Regresión Lineal Múltiple: Inferencia Estadística
Clase 18

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

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:

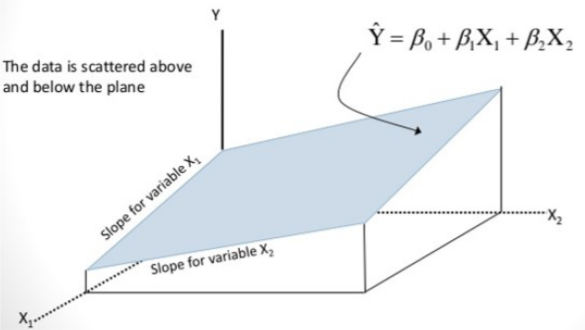
$$Y|X_1, \cdots, X_k \sim Normal\left(\beta_0 + \beta_1 X_1 + \cdots \beta_k X_k \ , \ \sigma^2\right)$$

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.

Planteamiento

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.

Planteamiento

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 $cov(\epsilon_i, \epsilon_j) = 0, i \neq j$

Planteamiento

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 $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

Planteamiento

Los supuestos en formato matricial quedan:

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

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

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

Outline

1. Regresión Lineal Múltiple
   - Planteamiento
   - Estimación
   - Propiedades del estimador
   - Pruebas de Hipótesis
   - Intervalos de Confianza y Predicción

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

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:

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

Outline

1. Regresión Lineal Múltiple
   - Planteamiento
   - Estimación
   - Propiedades del estimador
   - Pruebas de Hipótesis
   - Intervalos de Confianza y Predicción

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 $glE = n - k - 1 = n - \#\beta etas$.

Outline

Pruebas de Hipótesis: Significancia Individual

Una de las preguntas más importantes que queremos contestar, es Para saber si la variable X_j es o no relevante para explicar el comportamiento de Y. En nuestra notación, esto es equivalente a preguntar si:

$$
\begin{aligned}
H_0: &\quad \text{La Variable X}_\text{j} \text{ NO es significativa} \Leftrightarrow \beta_j = 0 \\
H_1: &\quad \text{La Variable X}_\text{j} \text{ SI es significativa} \Leftrightarrow \beta_j \neq 0
\end{aligned}
$$

$$
t = \frac{\hat{\beta}_j}{\sqrt{\hat{Var}(\hat{\beta}_j)}} \sim t_{glE}
$$

donde $\hat{Var}(\hat{\beta}_j)$ es la j-ésima posición de la diagonal de la matriz $MSE(\mathbf{X}^T\mathbf{X})^{-1}$

Pruebas de Hipótesis:Significancia Global

La pregunta anterior prueba una a una las variables incorporadas al modelo. ¿Qué sucede si queremos preguntar por el global del modelo? Es decir, se quiere probar si las variables como grupo sirven o no para explicar el comportamiento de Y .

$$
\begin{aligned}
H_0 : &\quad \text{El Modelo NO es significativo} \Leftrightarrow \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\
H_1 : &\quad \text{El Modelo SI es significativo} \Leftrightarrow \text{Algún } \beta_j \neq 0, j \in \{1, \cdots, k\}
\end{aligned}
$$

En forma general, esta hipótesis dice si el modelo con las $X's$ que se incluyeron, sirven o no para predecir Y.

¿No les resulta familiar la hipótesis?

Pruebas de Hipótesis: Significancia Global

Efectivamente, es muy parecida al hipótesis del ANOVA! Vamos a construir su equivalente para regresión. Para eso definimos:

Las sumas de cuadrados

$$\text{Suma de cuadrados total } SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \mathbf{Y}^T\mathbf{Y} - n\bar{Y}^2$$

$$\text{Suma de cuad. de la regresión } SSR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{\beta}^T(\mathbf{X}^T\mathbf{Y}) - n\bar{Y}^2$$

$$\text{Suma de cuadrados del la error } SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \mathbf{Y}^T\mathbf{Y} - \hat{\beta}^T(\mathbf{X}^T\mathbf{Y})$$

Pruebas de Hipótesis: Significancia Global

La interpretación de las sumas de cuadrados es la misma:

- SST es la variación total observada en el comportamiento de Y.
- SSR es la variación de Y que es explicada por el modelo de regresión, es decir, por la inclusión de las variables explicativas.
- SSE es la parte de la variación total de Y, que no se puede explicar con el modelo de regresión.

