Modelos Lineales

Regresión Lineal Múltiple

Fernando Massa; Bruno Bellagamba

jueves 21 de marzo 2024

FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN
IESTA INSTITUTO DE ESTADÍSTICA
UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Temario

- Definiremos la matriz de proyección H.
- Analizaremos sus propiedades.
- Obtendremos los momentos de $\hat{Y}$ y $\hat{\varepsilon}$.
- Retomaremos la descomposición de la variabilidad de Y pero matricialmente.

Vimos que podíamos plantear la siguiente descomposición ortogonal de Y.

$$Y = \hat{Y} + \hat{\varepsilon}$$

Siendo $\hat{Y}$ y $\hat{\varepsilon}$ los los vectores de predichos y residuos. El primero es la proyección ortogonal de Y en $\mathbb{C}(X)$ y el segundo es la proyección sobre el complemento ortogonal de $\mathbb{C}(X)$.

Vimos que podíamos plantear la siguiente descomposición ortogonal de Y.

$$Y = \hat{Y} + \hat{\varepsilon}$$

Siendo $\hat{Y}$ y $\hat{\varepsilon}$ los los vectores de predichos y residuos. El primero es la proyección ortogonal de Y en $\mathbb{C}(X)$ y el segundo es la proyección sobre el complemento ortogonal de $\mathbb{C}(X)$.

Si ingresamos el estimador de mínimos cuadrados dentro de esta ecuación obtenemos:

$$\begin{aligned} Y &= \hat{Y} + \hat{\varepsilon} \\ &= X\hat{\beta}_{MCO} + \left(Y - X\hat{\beta}_{MCO}\right) \\ &= X\left(X'X\right)^{-1}X'Y + \left(Y - X\left(X'X\right)^{-1}X'Y\right) \end{aligned}$$

En este momento definimos

$$H = X\left(X^{'}X\right)^{-1}X^{'}$$

obtenemos:

$$\begin{aligned} Y &= \underbrace{X\left(X^{'}X\right)^{-1}X^{'}}_{H}Y + (Y - \underbrace{X\left(X^{'}X\right)^{-1}X^{'}}_{H}Y) \\ Y &= \underbrace{HY}_{\hat{Y}} + \underbrace{(I-H)Y}_{\hat{\varepsilon}} \end{aligned}$$

A partir de esta descomposición es que en inglés se suele referir a esta matriz como la *hat matrix* ya que es la que le “pone el gorro” a la Y.

Definición

Decimos que una cierta matriz P es una matriz de proyección sobre un subespacio W si cumple que:

- $\forall Y \in W \quad PY = Y$
- $\forall Y \in W^{\perp} \quad PY = 0$

Matriz de proyección

Definición

Decimos que una cierta matriz P es una matriz de proyección sobre un subespacio W si cumple que:

- $\forall Y \in W \quad PY = Y$
- $\forall Y \in W^{\perp} \quad PY = 0$

Análogamente podemos decir que si P es matriz de proyección sobre W, entonces $I_n - P$ es matriz de proyección ortogonal sobre $W^{\perp}$.

Matriz de proyección

Definición

Decimos que una cierta matriz P es una matriz de proyección sobre un subespacio W si cumple que:

- $\forall Y \in W \quad PY = Y$
- $\forall Y \in W^{\perp} \quad PY = 0$

Análogamente podemos decir que si P es matriz de proyección sobre W, entonces $I_n - P$ es matriz de proyección ortogonal sobre $W^{\perp}$.

A modo de ejercicio, si P es matriz de proyección, entonces $P(I - P)Y =?$

Dos propiedades que usaremos bastante son las siguientes:

P es matriz de proyección si y solo si cumple que:

- $P = P^{'}$, es simétrica.
- $PP = P$, es idempotente.

Popiedades

Dos propiedades que usaremos bastante son las siguientes:

P es matriz de proyección si y solo si cumple que:

- $P = P^{'}$, es simétrica.
- $PP = P$, es idempotente.

