Regresión lineal

(ver 2019-2020)

Emilio Letón y Elisa M. Molanes-López

3-Nov-2019

Página legal

REGRESIÓN LINEAL

Madrid, versión 2019-2020

ISBN electrónico: xxx.
Edición digital (epub): xxx.

Prefacio

Este material está diseñado con el paradigma del grupo de innovación docente miniXmodular de generar material en formato mini y modular, según se puede ver en la página web www.minixmodular.ia.uned.es, donde se introduce, entre otros, el concepto de mini-libro electrónico modular.

En este mini-libro se verán algunos conceptos básicos relativos a la regresión lineal que está indicada en la modelización de una variable respuesta Y (también llamada dependiente, explicada, "output" o resultado) continua a través de varias variables explicativas $X_1, X_2, \cdots, X_p$ (también llamadas independientes, predictoras, características, "input" o entradas).

En el mini-capítulo 1 se especifican los métodos de estimación puntual del vector de parámetros poblacional $\boldsymbol{\beta}$ que interviene en el modelo de regresión lineal. Conviene mencionar que a lo largo de este mini-libro se adoptará el convenio clásico de denotar a un vector aleatorio y al valor que toma dicho vector aleatorio con la misma letra, indicando explícitamente la diferencia sólo en el caso de ambigüedad.

En este mini-libro no se distinguirá el caso con $p = 1$, correspondiente a la regresión lineal simple, del caso con $p > 1$, correspondiente a la regresión lineal múltiple. No obstante, se recomienda a los lectores que se adentren por primera vez en la regresión que consulten el capítulo 3 de [1] para una primera toma de contacto con la regresión lineal simple.

También se recomienda refrescar los conceptos básicos de esperanza y varianza de una variable aleatoria (v.a.) en, por ejemplo, [2].

Mini-capítulo 1

Estimación puntual de los parámetros

Desde un punto de vista poblacional, el modelo de regresión lineal establece para cada individuo i, que

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \mathcal{E}_i$$

donde $\mathcal{E}_i$ son v.a. i.i.d $\sim N(0, \sigma)$, siendo $i = 1, \ldots, n$, con n el número de individuos. Una vez observados los datos, el modelo anterior se traduce, desde un punto de vista muestral, en la siguiente expresión matricial

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}$$

o, equivalentemente,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

donde $\mathbf{y} \in \mathfrak{M}_{n\mathrm{x}1}$, $\mathbf{X} \in \mathfrak{M}_{n\mathrm{x}(p+1)}$, $\boldsymbol{\beta} \in \mathfrak{M}_{p\mathrm{x}1}$ y $\varepsilon \in \mathfrak{M}_{n\mathrm{x}1}$. Por razones que se verán más adelante, es importante señalar que es imprescindible que haya más datos que variables explicativas ($n > p$) y que el rango de $\mathbf{X}$ sea completo ($rg(\mathbf{X}) = p + 1$).

Existen dos métodos para estimar el vector poblacional de parámetros $\boldsymbol{\beta}$: el método de mínimos cuadrados y el método de máxima verosimilitud.

1.1. Mínimos cuadrados

La obtención de los estimadores mínimo cuadráticos de los parámetros $\boldsymbol{\beta}$, se hace minimizando la suma de los cuadrados de los residuos, dada por $SCR(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}))^2$ y que en términos matriciales se puede escribir como $\mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$.

Para minimizar $SCR(\boldsymbol{\beta})$ se deriva la expresión anterior de $SCR(\boldsymbol{\beta})$ con respecto a β y se iguala a cero, obteniéndose la llamada ecuación normal del modelo de regresión lineal dada por $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$. Si se resuelve la ecuación normal se obtiene $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, donde la inversa de $\mathbf{X}^T\mathbf{X}$ existe por tratarse de una matriz cuadrada de rango completo (ya que $rg(\mathbf{X}^T\mathbf{X}) = rg(\mathbf{X})$ y $rg(\mathbf{X}) = p + 1$). Por último, habría que demostrar que efectivamente se trata de un mínimo estudiando la segunda derivada.

Ejemplo

1. Demostrar la fórmula de la suma de cuadrados residuales en función de los parámetros.
   SOLUCIÓN:
   La expresión inicial $SCR(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}))^2$ se sustituye por su expresión matricial y se opera, es decir

$$\begin{aligned} SCR(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y}^T - (\mathbf{X}\boldsymbol{\beta})^T)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y}^T - \boldsymbol{\beta}^T\mathbf{X}^T)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T\mathbf{y} - (\mathbf{y}^T\mathbf{X}\boldsymbol{\beta})^T - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T\mathbf{y} - (\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}^{TT}) - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}, \end{aligned}$$

donde se ha usado el hecho de que $\mathbf{y}^T\mathbf{X}\boldsymbol{\beta}$ es un número y, por tanto, coincide con su traspuesto.