Ecuación fundamental de las sumas de cuadrados

$$SST = SSR + SSE$$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Pruebas de Hipótesis: Significancia Global

Recordando que $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \beta_k X_k$, los grados de libertad son:

Grados de libertad

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \Rightarrow gl_T = n - 1$$

$$SSR = \sum_{i=1}^{n} \left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \beta_k X_k - \bar{Y}\right)^2 \Rightarrow gl_R = k$$

$$SSE = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \beta_k X_k\right)^2 \Rightarrow gl_E = n - k - 1$$

$$gl_T = gl_R + gl_E$$

Pruebas de Hipótesis: Significancia Global

Por lo tanto el estadístico F queda:

$$
\begin{aligned}
H_0 &: \quad \text{El Modelo NO es significativo} \Leftrightarrow \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\
H_1 &: \quad \text{El Modelo SI es significativo} \Leftrightarrow \text{Algún } \beta_j \neq 0, j \in \{1, \cdots, k\}
\end{aligned}
$$

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} \sim F_{k,n-k-1}$$

$$RH_0 \Leftrightarrow F \geq F_{1-\alpha,k,n-k-1}$$

Nota

Si bien no se menciona, los pasos intermedios que hicimos cuando vimos la prueba F para DOE, así como la aplicación del teorema de Cochran, se mantienen para la equivalente aquí en regresión.

Combinaciones lineal de betas

Algunas cuentas de interés se pueden dar en una expresión de este estilo:

$$\theta = \sum_{j=0}^{k} c_j \beta_j$$

donde los c_j son constantes fijas, asociadas a costos, ingresos, etc. El objetivo es realizar procedimientos de inferencia estadística (pruebas de hipótesis e intervalos de confianza).

El estimador natural de esta expresión estaría dado por:

$$\hat{\theta} = \sum_{j=0}^{k} c_j \hat{\beta}_j = \mathbf{c}^T \hat{\beta}$$

donde $\mathbf{c}^T = (c_0, c_1, \cdots, c_k)$ es un vector

Combinaciones lineal de betas

De esto se deduce que:

$$\hat{\theta} = \sum_{j=0}^{k} c_j \hat{\beta}_j \sim N\left(\sum_{j=0}^{k} c_j \beta_j, \sum_{j=0}^{k} c_j^2 Var(\hat{\beta}_j) + 2\sum_{j=0}^{k}\sum_{i=0}^{j-1} c_i c_j Cov(\hat{\beta}_i, \hat{\beta}_j)\right)$$

o matricialmente

$$\hat{\theta} = \mathbf{c}^T \hat{\beta} \sim N\left(\mathbf{c}^T \beta, \mathbf{c}^T Var(\hat{\beta})\mathbf{c}\right)$$

donde $Var(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$

Combinaciones lineal de betas

Luego estimando la varianza se tiene que:

Pruebas de hipótesis

$$t = \frac{\mathbf{c}^T\hat{\beta} - \theta_0}{\sqrt{\mathbf{c}^T \hat{Var}(\hat{\beta})\mathbf{c}}} \sim t_{glE}$$

Intervalos de Confianza

$$IC_{1-\alpha}(\theta) = \mathbf{c}^T\hat{\beta} \pm t_{1-\alpha/2, gl_E}\sqrt{\mathbf{c}^T \hat{Var}(\hat{\beta})\mathbf{c}}$$

donde $\hat{Var}(\hat{\beta}) = MSE(\mathbf{X}^T\mathbf{X})^{-1}$

Outline

1. Regresión Lineal Múltiple
   - Planteamiento
   - Estimación
   - Propiedades del estimador
   - Pruebas de Hipótesis
   - Intervalos de Confianza y Predicción

Intervalos de Confianza y Predicción

Otra pregunta de interés, está asociada a si se desea conocer si el valor medio de Y toma un valor particular, dado cierto valor de X. El mejor estimador de $E(Y|X)$ está dado por la recta de regresión:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k = \mathbf{x}^T \hat{\beta}$$

donde $x^T = c(1, x_1, \cdots, x_k)$. Por lo que ya se dedujo del caso anterior, se tiene que:

$$\hat{Y} = \mathbf{x}^T \hat{\beta} \sim N\left(\mathbf{x}^T \beta, \mathbf{x}^T Var(\hat{\beta})\mathbf{x}\right)$$

donde $Var(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$

Intervalos de Confianza

Intervalo de confianza para la media de Y dado X

$$IC_{1-\alpha}\left(E(Y|X_1,\cdots,X_k)\right) = \mathbf{x}^T\hat{\beta} \pm t_{gl_E;1-\frac{\alpha}{2}}\sqrt{\mathbf{x}^T\hat{Var}(\hat{\beta})\mathbf{x}}$$

Intervalos de predicción para Y

$$IP_{1-\alpha}\left(E(Y|X_1,\cdots,X_k)\right) = \mathbf{x}^T\hat{\beta} \pm t_{gl_E;1-\frac{\alpha}{2}}\sqrt{MSE + \mathbf{x}^T\hat{Var}(\hat{\beta})\mathbf{x}}$$

El significado de este intervalo es completamente diferente, dado que no se infiere sobre el valor de un parámetro (constante).

donde $\hat{Var}(\hat{\beta}) = MSE(\mathbf{X}^T\mathbf{X})^{-1}$