Modelado Estocástico

Clase 4

Regresión Múltiple

Vamos a analizar el caso en el que se agrega una nueva variable explicativa.

Nuestro modelo ahora va a ser:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

Donde los errores, u_i, se definen igual que antes: $u_i = y_i - E(y_i)$. Notemos que por el supuesto 1, $E(u_i) = 0$.

Los residuos, e_i, se definen también igual que antes: $e_i = y_i - \hat{y}_i$.

Notemos que ahora el valor predicho será $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$

El método de mínimos cuadrados ordinarios minimiza la suma de los residuos al cuadrado. Sea

$$S(\tilde{\alpha}, \tilde{\beta}_1, \tilde{\beta}_2) = \sum_{i=1}^{n} {e_i}^2$$

$$S(\tilde{\alpha}, \tilde{\beta}_1, \tilde{\beta}_2) = \sum_{i=1}^{n} (y_i - \tilde{\alpha} - \tilde{\beta}_1\, x_{1i} - \tilde{\beta}_2\, x_{2i})^2$$

Las condiciones de primer orden de este problema de minimización son:

$$\frac{\partial S}{\partial \tilde{\alpha}} = 0 \qquad \frac{\partial S}{\partial \tilde{\beta}_1} = 0 \qquad \frac{\partial S}{\partial \tilde{\beta}_2} = 0$$

donde

$$\frac{\partial S}{\partial \tilde{\alpha}} = (-2) \sum_{i=1}^{n} \left(y_i - \tilde{\alpha} - \tilde{\beta}_1\, x_{1i} - \tilde{\beta}_2\, x_{2i}\right)$$

$$\frac{\partial S}{\partial \tilde{\beta}_1} = (-2) \sum_{i=1}^{n} \left(y_i - \tilde{\alpha} - \tilde{\beta}_1\, x_{1i} - \tilde{\beta}_2\, x_{2i}\right) x_{1i}$$

$$\frac{\partial S}{\partial \tilde{\beta}_2} = (-2) \sum_{i=1}^{n} \left(y_i - \tilde{\alpha} - \tilde{\beta}_1\, x_{1i} - \tilde{\beta}_2\, x_{2i}\right) x_{2i}$$

Note que al igualar las derivadas parciales a cero, y haciendo uso que $e_i = y_i - \tilde{\alpha} - \tilde{\beta}_1\, x_{1i} - \tilde{\beta}_2\, x_{2i}$, se obtiene

$$\sum_{i=1}^{n} e_i = 0 \qquad \sum_{i=1}^{n} e_i x_{1i} = 0 \qquad \sum_{i=1}^{n} e_i x_{2i} = 0$$

Recordemos que estas ecuaciones se llaman las **ecuaciones normales** (*normal equations*, en inglés) y la palabra "normal" no tiene nada que ver con la distribución normal.

Reescribiéndolas,

$$\textstyle\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}_1\, x_{1i} - \hat{\beta}_2\, x_{2i}\right) = 0 \qquad (1)$$

$$\textstyle\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}_1\, x_{1i} - \hat{\beta}_2\, x_{2i}\right)x_{1i} = 0 \qquad (2)$$

$$\textstyle\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}_1\, x_{1i} - \hat{\beta}_2\, x_{2i}\right)x_{2i} = 0 \qquad (3)$$

Sean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i}$,

$$\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2i}$$

De (1),

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\alpha} - \hat{\beta}_1 \sum_{i=1}^{n} x_{1i} - \hat{\beta}_2 \sum_{i=1}^{n} x_{2i} = 0$$

$$n\bar{y} - n\hat{\alpha} - \hat{\beta}_1 n\bar{x}_1 - \hat{\beta}_2 n\bar{x}_2 = 0$$

Consecuentemente,

$$\hat{\alpha} = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 \qquad (1')$$

Reemplazando (1´) en (2) se obtiene

$$\sum_{i=1}^{n}\left(y_i - \bar{y} + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}\right)x_{1i} = 0$$

Reordenando,

$$\sum y_i x_{1i} - n\bar{x}_1\bar{y} = \hat{\beta}_1\left(\sum x_{1i}^2 - n\bar{x}_1^2\right) +$$
$$+\hat{\beta}_2\left(\sum x_{1i}x_{2i} - n\bar{x}_1\bar{x}_2\right) \qquad (2')$$

