograr evitar el sobreajuste debido al gran número de variables explicativas, se imponine una penalización o término de penalización, que obligaría que alguna componente del vector de parámetros
β sea cero.

La elección del parámetro de penalización es de gran importancia: Es necesario un procedimiento que estime el valor de dicho parámetro a partir de los datos. Por tanto, en un intento de seleccionar las variables y de estimar los parámetros de forma automática y simultánea, se propone un enfoque unificado a través de mínimos cuadrados penalizados, que consiste en estimar el vector de parámetros, minimizando la siguiente expresión:

$$\sum_{i=1}^{n}(Y_i - X_i^t\beta)^2 + \lambda\sum_{j=1}^{p} P_\lambda(|\beta_j|), \tag{1.2}$$

donde P_λ es la función de penalización que será diferente para cada método y λ es el parámetro de penalización. Ademas, el parámetro λ debe ser elegido a través de algún procedimiento basado en los datos muestrales.

De este modo, se estima el vector de parámetros β como aquel que minimiza la expresión (1.2) y se denotará como $\hat{b_n}$. Naturalmente si $\lambda = 0$ este estimador se corresponde con el estimador de mínimos cuadrados ordinarios, que es denotado por $\hat{\beta}_{OLS}$.

A continuación se proponen tres condiciones deseables que un método de penalización debería cumplir:

1. Esparsidad: Realizar selección de variables automáticamente, es decir, tener la capacidad de

   fijar coeficientes a cero.

2. Continuidad: Ser continuo en los datos para evitar inestabilidad en la predicción.

3. Insesgadez: Tener bajo sesgo, especialmente para valores grandes de los coeficientes β_j .

A continuación se describen 2 métodos de regresión penalizada. El método LASSO y el métodos Ridge, los cuales se diferencian en el tipo de penalización (P_λ) utilizada.

1.4.1. Método LASSO

En 1996 es cuando se introduce un nuevo método de análisis de regresión llamado Least Absolute Shrinkage and Selection Operator (LASSO) basado en el método de mínimos cuadrados utilizando penalización (Tibshirani, 1996).

El LASSO en la estadística clásica desarrollado por Robert Tisbshirani es una técnica de regresión y selección de variables por mínimos cuadrados penalizados basado en el método de Garrote no negativo de Leo Breiman. LASSO considera restricciones con la norma 1 ($\mathcal{L}_1$) en los coeficientes de regresión estimados.

Así siendo β el vector de coeficientes, el método LASSO lo estima minimizando el siguiente problema de mínimos cuadrados penalizados

$$\begin{array}{c} \text{mín}_\beta \left\{ \frac{1}{n} \|(Y - X\beta)\|_2^2 \right\} \\ Sujeto\ a : \|\beta\|_1 \leq t \end{array}, \tag{1.3}$$

siendo t el parámetro de regularización o de penalización. De aquí se tiene que el estimador obtenido al resolver el problema dado en la ecuación (1.3), correspondiente a LASSO clásico es:

$$\hat{\beta}_{LASSO} = argmin_\beta \{(\tilde{y} - \beta X)'(\tilde{y} - \beta X) + \lambda \|\beta\|_1\}, \tag{1.4}$$

donde, $\tilde{Y} = Y - \bar{y}1_n$ y $\lambda \geq 0$ el cual determina la influencia de la penalización en la estimación.

Para valores grandes de λ o valores pequeños de t, los coeficientes β_j se contraen hacia cero y alguno se anula, por eso se dice que LASSO produce estimación de parámetros y selección de variables simultáneamente.