Veamos como nuestra matriz H cumple con ambas:

Dos propiedades que usaremos bastante son las siguientes:

P es matriz de proyección si y solo si cumple que:

- $P = P^{'}$, es simétrica.
- $PP = P$, es idempotente.

Veamos como nuestra matriz H cumple con ambas:

1) Simetría

$$\begin{aligned}
H &= X\left(X^{'}X\right)^{-1}X^{'} \\
H^{'} &= \left[X\left(X^{'}X\right)^{-1}X^{'}\right]^{'} \\
H^{'} &= (X^{'})^{'}\left[\left(X^{'}X\right)^{-1}\right]^{'}X^{'} \\
H^{'} &= X\left(X^{'}X\right)^{-1}X^{'}
\end{aligned}$$

2) Idempotencia

$$
\begin{aligned}
H &= X\left(X^{'}X\right)^{-1}X^{'} \\
HH &= X\left(X^{'}X\right)^{-1}X^{'}X\left(X^{'}X\right)^{-1}X^{'} \\
HH &= X\left(X^{'}X\right)^{-1}X^{'}
\end{aligned}
$$

Gracias a esta propiedad es fácil demostrar que:

- $H^k = H$
- $I_n - H$ también es idempotente
- $H(I_n - H) = 0_n$

Propiedades

Todos los valores propios de H son 0 o 1.

Propiedades

Todos los valores propios de H son 0 o 1. Sean λ y v un valor y un vector propio de H.

Entonces se cumple que:

$$Hv = \lambda v$$

Propiedades

Todos los valores propios de H son 0 o 1. Sean λ y v un valor y un vector propio de H. Entonces se cumple que:

$$Hv = \lambda v$$

Pero como H es idempotente:

$$HHv = \lambda v$$

Propiedades

Todos los valores propios de H son 0 o 1. Sean λ y v un valor y un vector propio de H.

Entonces se cumple que:

$$Hv = \lambda v$$

Pero como H es idempotente:

$$HHv = \lambda v$$

Sustituyendo Hv por λv:

$$\lambda Hv = \lambda v$$

PROPIEDADES

Todos los valores propios de H son 0 o 1. Sean λ y v un valor y un vector propio de H.

Entonces se cumple que:

$$Hv = \lambda v$$

Pero como H es idempotente:

$$HHv = \lambda v$$

Sustituyendo Hv por λv:

$$\lambda Hv = \lambda v$$

Y volviendo a sustituir Hv por λv:

$$\begin{aligned} \lambda\lambda v &= \lambda v \\ \lambda^2 v &= \lambda v \end{aligned}$$

Todos los valores propios de H son 0 o 1. Sean λ y v un valor y un vector propio de H. Entonces se cumple que:

$$Hv = \lambda v$$

Pero como H es idempotente:

$$HHv = \lambda v$$

Sustituyendo Hv por λv:

$$\lambda Hv = \lambda v$$

Y volviendo a sustituir Hv por λv:

$$\begin{aligned} \lambda \lambda v &= \lambda v \\ \lambda^2 v &= \lambda v \end{aligned}$$

Como $v \neq \mathbf{0}$ la ecuación anterior debe cumplir que $\lambda(1-\lambda) = 0$, que solo admite soluciones $\lambda = 0$ y $\lambda = 1$.

De hecho, $k+1$ valores propios son 1 y los restantes $n-k-1$ son 0.

De hecho, $k+1$ valores propios son 1 y los restantes $n-k-1$ son 0.

Para observar esto debemos tener en cuenta que todas las columnas de X *"viven"* en $\mathbb{C}(X)$, entonces, al proyectar una cierta columna x_j sobre $\mathbb{C}(X)$, se vuelve a obtener x_j, es decir:

$$Hx_j = x_j$$

Lo cual indica que las columnas de X son los vectores propios de H y cada una de ellas tiene como valor propio, al 1.

De hecho, $k+1$ valores propios son 1 y los restantes $n-k-1$ son 0.

