Regresión Lineal Múltiple: Estimación

Clase 17

Nicolás Mejía M.
n.mejia10@uniandes.edu.co

Probabilidad y Estadística II
Departamento de Ingeniería Industrial
Universidad de Los Andes, Bogotá, Colombia

2020-19

Outline

1. Remember, Remember...

2. Regresión Lineal Múltiple
   - Planteamiento
   - Estimación
   - Propiedades del estimador

Motivación

Que se logra con un modelo de regresión?

1. Explicar: Se pueden responder preguntas científicas (parecidas al diseño de experimentos) sobre la influencia de las variables independientes sobre la respuesta:
   - Aumenta el salario de un egresado obtener un mejor promedio académico?
   - Vale la pena estudiar más para el parcial?
2. Predecir: Se puede predecir el comportamiento de la variable de respuesta si se fijan los niveles de X:
   - Si mi promedio es 3.99, qué puedo esperar de mi salario?
   - Qué pasará si estudio sólamente 4 horas para el próximo parcial?
3. Precisión en Inferencia: Parecido al diseño de experimentos, al tener en cuenta el efecto de las X's, se reduce la varianza, con lo cual la inferencia acerca de Y es más precisa (con más potencia).
   - Al hacer un prueba para saber si mi salario será mayor a $2 millones, esta será mucho más exacta a respuesta si conozco mi promedio.

Planteamiento

Vamos a concentrarnos hoy en el caso en que solo se tiene una variable explicativa. Este caso se denomina el modelo de regresión lineal simple:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Donde $\epsilon \sim_{iid} N(0, \sigma^2)$. Bajo estos supuestos se considera que:

$$Y|X_1 \sim Normal\left(\beta_0 + \beta_1 X_1\ ,\ \sigma^2\right)$$

Es decir, la media de Y dado X es $E(Y|X) = \beta_0 = \beta_1 X_1$, luego:

- β_0 es el valor esperado de Y dado un valor nulo de la variable X
- β_1 es el aumento en Y, dado un aumento de 1 en X

Outline

Planteamiento

En un modelo de regresión lineal se plantea la relación:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon$$

Donde $\epsilon \sim_{iid} N(0, \sigma^2)$, el cual corresponde al error aleatorio en la observación, que es tambien homocedastico. Bajo estos supuestos se considera que:

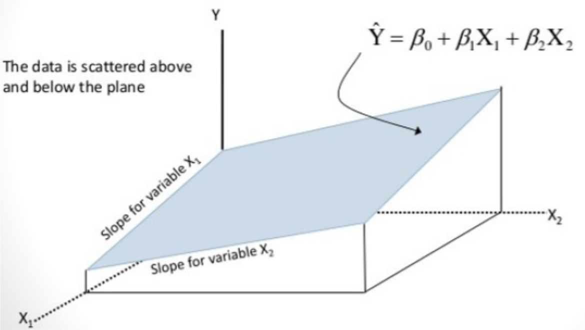
$$Y|X_1, \cdots, X_k \sim Normal\left(\beta_0 + \beta_1 X_1 + \cdots \beta_k X_k\ ,\ \sigma^2\right)$$

Es decir, la media de Y dado $X_1, \cdots, X_k$ se plantea como una relación lineal $E(Y|X_1, \cdots, X_k) = \beta_0 = \beta_1 X_1 + \cdots + \beta_k X_k$, luego:

- β_0 es el valor esperado de Y dado un valor nulo de las X's
- β_j es el aumento en Y, dado un aumento de 1 en X_j

Planteamiento

Si se tienen 2 variables explicativas, gráficamente se tendría una situación de este estilo:



El error aleatorio simplemente describe las desviaciones sobre la recta que no se pueden explicar por medio de X.

Outline

1. Remember, Remember...
2. Regresión Lineal Múltiple
   - Planteamiento
   - Estimación
   - Propiedades del estimador

Estimación

En la práctica, no tenemos conocimiento de los coeficientes $\beta_1, \beta_1, \cdots, \beta_k$ por tanto debemos estimarlos.

Dada una muestra de n datos, que corresponden a tuplas de observaciones de la variable de interés Y, y las variables explicativas $X_1, ..., X_k$:

i	Y	X_1	$\dots$	X_j	$\dots$	X_k
1	Y_1	X_{11}	$\dots$	X_{j1}	$\dots$	X_{k1}
2	Y_2	X_{12}	$\dots$	X_{j2}	$\dots$	X_{k2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	Y_i	X_{1i}	$\dots$	X_{ji}	$\dots$	X_{ki}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	Y_n	X_{1n}	$\dots$	X_{jn}	$\dots$	X_{kn}

El objetivo es tratar de recrear la relación entre ellas.

