

Índice general

1. Modelos lineales generalizados	3
1.1. Introducción	3
1.2. Modelos lineales	3
Bibliografía	5

Capítulo 1

Modelos lineales generalizados

1.1. Introducción

Como bien introdujimos antes, los modelos estadísticos pretenden explicar la relación entre dos o mas variables, en particular, tratan de describir el comportamiento de una variable respuesta (o dependiente), que se suele denotar por Y , mediante la información que otorgan las variables predictoras (o independientes), que se suelen denotar como X_1, \dots, X_p .

La forma más general de expresar matemáticamente la intención de los modelos estadísticos es la siguiente:

$$Y = f(X_1, \dots, X_p) + \epsilon$$

donde f es una función desconocida cuyo propósito es el de expresar de la “mejor”¹ manera posible la relación entre las variables X_1, \dots, X_p e Y , y ϵ es un error aleatorio independiente de las variables predictoras que deberá cumplir ciertas condiciones según el tipo de modelo que estemos tratando.

Este trabajo tiene como finalidad el definir un conjunto de estrategias y técnicas que proporcionan un ajuste óptimo de la función f , es decir, que asemeje lo mejor posible la relación de las variables al fenómeno que se esté estudiando. Además, se incluye una posterior aplicación práctica de dichos resultados en el ámbito del cambio climático.

En lo que a este capítulo respecta, partiremos definiendo los modelos lineales de una forma breve y más general, ya que se ve de manera más extensa en varias asignaturas durante el grado. Tras ello daremos varios resultados básicos para el entendimiento y desarrollo de los Modelos Lineales Generalizados.

1.2. Modelos lineales

El modelo lineal ocupa un lugar clave en el manual de herramientas de todo estadístico aplicado. Esto se debe a su simple estructura, a la fácil interpretación de sus resultados y al sencillo desarrollo de la teoría de mínimos cuadrados. Sin embargo, a la hora de dar su definición se deben tener en cuenta ciertas restricciones que deben cumplir las

¹Escribimos mejor entre comillas pues hay diversas formas de evaluar los modelos.

variables y los errores del modelo. Estas condiciones hacen que el modelo no sea capaz de adaptarse bien a todos los fenómenos que uno se propone describir pero, a cambio, otorga esa sencillez de visualización antes mencionada.

Definición 1.2.1 (Modelo Lineal). Sean X_1, \dots, X_p un conjunto de p vectores aleatorios de n componentes (con $p \leq n$) e $Y = (Y_1, \dots, Y_n)^T$ un vector aleatorio de p componentes tal que $E[Y] = \mu$. Entonces, se entiende por modelo lineal (multivariante) aquel que determina la relación entre los vectores aleatorios mediante una combinación lineal de parámetros de la siguiente forma:

$$\mu = X\beta$$

$$Y = \mu + \epsilon$$

O vectorialmente como:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{(1,1)} & \cdots & x_{(1,j)} & \cdots & x_{(1,p)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{(i,1)} & \cdots & x_{(i,j)} & \cdots & x_{(i,p)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{(n,1)} & \cdots & x_{(n,j)} & \cdots & x_{(n,p)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Donde: β es un vector de parámetros a determinar y reflejan la magnitud del efecto lineal (constante) de los incrementos unitarios en las variables explicativas X_i sobre la variable explicada Y . ϵ es un vector aleatorio tal que los ϵ_i son variables aleatorias independientes e idénticamente distribuidas por una distribución normal de esperanza nula y varianza σ^2 ($\epsilon_i \sim N(0, \sigma^2)$). Representa el término de error del modelo y corresponde con $\epsilon = Y - E[Y]$.

Observación 1.2.1. Gracias a lo notado en el párrafo anterior podemos ver que: $Y \sim N(\mu, \sigma^2 I_n)$ ya que:

$$E[Y] = E[X\beta + \epsilon] = X\beta + E[\epsilon] = X\beta = \mu$$

$$Cov(Y) = Cov(X\beta + \epsilon) = Cov(\epsilon) = \sigma^2 I_n$$

Veamos ahora cómo se pueden obtener los valores de los parámetros β .