Para observar esto debemos tener en cuenta que todas las columnas de X *"viven"* en $\mathbb{C}(X)$, entonces, al proyectar una cierta columna x_j sobre $\mathbb{C}(X)$, se vuelve a obtener x_j, es decir:

$$Hx_j = x_j$$

Lo cual indica que las columnas de X son los vectores propios de H y cada una de ellas tiene como valor propio, al 1.

El resto de los valores propios son todos 0, debido a que:

1. No existen más vectores que cumplan que $Hv = v$.
2. Los valores propios solo pueden ser 0 o 1.

En virtud de las propiedades anteriores, es posible obtener que:

$$tr(H) = k + 1$$

Esto se debe a que la traza de una matriz simétrica es la suma de sus valores propios.

$tr(H)$ y $rg(H)$

En virtud de las propiedades anteriores, es posible obtener que:

$$tr(H) = k + 1$$

Esto se debe a que la traza de una matriz simétrica es la suma de sus valores propios.

Una propiedad similar indica que el rango de una matriz simétrica coincide con el número de valores propios distintos de cero, entonces:

$$rg(H) = k + 1$$

Esperanza de $\hat{Y}$ y $\hat{\varepsilon}$

Recordemos que:

$$\begin{aligned} \hat{Y} &= HY \\ \hat{\varepsilon} &= (I_n - H)Y \end{aligned}$$

Esperanza de $\hat{Y}$ y $\hat{\varepsilon}$

Recordemos que:

$$\begin{aligned}\hat{Y} &= HY \\ \hat{\varepsilon} &= (I_n - H)Y\end{aligned}$$

Las esperanzas de los vectores de predichos y residuos son:

$$\begin{aligned}\mathbb{E}(\hat{Y}) &= \mathbb{E}(HY) \\ \mathbb{E}(\hat{Y}) &= H\mathbb{E}(Y) \\ \mathbb{E}(\hat{Y}) &= HX\beta \\ \mathbb{E}(\hat{Y}) &= X\beta \\ \mathbb{E}(\hat{Y}) &= Y\end{aligned}$$

$$\begin{aligned}\mathbb{E}(\hat{\varepsilon}) &= \mathbb{E}[(I_n - H)Y] \\ \mathbb{E}(\hat{\varepsilon}) &= (I_n - H)\mathbb{E}(Y) \\ \mathbb{E}(\hat{\varepsilon}) &= (I_n - H)X\beta \\ \mathbb{E}(\hat{\varepsilon}) &= X\beta - HX\beta \\ \mathbb{E}(\hat{\varepsilon}) &= X\beta - X\beta \\ \mathbb{E}(\hat{\varepsilon}) &= \mathbf{0}\end{aligned}$$

VARIANZA DE $\hat{Y}$ Y $\hat{\varepsilon}$

... y sus matrices de covarianzas:

$$
\begin{aligned}
\mathbb{C}ov(\hat{Y}) &= \mathbb{C}ov(HY) \\
\mathbb{C}ov(\hat{Y}) &= H\mathbb{C}ov(Y)H^{'} \\
\mathbb{C}ov(\hat{Y}) &= \sigma^2 HH^{'} \\
\mathbb{C}ov(\hat{Y}) &= \sigma^2 HH \\
\mathbb{C}ov(\hat{Y}) &= \sigma^2 H
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{C}ov(\hat{\varepsilon}) &= \mathbb{C}ov\left[(I_n - H)Y\right] \\
\mathbb{C}ov(\hat{\varepsilon}) &= (I_n - H)\mathbb{C}ov(Y)(I_n - H)^{'} \\
\mathbb{C}ov(\hat{\varepsilon}) &= \sigma^2 (I_n - H)(I_n - H)^{'} \\
\mathbb{C}ov(\hat{\varepsilon}) &= \sigma^2 (I_n - H)(I_n - H) \\
\mathbb{C}ov(\hat{\varepsilon}) &= \sigma^2 (I_n - H)
\end{aligned}
$$