Reemplazando (1´) en (3) y reordenando se obtiene:

$$\sum y_i x_{2i} - n\bar{x}_2\bar{y} = \hat{\beta}_1\left(\sum x_{1i}x_{2i} - n\bar{x}_1\bar{x}_2\right) +$$
$$+\hat{\beta}_2\left(\sum x_{2i}^2 - n\bar{x}_2^2\right) \qquad (3')$$

Sean:

$$S_{y1} = \sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(y_i - \bar{y}) = \sum y_i x_{1i} - n\bar{x}_1\bar{y}$$

$$S_{y2} = \sum_{i=1}^{n}(x_{2i} - \bar{x}_2)(y_i - \bar{y}) = \sum y_i x_{2i} - n\bar{x}_2\bar{y}$$

$$S_{11} = \sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2 = \sum x_{1i}^2 - n\bar{x}_1^2$$

$$S_{12} = \sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) = \sum x_{1i}x_{2i} - n\bar{x}_1\bar{x}_2$$

$$S_{22} = \sum_{i=1}^{n}(x_{2i} - \bar{x}_2)^2 = \sum x_{2i}^2 - n\bar{x}_2^2$$

Entonces, (2´) puede escribirse como

$$S_{y1} = \hat{\beta}_1 S_{11} + \hat{\beta}_2 S_{12} \qquad (2´´)$$

También, (3´) puede escribirse como

$$S_{y2} = \hat{\beta}_1 S_{12} + \hat{\beta}_2 S_{22} \qquad (3´´)$$

(2´´) y (3´´) forman un sistema de 2 ecuaciones con dos incógnitas. Despejando uno de los beta sombrero de una ecuación y reemplazándolo en la otra se encuentra la solución:

$$\hat{\beta}_1 = \frac{S_{y1}S_{22} - S_{y2}S_{12}}{S_{11}S_{22} - {S_{12}}^2}$$

$$\hat{\beta}_2 = \frac{S_{y2}S_{11} - S_{y1}S_{12}}{S_{11}S_{22} - {S_{12}}^2}$$

¿Existen siempre $\hat{\beta}_1$ y $\hat{\beta}_2$?

Insesgadez: se puede demostrar que los estimadores son insesgados.

Note que:

$$
\begin{aligned}
S_{y1} &= \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)y_i = \\
&= \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(\alpha+\beta_1 x_{1i}+\beta_2 x_{2i}+u_i) \\
&= \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(\beta_1 x_{1i}+\beta_2 x_{2i}+u_i) \\
&= \beta_1 S_{11} + \ \beta_2 S_{12} + \sum(x_{1i}-\bar{x}_1)\,u_i
\end{aligned}
$$

Del mismo modo,

$$
\begin{aligned}
S_{y2} &= \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)y_i = \\
&= \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(\alpha+\beta_1 x_{1i}+\beta_2 x_{2i}+u_i) \\
&= \beta_1 S_{12} + \beta_2 S_{22} + \sum(x_{2i}-\bar{x}_2)\,u_i
\end{aligned}
$$

Usamos esto en la fórmula del estimador $\hat{\beta}_1 = \frac{S_{y1}S_{22}-S_{y2}S_{12}}{S_{11}S_{22}-{S_{12}}^2}$ para obtener:

$$\hat{\beta}_1 = \frac{\begin{array}{c}[\beta_1 S_{11} + \beta_2 S_{12} + \sum(x_{1i} - \bar{x}_1)\, u_i] S_{22} - \\ -[\beta_1 S_{12} + \beta_2 S_{22} + \sum(x_{2i} - \bar{x}_2)\, u_i] S_{12}\end{array}}{S_{11} S_{22} - {S_{12}}^2}$$

$$\hat{\beta}_1 = \beta_1 + \frac{S_{22} \sum(x_{1i} - \bar{x}_1)\, u_i - S_{12} \sum(x_{2i} - \bar{x}_2)\, u_i}{S_{11} S_{22} - {S_{12}}^2}$$

Esta es la descomposición del estimador en el parámetro más el error muestral del estimador (*sampling error*). Es fácil demostrar a partir de esta última expresión que $\hat{\beta}_1$ es insesgado, es decir, $E(\hat{\beta}_1) = \beta_1$.

Análogamente, se puede demostrar que $E(\hat{\beta}_2) = \beta_2$.