1.4.2. LASSO bayesiano

En 2008, se estudió una variación de este método con enfoque en la estadística bayesiana, el modelo denominado LASSO bayesiano, es aquel el cual usa una distribución a priori doble-exponencial (Park y Casella, 2008). Quedando el estimador para el método de LASSO bayesiano de la siguiente manera:

$$\hat{\beta}_{LASSO} = argmax_{\beta} \left\{ f(\beta|Y, \sigma^2, \lambda) \right\}, \tag{1.5}$$

donde, σ^2 y λ son parámetros de la doble-exponencial. El estimador $\hat{\beta}_{LASSO}$ puede pensarse como la moda de la distribución a posteriori de β. Además, la priori se representa por

$$\beta \sim f(\beta|\sigma^2), \tag{1.6}$$

ademas se considera una priori marginal no informativa $f(\sigma^2) \sim \frac{1}{\sigma^2}$ para σ^2 o cualquier distribución gamma inversa, para mantener la conjugación (Park y Casella, 2008).

1.4.3. Ridge

El método de Ridge fue propuesto originalmente por Hoerl y Kennard en 1970, como un método para eludir los efectos adversos del problema de colinealidad en un modelo lineal estimado por mínimos cuadrados, en el contexto $p < n$, esta basado también en el método de mínimos cuadrados penalizados, donde la función de penalización presente en la ecuación (1.2), es $P_{\lambda}(|\beta_j|) = \beta_j^2$ con los β_j estandarizados, quedando el problema como:

$$\sum_{i=1}^{n}(Y_i - X_i^t\beta)^2 + \lambda\sum_{j=1}^{p}\beta_j^2, \tag{1.7}$$

donde $\lambda \geq 0$ es el parámetro de contracción que se determinará por separado.

En general, regresión Ridge produce predicciones más precisas que OLS y selección de variables.

Uno de los inconvenientes de este método es que contrae todos los coeficientes hacia cero, pero sin conseguir la nulidad de ninguno de ellos. Por tanto, no se produce selección de variables, permaneciendo en el modelo todas las variables.

1.5. Métodos para seleccionar el parámetro de penalización λ

En esta sección explicaremos los métodos para escoger el parámetro de penalización λ en los métodos de mínimos cuadrados penalizados.

Como puede observarse todas las técnicas de mínimos cuadrados penalizados dependen de un parámetro de penalización λ, que controla la importancia dada a dicha penalización en el proceso de optimización. Ademas, cuanto mayor es λ mayor es la penalización en los coeficientes de β de la regresión y más son contraídos estos hacia cero. Note también que si $\lambda = 0$ la estimación coincide con la de mínimos cuadrados ordinarios.

Algunos de los métodos para seleccionar el parámetro λ son los siguientes:

1. Una propuesta inicial y que continúa siendo sugerida por algunos autores es la utilización de una traza Ridge para determinar λ.Esto consiste en gráficar simultáneamente los coeficientes de regresión estimados en función de λ, y elegir el valor más pequeño del parámetro para el cuál se estabilizan dichos coeficientes.

2. Un método más automático, pero intensivo computacionalmente, consiste en estimar λ mediante validación cruzada.

1.5.1. CV (Validación cruzada):

Uno de los métodos más utilizados para estimar el parámetro λ es el método k-fold cross-validation.

El método de validación cruzada consiste en dividir el modelo en un set de entrenamiento (training set) para ajustar un modelo y un set de prueba (test set) para evaluar su capacidad predictiva, mediante el error de predicción u otra medida.

La forma en que se aplica la validación cruzada es mediante la división del conjunto de datos disponibles de manera aleatoria en k subconjuntos o pliegues de igual tamaño y mutuamente excluyentes.

Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado.De aqui que el valor del parámetro será el que de el mínimo error. Este método es muy preciso puesto que evaluamos a partir de K combinaciones de datos de entrenamiento y de prueba, pero aun así tiene una desventaja, y es que es lento desde el punto de vista computacional.

1.6. Medidas de bondad de ajuste

En esta sección se explican los métodos utilizados para comparar los ajustes de modelos obtenidos a partir de los métodos de regresión antes expuestos.