Varianza de $\hat{Y}$ y $\hat{\varepsilon}$

... y sus matrices de covarianzas:

$$
\begin{aligned}
\mathbb{C}ov(\hat{Y}) &= \mathbb{C}ov(HY) \\
\mathbb{C}ov(\hat{Y}) &= H\mathbb{C}ov(Y)H^{'} \\
\mathbb{C}ov(\hat{Y}) &= \sigma^2 HH^{'} \\
\mathbb{C}ov(\hat{Y}) &= \sigma^2 HH \\
\mathbb{C}ov(\hat{Y}) &= \sigma^2 H
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{C}ov(\hat{\varepsilon}) &= \mathbb{C}ov\left[(I_n - H)Y\right] \\
\mathbb{C}ov(\hat{\varepsilon}) &= (I_n - H)\mathbb{C}ov(Y)(I_n - H)^{'} \\
\mathbb{C}ov(\hat{\varepsilon}) &= \sigma^2 (I_n - H)(I_n - H)^{'} \\
\mathbb{C}ov(\hat{\varepsilon}) &= \sigma^2 (I_n - H)(I_n - H) \\
\mathbb{C}ov(\hat{\varepsilon}) &= \sigma^2 (I_n - H)
\end{aligned}
$$

De esta manera, observamos que a la interna de ambos vectores, los valores NO están incorrelacionados, salvo en el caso particular de que H sea una matriz diagonal. Esto es particularmente llamativo en el caso de $\hat{\varepsilon}$.

Estimación de σ^2

En el modelo de RLS planteamos un estimador de σ^2 basado en la varianza muestral de los residuos.

Planteando la *SCRes* en forma matricial obtenemos:

$$
\begin{aligned}
SCRes &= \hat{\varepsilon}'\hat{\varepsilon} \\
&= [(I_n - H)Y]'(I_n - H)Y \\
&= Y'(I_n - H)'(I_n - H)Y \\
&= Y'(I_n - H)Y
\end{aligned}
$$

En el modelo de RLS planteamos un estimador de σ^2 basado en la varianza muestral de los residuos.

Planteando la *SCRes* en forma matricial obtenemos:

$$\begin{aligned} SCRes &= \hat{\varepsilon}'\hat{\varepsilon} \\ &= [(I_n - H)Y]'(I_n - H)Y \\ &= Y'(I_n - H)'(I_n - H)Y \\ &= Y'(I_n - H)Y \end{aligned}$$

Calculemos ahora el valor esperado de esta cantidad:

$$\mathbb{E}(SCRes) = \mathbb{E}[Y'(I_n - H)Y]$$

Estimación de σ^2

Teorema

Sea Y un vector aleatorio, $\mathbb{E}(Y) = \mu$ y $\mathbb{C}ov(Y) = \Sigma$, entonces:

$$\mathbb{E}(Y^{'}AY) = tr(A\Sigma) + \mu^{'}A\mu$$

Estimación de σ^2

Teorema

Sea Y un vector aleatorio, $\mathbb{E}(Y) = \mu$ y $\mathbb{C}ov(Y) = \Sigma$, entonces:

$$\mathbb{E}(Y^{'}AY) = tr(A\Sigma) + \mu^{'}A\mu$$

Usando $I_n - H$ en el lugar de A y recordando que $\mathbb{E}(Y) = X\beta$ y $\mathbb{C}ov(Y) = \sigma^2 I_n$:

$$\begin{aligned}
\mathbb{E}(SCRes) &= \mathbb{E}[Y^{'}(I_n - H)Y] \\
&= tr[(I_n - H)\sigma^2 I_n] + (X\beta)^{'}(I_n - H)X\beta \\
&= \sigma^2 tr[(I_n - H)] \\
&= \sigma^2 [tr(I_n) - tr(H)] \\
&= \sigma^2 (n - k - 1)
\end{aligned}$$