Para calcular la varianza de $\hat{\beta}_1$, partimos de esta misma descomposición:

$$\hat{\beta}_1 = \beta_1 + \frac{S_{22} \sum(x_{1i} - \bar{x}_1)\, u_i - S_{12} \sum(x_{2i} - \bar{x}_2)\, u_i}{S_{11} S_{22} - {S_{12}}^2}$$

$$Var(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2 =$$
$$= E\left(\frac{S_{22}\sum(x_{1i} - \bar{x}_1)\, u_i - S_{12}\sum(x_{2i} - \bar{x}_2)\, u_i}{S_{11}S_{22} - {S_{12}}^2}\right)^2$$

Sea $A = S_{11}S_{22} - {S_{12}}^2$

$$= \frac{1}{A^2} E\left(S_{22}\sum(x_{1i} - \bar{x}_1)\, u_i - S_{12}\sum(x_{2i} - \bar{x}_2)\, u_i\right)^2$$
$$= \frac{1}{A^2} E\left[{S_{22}}^2\left(\sum(x_{1i} - \bar{x}_1)\, u_i\right)^2 + {S_{12}}^2\left(\sum(x_{2i} - \bar{x}_2)\, u_i\right)^2 - 2S_{22}S_{12}\left(\sum(x_{1i} - \bar{x}_1)\, u_i\right)\left(\sum(x_{2i} - \bar{x}_2)\, u_i\right)\right]$$
$$= \frac{\sigma^2}{A^2}\left[{S_{22}}^2 S_{11} + {S_{12}}^2 S_{22} - 2S_{22}S_{12}S_{12}\right] =$$
$$= \frac{\sigma^2}{A^2}\left[{S_{22}}^2 S_{11} - {S_{12}}^2 S_{22}\right] = \frac{\sigma^2 S_{22}}{A}$$

O sea, $Var(\hat{\beta}_1) = \frac{\sigma^2 S_{22}}{S_{11}S_{22} - {S_{12}}^2}$

Análogamente, $Var(\hat{\beta}_2) = \frac{\sigma^2 S_{11}}{S_{11}S_{22} - {S_{12}}^2}$

Sea r_{12}^2 el coeficiente de correlación muestral al cuadrado entre las variables explicativas x_1 y x_2. Es decir,

$$r_{12}^2 = \frac{{S_{12}}^2}{S_{11}S_{22}}$$

Notemos que,

$$\begin{aligned} Var(\hat{\beta}_1) &= \frac{\sigma^2 S_{22}}{S_{11}S_{22} - {S_{12}}^2} \\ &= \frac{\sigma^2 S_{22}}{S_{11}S_{22}(1 - {r_{12}}^2)} \\ &= \frac{\sigma^2}{S_{11}(1 - {r_{12}}^2)} \end{aligned}$$

Cuanto mayor sea la correlación entre las dos variables explicativas, mayor será la varianza de los estimadores de las pendientes. A este problema se lo conoce con el nombre de multicolinealidad (alta).

Cuando $r_{12}{}^2 = 1$, entonces no es posible obtener un único valor de la solución de $\hat{\beta}_1$ ó de $\hat{\beta}_2$ y a esta situación se la llama multicolinealidad perfecta.

En otras palabras, el problema de la multicolinealidad perfecta consiste en tener una de las variables explicativas como una combinación lineal de otra/otras. En este caso, no es posible obtener una única solución ya que no es posible atribuir la explicación de la variable dependiente a una variable o a la(s) otra(s). Para entenderlo intuitivamente, es como si incluyéramos la variable "x" es pesos y en centavos. ¿a cuál le atribuimos la explicación de la variable "y"?

Por otro lado, el problema de la multicolinealidad alta se da cuando la correlación entre dos variables explicativas es alta pero no es exactamente igual a +1 o bien igual a -1. Cuando esto ocurre, se obtiene una única solución para los estimadores beta sombrero, pero las varianzas de los beta sombrero son altas. Esto se detecta en la segunda columna de las salidas de los software, que es el *standard error* (es la raíz cuadrada de la varianza estimada).

Solución: la solución es eliminar de a una las variables sospechadas de producir multicolinealidad.
Es decir, si en una salida de la regresión se observan a dos o más variables con estadísticos t bajos, podemos estar bajo un problema de multicolinealidad o bien en una situación en la que dichas variables no explican a la variable dependiente. Si al eliminarlas de a una (siempre de este grupo de variables con estadístico t bajo) observamos que la que dejamos pasa ahora a tener un estadístico t alto (mayor a 1,645, por ejemplo), entonces esto quiere decir que

estábamos frente a un problema de multicolinealidad.