1.6.1. Error Cuadrático Medio con Validación Cruzada Usando K Grupos

El error cuadrático medio (ECM) es una forma de evaluar la diferencia entre un estimador y el valor real de la cantidad que se quiere calcular. El ECM mide el promedio del cuadrado del “error”, siendo el error el valor en la que el estimador difiere de la cantidad a ser estimada. En otras palabras, se esta construyendo es estimador muestral de $E((Y - X\beta)^2)$ como

$$ECM = \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{N}, \tag{1.8}$$

donde N es la cantidad de datos, $Y \hat{-} i$ son los valores estimados y Y_i los valores reales. Para obtener valores reales se procedió a realizar una validación cruzada, donde se definieron grupos con el 50 % de datos con los cuales realizar la predicción y grupos con el 50 % de los datos con los cuales realizar la comparación para los cálculos realizados por cada brote.

Mientras más pequeña sea esta medida de error, mejor es el ajuste del modelo.

1.6.2. Coeficiente de determinación R^2

El Coeficiente de Determinación R^2 da la proporción de variación de la variable y que es explicada por la variable X (variable predictora o explicativa). Si la proporción es igual a 0, significa que la variable predictora no tiene ninguna capacidad predictiva de la variable respuesta (Y). Cuanto mayor sea R^2, mejor sería la predicción. Si llegara a ser igual a 1 la variable predictora explicaría perfectamente la variación de Y, y las predicciones no tendrían error.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \bar{Y}_i)^2}, \tag{1.9}$$

donde $\hat{Y}_i$ son los valores estimados, Y_i son los valores reales y $\bar{Y}_i$ es el promedio de los Y_i

Las medidas anteriores indican la bondad del ajuste sobre los propios elementos, pero no dan información sobre la bondad del ajuste para una observación diferente de las de la muestra.

CAPÍTULO II

MARCO METODOLÓGICO

2.1. Los Datos

Los datos utilizados en esta investigación tratan acerca de los Brotes de ETA presentados en Chile durante 2017. Esta base de datos fue extraída de la pagina del Gobierno de Chile. La misma están compuestas de 133 variables obtenidas a través del sistema de vigilancia de ETA desarrollado en el país, de las cuales se considera solo las primeras 5 semanas de 2017 y las siguientes 57 de las 133 variables:

- Variables consideradas en el estudio

  Semana estadistica:] Comprende las semanas de la 1 a la 5.

  Periodo de Incubación Días: Días que tarda en presentarse los síntomas.

  Duración Brote Días: Periodo de duración del brote.

  Region de notificación: Comprende la región donde se notifico el brote de las 15 regiones del país.

  Región de consumo: Comprende la región donde dió el consumo del alimento sospechoso de originar el brote de las 15 regiones del país.

  Región de notificación es la misma que la de consumo: Variable lógica sobre la comparación entre las regiones de consumo y notificación.

  Expuestos: Cantidad de expuestos al brote.

  Enfermos: Cantidad de enfermos por el brote.

  Tasa Ataque: Tasa en la que los expuestos enfermaron.

  mediana enfermos: Mediana para los enfermos

  Total ambulatorios: Cantidad de personas enfermas atendidas en ambulatorios.

Total amb k: Cantidad de personas enfermas atendidas en ambulatorios en los rangos de edad k: menor a 1 año, entre 1 y 4, entre 5 y 14, entre 15 y 44, entre 45 y 64, y mayores de 65.

total de Hospitalizados: Cantidad de personas enfermas atendidas en hospitales.

total de Hosp k: Cantidad de personas enfermas atendidas en hospitales en los rangos de edad k: entre 1 y 4, entre 5 y 14, entre 15 y 44, entre 45 y 64, y mayores de 65.

total sin atención: Cantidad de personas enfermas que no recibieron atención medica.

total S medica k: Cantidad de personas enfermas que no recibieron atención medica en los rangos de edad k: entre 1 y 4, entre 5 y 14, entre 15 y 44, entre 45 y 64, y mayores de 65.