Teorema

Sea Y un vector aleatorio, $\mathbb{E}(Y) = \mu$ y $\mathbb{C}ov(Y) = \Sigma$, entonces:

$$\mathbb{E}(Y^{'}AY) = tr(A\Sigma) + \mu^{'}A\mu$$

Usando $I_n - H$ en el lugar de A y recordando que $\mathbb{E}(Y) = X\beta$ y $\mathbb{C}ov(Y) = \sigma^2 I_n$:

$$\begin{aligned}
\mathbb{E}(SCRes) &= \mathbb{E}[Y^{'}(I_n - H)Y] \\
&= tr[(I_n - H)\sigma^2 I_n] + (X\beta)^{'}(I_n - H)X\beta \\
&= \sigma^2 tr[(I_n - H)] \\
&= \sigma^2[tr(I_n) - tr(H)] \\
&= \sigma^2(n - k - 1)
\end{aligned}$$

De esta manera, un estimador insesgado para σ^2 sería:

$$\hat{\sigma}^2 = \frac{Y^{'}(I_n - H)Y}{n - k - 1}$$

$SCT = SCExp + SCRes$

Expresemos la ecuación de la descomposición de la variabilidad en forma escalar.

$$\begin{aligned} SCT &= SCExp + SCRes \\ \sum_i (y_i - \bar{y})^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y})^2 \\ \sum_i y_i^2 - n\bar{y}^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y})^2 \end{aligned}$$

$SCT = SCExp + SCRes$

Expresemos la ecuación de la descomposición de la variabilidad en forma escalar.

$$\begin{aligned} SCT &= SCExp + SCRes \\ \sum_i (y_i - \bar{y})^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y})^2 \\ \sum_i y_i^2 - n\bar{y}^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y})^2 \end{aligned}$$

Una manera de llegar a una expresión matricial equivalente a la anterior parte de plantear la suma de cuadrados (sin centrar) de Y:

$$\begin{aligned} Y'Y &= [(H + (I_n - H))Y]'[(H + (I_n - H))Y] \\ &= Y'HY + Y'(I_n - H)Y \end{aligned}$$

Expresemos la ecuación de la descomposición de la variabilidad en forma escalar.

$$\begin{aligned}
SCT &= SCExp + SCRes \\
\sum_i (y_i - \bar{y})^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y})^2 \\
\sum_i y_i^2 - n\bar{y}^2 &= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y})^2
\end{aligned}$$

Una manera de llegar a una expresión matricial equivalente a la anterior parte de plantear la suma de cuadrados (sin centrar) de Y:

$$\begin{aligned}
Y^{'}Y &= [(H + (I_n - H))Y]^{'}[(H + (I_n - H))Y] \\
&= Y^{'}HY + Y^{'}(I_n - H)Y
\end{aligned}$$

A continuación debemos corregir esta ecuación de modo que a la izquierda del igual se obtenga $\sum_i (y_i - \bar{y})^2$

$SCT = SCExp + SCRes$

El vector de predichos de un modelo que solo contiene la constante $(y_i = \beta_0 + \varepsilon)$:

$$\begin{aligned}\hat{Y} &= H_{\mathbb{1}} Y \\ \hat{Y} &= \mathbb{1} \underbrace{(\mathbb{1}^{'} \mathbb{1})^{-1}}_{\frac{1}{n}} \underbrace{\mathbb{1}^{'} Y}_{\sum_i y_i} \\ \hat{Y} &= \mathbb{1} \bar{y}\end{aligned}$$

Siendo $H_{\mathbb{1}}$ la matriz de proyeccción sobre el subespacio generado por el vector de unos.

$SCT = SCExp + SCRes$

El vector de predichos de un modelo que solo contiene la constante ($y_i = \beta_0 + \varepsilon$):

$$
\begin{aligned}
\hat{Y} &= H_{\mathbb{1}} Y \\
\hat{Y} &= \mathbb{1} \underbrace{(\mathbb{1}^{'}\mathbb{1})^{-1}}_{\frac{1}{n}} \underbrace{\mathbb{1}^{'} Y}_{\sum_i y_i} \\
\hat{Y} &= \mathbb{1}\bar{y}
\end{aligned}
$$

Siendo $H_{\mathbb{1}}$ la matriz de proyeccción sobre el subespacio generado por el vector de unos.