Estimación

Escribiendo el modelo para cada observación se tendría que:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \cdots + \beta_k X_{k1} + \epsilon$$
$$Y_2 = \beta_0 + \beta_1 X_{12} + \cdots + \beta_k X_{k2} + \epsilon$$
$$\vdots$$
$$Y_n = \beta_0 + \beta_1 X_{1n} + \cdots + \beta_k X_{kn} + \epsilon$$

donde $\epsilon \sim N(0, \sigma^2)$ y $cov(\epsilon_i, \epsilon_j) = 0, i \neq j$

Estimación

Lo cual en formato matricial quedaría:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

donde $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$ y $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

Estimación

Los supuestos en formato matricial quedan:

$$E(\epsilon) = \begin{bmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

$$Var(\epsilon) = \begin{bmatrix} Var(\epsilon_1) & Cov(\epsilon_1, \epsilon_2) & \cdots & Cov(\epsilon_1, \epsilon_n) \\ Cov(\epsilon_1, \epsilon_2) & Var(\epsilon_2) & \cdots & Cov(\epsilon_2, \epsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(\epsilon_1, \epsilon_n) & Cov(\epsilon_2, \epsilon_n) & \cdots & Var(\epsilon_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}$$

Estimación

Cómo se hace el ajuste? Mínimos Cuadrados Ordinarios

El hiperplano ajustado se puede denotar como $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k$, luego el error de estimación es $e_i = Y_i - \hat{Y}_i$. La idea es minimizar la suma de cuadrados del error:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)^2$$

La suma simplemente agrega el error de todas las observaciones, y el cuadrado hace desaparecer el signo para así minimimizar la magnitud del error.

Estimación

Cómo hallamos el mínimo? Derivando, igualando a 0 y despejando

Las ecuaciones normales

$$\frac{\partial SSE}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}) = 0$$

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{n} X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}) = 0$$

$$\vdots$$

$$\frac{\partial SSE}{\partial \beta_k} = -2\sum_{i=1}^{n} X_{ki}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \cdots - \hat{\beta}_k X_{ki}) = 0$$

Estimación

Reorganizando el sistema queda como:

Las ecuaciones normales

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_1 X_{1i} - \cdots - \sum_{i=1}^{n} \hat{\beta}_k X_{ki} = 0$$

$$\sum_{i=1}^{n} X_{1i} Y_i - \sum_{i=1}^{n} \hat{\beta}_0 X_{1i} - \sum_{i=1}^{n} \hat{\beta}_1 X_{1i} X_{1i} - \cdots - \sum_{i=1}^{n} \hat{\beta}_k X_{1i} X_{ki} = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ki} Y_i - \sum_{i=1}^{n} \hat{\beta}_0 X_{ki} - \sum_{i=1}^{n} \hat{\beta}_1 X_{ki} X_{1i} - \cdots - \sum_{i=1}^{n} \hat{\beta}_k X_{ki} X_{ki} = 0$$

Estimación

Reorganizando el sistema queda como:

Las ecuaciones normales

$$\hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i} + \cdots - \hat{\beta}_k \sum_{i=1}^{n} X_{ki} = \sum_{i=1}^{n} Y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_{1i} + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \cdots - \hat{\beta}_k \sum_{i=1}^{n} X_{1i} X_{ki} = \sum_{i=1}^{n} X_{1i} Y_i$$

$$\vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} X_{ki} + \hat{\beta}_1 \sum_{i=1}^{n} X_{1i} X_{ki} + \cdots - \hat{\beta}_k \sum_{i=1}^{n} X_{ki}^2 = \sum_{i=1}^{n} X_{ki} Y_i$$

Estimación

Matricialmente queda:

$$\begin{bmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \cdots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{2i}X_{1i} & \cdots & \sum_{i=1}^n X_{ki}X_{1i} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}^2 & \cdots & \sum_{i=1}^n X_{ki}X_{2i} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{1i}X_{ki} & \sum_{i=1}^n X_{2i}X_{ki} & \cdots & \sum_{i=1}^n X_{ki}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{1i}Y_i \\ \sum_{i=1}^n X_{2i} \\ \vdots \\ \sum_{i=1}^n X_{ki}Y_i \end{bmatrix}$$

Que en nuestras definiciones sería equivalente a:

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

y despejando queda que:

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

Estimación

El estimador ya no es una expresión puntual particular, debe encontrarse numéricamente! Importante que recuerden la estructura de las matrices TODO LO QUE NECESITAN ESTÁ AHÍ:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \cdots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{2i}X_{1i} & \cdots & \sum_{i=1}^n X_{ki}X_{1i} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}^2 & \cdots & \sum_{i=1}^n X_{ki}X_{2i} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{1i}X_{ki} & \sum_{i=1}^n X_{2i}X_{ki} & \cdots & \sum_{i=1}^n X_{ki}^2 \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{1i}Y_i \\ \sum_{i=1}^n X_{2i} \\ \vdots \\ \sum_{i=1}^n X_{ki}Y_i \end{bmatrix}$$

Outline

Propiedades del estimador

El estimador de mínimos cuadrados tiene la siguiente distribución multivariada (recuerde que es un vector):

$$\hat{\beta} \sim N\left(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

es decir, el estimador es Normal, insesgado y consistente.

Teorema de Gauss-Markov

El estimador de mínimos cuadrados es el estimador de varianza mínima en la clase de los estimadores lineales insesgados, cuando se satisfacen los supuestos de regresión. En palabras simples, es el estimador más eficiente.

Estimación de la varianza

Las expresiones anteriores involucran a σ^2, luego es necesario estimarlo. Para este punto ya sabemos como!

$$\hat{\sigma}^2 = MSE = \frac{SSE}{gl_E}$$

Ya vimos que el SSE se define como:

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)^2$$

Luego el total de grados de libertad es $glE = n - k - 1 = n - \#\beta etas$.