También es posible detectar multicolinealidad mediante un test que veremos en breve, que se llama test F, cuando testeamos la significatividad conjunta de los parámetros que multiplican a las variables de las que sospechamos multicolinealidad.

Interpretación de las pendientes: matemáticamente son derivadas parciales, de manera que en su interpretación corresponde aclarar que es en cuánto cambia $\hat{y}_i$ frente a un cambio en una unidad de la variable explicativa, ***ceteris paribus***, que significa "manteniendo todo lo demás constante". Los beta sombrero se interpretan como derivadas parciales.

Gauss-Markov: se sigue cumpliendo en el modelo de regresión múltiple con dos o más variables explicativas. O sea, $\hat{\alpha}, \hat{\beta}_1 \, y \, \hat{\beta}_2$ y cualquier combinación lineal entre ellos es MELI (BLUE).

Punto de medias: notar que la regresión va a pasar por el punto de medias como en el modelo de regresión simple. ¿Puede mostrarlo?

Ecuaciones normales: notemos también que de las ecuaciones normales surge que $\sum_{i=1}^{n} e_i \, \hat{y}_i = 0$

Suma de Cuadrados: se sigue cumpliendo que se puede descomponer el TSS en el ESS y el RSS.

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} {e_i}^2$$

$$TSS = \quad ESS \quad + RSS$$

RSS quiere decir "*residual sum of squares*", TSS quiere decir "*total sum of squares*" y ESS quiere decir "*explained sum of squares*". Note que

$$RSS = \sum_{i=1}^{n} {e_i}^2$$

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Si dividimos la expresión por TSS, obtenemos que

$$1 = \mathrm{R}^2 + \frac{RSS}{TSS}$$

Definición: el coeficiente de determinación R^2 se define como $\mathrm{R}^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}}$. Se puede demostrar que en el caso de regresión múltiple, el R^2 es igual a la correlación al cuadrado entre y_i e $\hat{y}_i$.

Cuando agregamos más variables explicativas a nuestro modelo, por más que sean irrelevantes, el RSS cae y el R^2 aumenta. Por eso, para comparar entre dos modelos donde uno incluye las mismas variables que otro y otras adicionales, corresponde usar el concepto de R^2 ajustado, que se denota $\overline{\mathrm{R}}^2$ y se define:

$\overline{\mathrm{R}}^2$ ajustado:

$$\overline{\mathrm{R}}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$$

donde n es el número de observaciones y k el número de parámetros a estimar (intercepto más pendientes).

Distribución del s^2: ahora, en regresión múltiple, $\frac{(n-k)s^2}{\sigma^2}$ tiene una distribución chi-cuadrado con $n-k$ grados de libertad, donde k el número de parámetros a estimar (intercepto más pendientes).

Test de Hipótesis:

Supongamos que nuestro modelo es

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

y queremos testear estadísticamente si x_2 afecta a y. Esto puede plantearse mediante un test de hipótesis:

$$H_0 : \beta_2 = 0$$
$$H_A : \beta_2 \neq 0$$

Este es un test de hipótesis a dos colas. Funciona de la misma manera que en el caso de regresión simple, pero con la única diferencia que el estadístico va a tener una distribución t de Student con la misma cantidad de grados de libertad que la chi-cuadrada de su denominador. En este caso, sería una t de Student con $n - 3$ grados de libertad y en general, la cantidad de grados de libertad es $n - k$, donde k el número de parámetros a estimar (intercepto más pendientes). Esto es importante tenerlo en cuenta a la hora de buscar los puntos críticos.

Intervalos de Confianza:

Un intervalo de confianza del 95% para el parámetro γ se construye de la siguiente manera:

$$(\hat{\gamma} - |CV_{0.025}|\, se(\hat{\gamma}), \hat{\gamma} + CV_{0.975}\, se(\hat{\gamma}))$$

léase por γ los parámetros α, β (intercepto o pendientes).

Es exactamente como en el caso de regresión simple, pero con la única diferencia de que los grados de libertad de la t de Student de donde se obtienen los valores críticos cambia de la misma manera que para los test de hipótesis (ahora la cantidad de grados de libertad es $n-k$, donde k el número de parámetros a estimar (intercepto más pendientes).

Para "n" grande, uno puede aproximar por la distribución normal de modo que la cantidad de grados de libertad es más relevante cuando el tamaño de la muestra es chico.