Entonces, para obtener $n\bar{y}^2$ basta con plantear $\hat{Y}^{'}\hat{Y}$,

$SCT = SCExp + SCRes$

El vector de predichos de un modelo que solo contiene la constante ($y_i = \beta_0 + \varepsilon$):

$$\begin{aligned}
\hat{Y} &= H_{\mathbb{1}} Y \\
\hat{Y} &= \mathbb{1} \underbrace{(\mathbb{1}^{'} \mathbb{1})^{-1}}_{\frac{1}{n}} \underbrace{\mathbb{1}^{'} Y}_{\sum_i y_i} \\
\hat{Y} &= \mathbb{1} \bar{y}
\end{aligned}$$

Siendo $H_{\mathbb{1}}$ la matriz de proyeccción sobre el subespacio generado por el vector de unos.

Entonces, para obtener $n\bar{y}^2$ basta con plantear $\hat{Y}^{'} \hat{Y}$, obteniendo finalmente:

$$\begin{aligned}
n\bar{y}^2 &= \hat{Y}^{'} \hat{Y} \\
&= Y^{'} H_{\mathbb{1}}^{'} H_{\mathbb{1}} Y \\
&= Y^{'} H_{\mathbb{1}} Y
\end{aligned}$$

$SCT = SCExp + SCRes$

El vector de predichos de un modelo que solo contiene la constante ($y_i = \beta_0 + \varepsilon$):

$$
\begin{aligned}
\hat{Y} &= H_{\mathbb{1}} Y \\
\hat{Y} &= \mathbb{1} \underbrace{(\mathbb{1}^{'} \mathbb{1})^{-1}}_{\frac{1}{n}} \underbrace{\mathbb{1}^{'} Y}_{\sum_i y_i} \\
\hat{Y} &= \mathbb{1} \bar{y}
\end{aligned}
$$

Siendo $H_{\mathbb{1}}$ la matriz de proyeccción sobre el subespacio generado por el vector de unos.

Entonces, para obtener $n\bar{y}^2$ basta con plantear $\hat{Y}^{'} \hat{Y}$, obteniendo finalmente:

$$
\begin{aligned}
n\bar{y}^2 &= \hat{Y}^{'} \hat{Y} \\
&= Y^{'} H_{\mathbb{1}}^{'} H_{\mathbb{1}} Y \\
&= Y^{'} H_{\mathbb{1}} Y
\end{aligned}
$$

Empleando esta expresión junto con la de la diapositiva anterior, obtenemos la descomposición de la variabilidad.

Restando $Y^{'}H_{\mathbb{1}}Y$ a ambos lados de la diapositiva ante-anterior, se obtiene

$$\begin{aligned} SCT &= SCExp + SCRes \\ Y^{'}Y - Y^{'}H_{\mathbb{1}}Y &= Y^{'}H_XY - Y^{'}H_{\mathbb{1}}Y + Y^{'}(I_{n\times n} - H_X)Y \\ Y^{'}(I_n - H_{\mathbb{1}})Y &= Y^{'}(H_X - H_{\mathbb{1}})Y + Y^{'}(I_n - H_X)Y \end{aligned}$$

Siendo H_X la matriz de proyección sobre el subespacio generado por todas las columnas de X.

Restando $Y^{'}H_{\mathbb{1}}Y$ a ambos lados de la diapositiva ante-anterior, se obtiene

$$
\begin{aligned}
SCT &= SCExp + SCRes \\
Y^{'}Y - Y^{'}H_{\mathbb{1}}Y &= Y^{'}H_X Y - Y^{'}H_{\mathbb{1}}Y + Y^{'}(I_{n\times n} - H_X)Y \\
Y^{'}(I_n - H_{\mathbb{1}})Y &= Y^{'}(H_X - H_{\mathbb{1}})Y + Y^{'}(I_n - H_X)Y
\end{aligned}
$$

Siendo H_X la matriz de proyección sobre el subespacio generado por todas las columnas de X.