Nauseas; Vomitos Diarrea; Dolores Cólicos; Abdominales; Heces Sanguinolentas; Par Variable lógica sobre la Presencia de cada sintoma.

Grupo Alimento Sospechoso: Comprende los 14 grupos de contaminación.

local de elaboración: Comprende 5 tipos de locales de elaboración de alimentos.

local consumo: Comprende 5 tipos de locales de consumo de alimentos.

Factor contribuyente de Contaminación: Comprende 8 tipos de factores de contaminación de alimentos.

Factor contribuyente de Supervivencia: Comprende 5 tipos de factores de supervivencia de los contaminantes de los alimentos.

Factor contribuyente de Proliferación: Comprende 8 tipos de factores de Proliferación de los contaminantes de los alimentos.

Proceso de pérdida de inocuidad: Comprende 8 tipos de procesos de perdida de inocuidad de los alimentos.

Lugar Perdida Inocuidad: Comprende 3 tipos de lugares donde se dió la perdida de inocuidad de los alimentos.

código CIE-10: Comprende 32 tipos de diagnosticos de enfermedades.

Tipo de Diagnostico: Comprende 3 tipos de diagnosticos alcanzados.

conclusión del brote: Variable lógica sobre la Presencia de un brote de ETA.

Contempla Inspección: Variable lógica sobre la necesidad de inspección de un brote de ETA.

2.2. Análisis Exploratorio

CAPÍTULO III

ANÁLISIS DE LOS RESULTADOS

En el siguiente capítulo se analizarán y compararán los resultados obtenidos de los tres modelos propuestos anteriormente con respecto a información real.

3.1. Error Cuadrático Medio con Validación Cruzada Usando K Grupos

Para comparar el ajuste de los modelos con respecto a valores reales se procedió de calcular el Error Cuadrático Medio (ECM) con Validación Cruzada usando K Grupos de la siguiente manera:

$$ECM = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{3.1}$$

donde n es la cantidad de datos, $\hat{y}$ son los valores estimados y y los valores reales. Para obtener valores reales se procedió a realizar una validación cruzada, donde se definieron grupos con el 85 % de datos con los cuales realizar la predicción y grupos con el 15 % de los datos con los cuales realizar la comparación para los cálculos realizados por cada mes.

Mientras más pequeña sea esta medida de error, mejor es el ajuste del modelo. Los resultados obtenidos por cada modelo mes a mes pueden apreciarse en la Tabla 3.1, mientras que el ECM total de cada modelo se encuentra en al Tabla 3.2.

Al analizar el ECM de los modelos mes a mes se puede observar que el método de Kriging para los meses enero, marzo y mayo obtuvieron el menor resultado. En contraste, para los meses de julio y septiembre el método de Polinomios de Mínimos Cuadrados Ordinarios obtuvo el menor resultado. El método de funciones de Spline fue el que obtuvo los mayores resultados para 4 meses, exceptuando septiembre que lo obtuvo el método de Kriging.

Tabla 3.1: Error Cuadrático Medio de los Modelos mes a mes

Mes	Kriging	Spline	OLSP
Enero	12942.61	40167.58	14019.72
Marzo	1819.733	3402.114	2117.082
Mayo	954.0305	34386.25	1510.127
Julio	1437.514	47705.22	1310.573
Septiembre	48167.71	9492.748	1505.743

Tabla 3.2: Error Cuadrático Medio de los Modelos

Mes	Kriging	Spline	OLSP
Total	5335.489	27154.76	4092.65

Para el total del ECM de cada método, el método de **OLSP!** (**OLSP!**) obtuvo el menor resultado, mientras que el método de funciones Spline obtuvo el mayor resultado.

CONCLUSIONES

CAPÍTULO IV

REFERENCIAS

1. Hoerl, A.E. y Kennard, R.W. (1970). Ridge regression: biased estimation for non-orthogonal problems. Technometrics, 12(1): págs. 55-67.

REFERENCIAS