2. Demostrar la obtención de la ecuación normal

   Se trata de derivar $SCR(\boldsymbol{\beta})$ con respecto a $\boldsymbol{\beta}$ e igualar a cero, es decir

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}SCR(\boldsymbol{\beta}) &= \frac{\partial}{\partial\boldsymbol{\beta}}(\mathbf{y}^T\mathbf{y}-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}+\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})\\ &= \frac{\partial}{\partial\boldsymbol{\beta}}(\mathbf{y}^T\mathbf{y})+\frac{\partial}{\partial\boldsymbol{\beta}}(-2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y})+\frac{\partial}{\partial\boldsymbol{\beta}}(\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})\\ &= 0-2\mathbf{X}^T\mathbf{y}+2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}=-2(\mathbf{X}^T\mathbf{y}-\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})=0,\end{aligned}$$

   donde se ha usado el hecho de que $\mathbf{X}^T\mathbf{X}$ es simétrica a la hora de derivar.

1.2. Máxima verosimilitud

Para un individuo i, se tiene que

$$Y_i \sim N(\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip},\sigma),$$

con

$$f(y_i)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y_i-(\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}))^2},$$

con lo que

$$\begin{aligned}f(y_1,\ldots,y_n) &= \prod_{i=1}^n f(y_i)=\prod_{i=1}^n\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y_i-(\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}))^2}\\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-(\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}))^2}=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}SCR(\boldsymbol{\beta})}.\end{aligned}$$

Dicha función de densidad constituye la función de verosimilitud cuando se considera como función de los parámetros $\beta_0,\ldots,\beta_p,\sigma^2$. Dado que el logaritmo neperiano es una función monótona creciente, maximizar la función de verosmilitud equivale a maximizar su logaritmo neperiano, es decir la función dada por

$$Ln(f(y_1,\ldots,y_n))=-\frac{n}{2}Ln(2\pi)-\frac{n}{2}Ln(\sigma^2)-\frac{1}{2\sigma^2}SCR(\boldsymbol{\beta}),$$

con lo que el vector de parámetros $\boldsymbol{\beta}$ que maximiza la función de verosimilitud es el mismo que minimiza la suma de cuadrados residuales $SCR(\boldsymbol{\beta})$.

Referencias

[1] James, G., Witten, D., Hastie, T., & Tibshirani, R., *An introduction to statistical learning*, Springer (New York), 2013.

[2] Ramos, E., Vélez, R., Hernández, V., *Modelos probabilísticos y optimización*, Sanz y Torres (Madrid), 2019.

Acerca de

Emilio Letón

Figura 1.1: Emilio Letón

Emilio Letón nace en Madrid en 1966. Es Licenciado en Matemáticas por la UCM en 1989 y doctor en Matemáticas en la misma universidad en 2002. En la actualidad es profesor contratado doctor de la UNED en el departamento de Inteligencia Artificial, al que se incorporó en 2009. Anteriormente fue profesor del departamento de Estadística de la UC3M durante 5 años. Asimismo, ha trabajado durante 15 años en departamentos de Planificación y Estadística dentro del sector bancario y de la industria farmacéutica. Sus líneas de investigación incluyen el Análisis de Supervivencia, tests no paramétricos, PLS, Meta-Análisis, Bioestadística y B-Learning. Ha participado en más de 30 proyectos de innovación docente (siendo coordinador en más de 10 de ellos) colaborando con distintas universidades: UNED, UC3M, UCM y UPM. Ha recibido 1 premio en excelencia en publicaciones científicas (UC3M) y 5 premios en excelencia docente (1 en UC3M, 3 en UNED y 1 en OCW Consortium). En @emilioleton se pueden encontrar sus tweets y su página web personal con información ampliada de su curriculum. https://twitter.com/emilioleton

Elisa M. Molanes-López

Figura 1.2: Elisa M. Molanes-López

Elisa M. Molanes López nace en Vigo en 1976. Se licencia en Matemáticas por la Universidad de Santiago de Compostela en 2000 y consigue el grado de doctora, con acreditación europea, por la Universidad de A Coruña en 2007. Entre 2003 y 2007, al amparo de una beca FPI, realiza varias estancias de investigación en las siguientes universidades extranjeras: Universiteit Hasselt (Bélgica), Université Catholique de Louvain (Bélgica) y The University of Texas (EE.UU.). Durante 8 años, entre 2007 y 2015, es profesora visitante en el departamento de Estadística de la Universidad Carlos III de Madrid. Actualmente, desde octubre de 2015, es profesora ayudante doctora en la Unidad Departamental de Bioestadística del Departamento de Estadística e Investigación Operativa de la Universidad Complutense de Madrid. Durante su etapa docente ha participado en 10 proyectos de innovación docente. En diciembre de 2013, recibe un accésit a la mejor práctica docente en los Premios del Consejo Social de la UNED por su participación en el MOOC "Mini-vídeos docentes modulares: un elemento crítico en el diseño de un MOOC", y en abril de 2014, el OpenCourseWare Consortium le otorga un premio de excelencia por su participación en el curso OCW "Mini-vídeos docentes modulares para diseñar un MOOC". Sus líneas de investigación incluyen la estadística no paramétrica, el análisis de supervivencia, las curvas ROC y las funciones cópula. En la página web personal de Elisa M. Molanes-López, se puede encontrar información más detallada. https://elisamariamolanes.wixsite.com/emolanes

Índice general

Índice de figuras