Una posible aplicación de esta descomposición es la definición del estadístico R^2.

Restando $Y^{'} H_{\mathbb{1}} Y$ a ambos lados de la diapositiva ante-anterior, se obtiene

$$
\begin{aligned}
SCT &= SCExp + SCRes \\
Y^{'}Y - Y^{'}H_{\mathbb{1}}Y &= Y^{'}H_XY - Y^{'}H_{\mathbb{1}}Y + Y^{'}(I_{n\times n} - H_X)Y \\
Y^{'}(I_n - H_{\mathbb{1}})Y &= Y^{'}(H_X - H_{\mathbb{1}})Y + Y^{'}(I_n - H_X)Y
\end{aligned}
$$

Siendo H_X la matriz de proyección sobre el subespacio generado por todas las columnas de X.

Una posible aplicación de esta descomposición es la definición del estadístico R^2.

$$
\begin{aligned}
R^2 &= \frac{SCExp}{SCT} \\
&= \frac{Y^{'}(H_X - H_{\mathbb{1}})Y}{Y^{'}(I_n - H_{\mathbb{1}})Y} \\
&= \frac{Y^{'}H_xY - n\bar{y}^2}{Y^{'}Y - n\bar{y}^2}
\end{aligned}
$$

En el modelo de RLM, se le suele llamar *coeficiente de determinación*.

En la diapositiva anterior restamos $Y^{'}H_{\mathbb{1}}Y$ en ambos lados de la igualdad. Lo que logramos con esto fue **remover el efecto de la "variable" $\mathbb{1}$ de la variabilidad de Y**. Lo que es aún más importante, es que a la derecha de la igualdad, este término fue a parar a la *SCExp*.

Descomposición secuencial de *SCExp*

En la diapositiva anterior restamos $Y^{'}H_{\mathbb{1}}Y$ en ambos lados de la igualdad. Lo que logramos con esto fue **remover el efecto de la "variable" $\mathbb{1}$ de la variabilidad de Y**. Lo que es aún más importante, es que a la derecha de la igualdad, este término fue a parar a la *SCExp*.

$$SCExp = Y^{'}(H_X - H_{\mathbb{1}})Y$$

Lo que se puede interpretar de este término es que se está midiendo la cantidad de la variabilidad de Y **atribuible** a las variables incluidas en la matriz X **luego de remover el efecto de** $\mathbb{1}$.

En la diapositiva anterior restamos $Y^{'}H_{\mathbb{1}}Y$ en ambos lados de la igualdad. Lo que logramos con esto fue **remover el efecto de la "variable" $\mathbb{1}$ de la variabilidad de Y**. Lo que es aún más importante, es que a la derecha de la igualdad, este término fue a parar a la $SCExp$.

$$SCExp = Y^{'}(H_X - H_{\mathbb{1}})Y$$

Lo que se puede interpretar de este término es que se está midiendo la cantidad de la variabilidad de Y **atribuible** a las variables incluidas en la matriz X **luego de remover el efecto de $\mathbb{1}$**.

Gracias a esta observación podemos pensar en descomponer la variabilidad de $SCExp$ en la sumatoria de los aportes de cada variable explicativa.

Así como definimos $H_{\mathbb{1}}$ como la matriz de proyección sobre el vector de unos, podemos definir $X_{\mathbb{1}x_1}$ como la matriz de proyección sobre el vector de unos **y** la variable x_1.

Así como definimos $H_{\mathbb{1}}$ como la matriz de proyección sobre el vector de unos, podemos definir $X_{\mathbb{1}x_1}$ como la matriz de proyección sobre el vector de unos **y** la variable x_1.

Al sumar y restar $H_{\mathbb{1}x_1}$ en la expresión de $SCExp$:

$$
\begin{aligned}
SCExp &= Y^{'}(H_X - H_{\mathbb{1}})Y \\
&= Y^{'}(H_X - H_{\mathbb{1}x_1} + H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y \\
&= Y^{'}(H_X - H_{\mathbb{1}x_1})Y + Y^{'}(H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y
\end{aligned}
$$

Descomposición secuencial de *SCRreg*

Así como definimos $H_{\mathbb{1}}$ como la matriz de proyección sobre el vector de unos, podemos definir $X_{\mathbb{1}x_1}$ como la matriz de proyección sobre el vector de unos **y** la variable x_1.