Definición 1.2.2 (Estimador por mínimos cuadrados). Elegiremos los valores de β que minimicen la expresión:

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \mu_i)^2 = \sum_{i=1}^n (Y_i - (X\beta)_i)^2$$

Es decir:

$$\hat{\beta} = \arg_{\beta \in \mathbb{R}^p} \min ||Y - X\beta||^2$$

Donde $|| \cdot ||$ denota la norma euclídea. Para encontrar tal mínimo se razona derivando respecto de cada β_i y luego igualando a 0. De este modo los parámetros β_i vienen dados por la solución del siguiente sistema:

$$\begin{cases} \frac{\partial S}{\partial \beta_1} = 0 \\ \vdots \\ \frac{\partial S}{\partial \beta_p} = 0 \end{cases}$$

Desarrollando este sistema de ecuaciones llegamos a que es equivalente a $X^T X \hat{\beta} = X^T Y$, por lo que el estimador por mínimos cuadrados del vector de parámetros β viene dado por:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Observación 1.2.2. Notemos que esta expresión tiene sentido pues el producto $X^T X$ resulta en una matriz cuadrada de orden p y rango máximo.

Proposición 1.2.1 (Distribución del estimador por mínimos cuadrados). *El estimador por mínimos cuadrados del vector de parámetros β , $\hat{\beta}$, sigue una distribución del tipo normal p -variante de esperanza β y matriz de covarianzas $V_{\hat{\beta}} = \sigma^2(X^T X)^{-1}$. Es decir: $\hat{\beta} \sim N_p(\beta, \sigma^2(X^T X)^{-1})$*

Demostración. Partiremos viendo que el estimador $\hat{\beta}$ es insesgado:

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T Y] = (X^T X)^{-1} X^T E[Y] = (X^T X)^{-1} X^T X \beta = \beta$$

Por otro lado, calculemos la matriz de covarianzas de $\hat{\beta}$, para ello primero debemos notar que como los errores aleatorios ϵ_i son independientes e idénticamente distribuidos con esperanza nula y varianza σ^2 , $\forall i \neq j$:

$$E[\epsilon_i \epsilon_j] = E[\epsilon_i] + E[\epsilon_j] + Cov(\epsilon_i, \epsilon_j) = 0$$

y, por tanto:

$$E[\epsilon \epsilon^T] = E \begin{bmatrix} \epsilon_1^2 & \epsilon_1 \epsilon_2 & \cdots & \epsilon_1 \epsilon_n \\ \epsilon_1 \epsilon_2 & \epsilon_2^2 & \cdots & \epsilon_2 \epsilon_n \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_1 \epsilon_n & \cdots & \cdots & \epsilon_n^2 \end{bmatrix} = \begin{pmatrix} E[\epsilon_1^2] & E[\epsilon_1 \epsilon_2] & \cdots & E[\epsilon_1 \epsilon_n] \\ E[\epsilon_1 \epsilon_2] & E[\epsilon_2^2] & \cdots & E[\epsilon_2 \epsilon_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[\epsilon_1 \epsilon_n] & \cdots & \cdots & E[\epsilon_n^2] \end{pmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma^2 \end{bmatrix}$$

Luego, utilizando que $\hat{\beta} = \beta + (X^T X)^{-1} X^T \epsilon$, se tiene que:

$$Cov(\hat{\beta}) = E[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])^T] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = E[((X^T X)^{-1} X^T \epsilon)((X^T X)^{-1} X^T \epsilon)^T] = E[(X^T X)^{-1} X^T \epsilon \epsilon^T X (X^T X)^{-1}] = E[(X^T X)^{-1} X^T E[\epsilon \epsilon^T] X (X^T X)^{-1}] = E[(X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1}] = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

■