Al sumar y restar $H_{\mathbb{1}x_1}$ en la expresión de *SCExp*:

$$
\begin{aligned}
SCExp &= Y^{'}(H_X - H_{\mathbb{1}})Y \\
&= Y^{'}(H_X - H_{\mathbb{1}x_1} + H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y \\
&= Y^{'}(H_X - H_{\mathbb{1}x_1})Y + Y^{'}(H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y
\end{aligned}
$$

- $Y^{'}(H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y$ es la variabilidad de Y captada por x_1 en un modelo que ya incluía la constante. A este término se lo suele denominar $SCExp(X_1)$.
- $Y^{'}(H_X - H_{\mathbb{1}x_1})Y$ es la variabilidad de Y captada por las columnas de X en un modelo que ya incluye a x_1 y al vector de unos. A este término se lo suele denominar $SCExp(X_2, X_3, \ldots, X_k | X_1)$

EL R^2 PUEDE SER ENGAÑOSO

Retomando la última expresión:

$$SCExp(X_1, X_2, \ldots, X_k) = Y^{'}(H_X - H_{\mathbb{1}x_1})Y + Y^{'}(H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y$$

Se observa que los términos a la derecha del igual son formas cuadráticas, por ende, no negativos.

Retomando la última expresión:

$$SCExp(X_1, X_2, \ldots, X_k) = Y^{'}(H_X - H_{\mathbb{1}x_1})Y + Y^{'}(H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y$$

Se observa que los términos a la derecha del igual son formas cuadráticas, por ende, no negativos.

Retomando ahora la definición del R^2.

$$\begin{aligned} R^2 &= \frac{SCExp}{SCT} \\ &= \frac{SCExp(X_1)+SCExp(X_2|X_1)+SCExp(X_3|X_2,X_1)+\ldots}{Y^{'}Y-n\bar{y}^2} \end{aligned}$$

EL R^2 PUEDE SER ENGAÑOSO

Retomando la última expresión:

$$SCExp(X_1, X_2, \ldots, X_k) = Y^{'}(H_X - H_{\mathbb{1}x_1})Y + Y^{'}(H_{\mathbb{1}x_1} - H_{\mathbb{1}})Y$$

Se observa que los términos a la derecha del igual son formas cuadráticas, por ende, no negativos.

Retomando ahora la definición del R^2.

$$\begin{aligned} R^2 &= \frac{SCExp}{SCT} \\ &= \frac{SCExp(X_1) + SCExp(X_2|X_1) + SCExp(X_3|X_2, X_1) + \ldots}{Y^{'}Y - n\bar{y}^2} \end{aligned}$$

Esto nos muestra que al agregar variables explicativas al modelo (por poco o mucho que tengan que ver con la Y) el R^2 **siempre** aumenta (o al menos NO disminuye). Lo cual plantea algunos inconvenientes que veremos a la hora de comparar modelos.

Retomemos el ejemplo de las ventas en R. Usemos la matriz de proyección.

La próxima hablaremos de:

- Retomaremos la matriz de proyección para hablar sobre la parcialidad de los coeficientes $\hat{\beta}_{MCO}$.
- Observaremos qué sucede cuando las columnas de X son ortogonales.
- Plantearemos el problema de excluir variables.
- Veremos ejemplos.

Carmona, Francesc (2003). *Modelos Lineales (notas de curso).* Departament d'Estadística.

Farraway, Julian (2014). *Linear Models with R, second edition.* Chapman Hall/CRC.

Rencher, Alvin y Bruce Schaalje (2008). *Linear Models in Statistics, second edition.* John Wiley Sons, Inc.

¿Preguntas?

Muchas Gracias