Más sobre test de Hipótesis:

Uno puede estar interesado en testear alguna hipótesis que involucre a varios betas:

$$H_0: \beta_1 = \beta_2$$
$$H_A: \beta_1 \neq \beta_2$$

Esto es equivalente a testear:

$$H_0: \beta_1 - \beta_2 = 0$$
$$H_A: \beta_1 - \beta_2 \neq 0$$

U otros tests como:

$H_0: \beta_1 = \beta_2 = 0$
$H_A: \beta_1 \text{ó } \beta_2 \; o \; ambos \; diferentes \; de \; cero$

Estos tests se llevan a cabo con la distribución F. Esta distribución se caracteriza por tener un número de grados de libertad del numerador y otro del denominador. El estadístico se calcula de la siguiente manera:

$$F = \frac{(RRSS - URSS)/q}{URSS/(n-k)}$$

Donde "RRSS" es el "*restricted*" RSS, es decir el RSS que me quedaría si corriera la regresión imponiendo la hipótesis nula, H_0. "URSS" es el "*unrestricted*" RSS, es decir el RSS que me quedaría bajo la hipótesis alternativa, H_A. "n" es el número de observaciones, "k" es número de parámetros a estimar en el modelo no restringido (intercepto más pendientes) y "q" es el número de restricciones bajo H_0. El estadístico F tiene una distribución F con "q" grados de libertad en el numerador y la cantidad "$n-k$" de grados de libertad en el denominador. El valor crítico hay que buscarlo en la tabla F y rechazo H_0 si el estadístico es mayor al valor crítico (que surge de la tabla).

Variables Dummy

Las variables *dummy* son variables que toman dos valores, 0 ó 1. Nosotros las utilizaremos como regresores (variables explicativas) y no como la variable a explicar (que solemos denotar por Y). Por ejemplo, en el "Ejemplo Casa", las variables "Entrada", "Rec", "Sótano", "Calefacción Central", "Aire" y "Vecindario o NBHD" son ejemplos de variables *dummy*. La variable "Garage" no es *dummy*, ya que toma los valores 0, 1, 2 ó 3 y la variable pisos tampoco es *dummy*, ya que toma los valores 1 ó 2.

Nota al pie: supongamos que nuestro modelo fuera:

$$y_i = \alpha + u_i$$

Es fácil mostrar, minimizando la RSS, que el estimador por el método de MCO es $\hat{\alpha} = \bar{y}$.

Definamos las siguientes dos variables *dummy*:

$$M_i = \begin{cases} 1 & si\ la\ observación\ es\ mujer \\ 0 & si\ no \end{cases}$$

$$H_i = \begin{cases} 1 & si\ la\ observación\ es\ hombre \\ 0 & si\ no \end{cases}$$

M_i y H_i son variables *dummy*.

Además, si todas las observaciones en la muestra son clasificadas en alguna de estas dos categorías, se cumple que $M_i + H_i = 1$.

Si nuestro modelo (Modelo 1) ahora fuera:

$$y_i = \beta_0 + \beta_1 M_i + u_i$$

se puede reescribir como:

$$y_i = \begin{cases} \beta_0 + \beta_1 + u_i & si\ \ mujer \\ \beta_0 + u_i & sino \end{cases}$$

De manera que la inclusión de la variable en este modelo estaría permitiendo obtener un intercepto para las observaciones que son hombres (que sería β_0) y otro intercepto para las mujeres (que sería $\beta_0 + \beta_1$). En otras palabras, $\hat{\beta}_0 = \bar{y}_{hombres}$ y por otro lado, $\hat{\beta}_0 + \hat{\beta}_1 = \bar{y}_{mujeres}$, de modo que el beta que premultiplica la *dummy* nos queda, $\hat{\beta}_1 = \bar{y}_{mujeres} - \bar{y}_{hombres}$.

Notemos que en este último ejemplo no hay una variable explicativa “x”.

A continuación, supongamos que nuestro modelo (Modelo 2) fuera:

$$y_i = \gamma_0 + \gamma_1 H_i + u_i$$

se puede reescribir como:

$$y_i = \begin{cases} \gamma_0 + \gamma_1 + u_i & si\ \ hombre \\ \gamma_0 + u_i & sino \end{cases}$$

Siguiendo el mismo razonamiento que en el caso anterior, podemos afirmar que $\hat{\gamma}_0 = \bar{y}_{mujeres}$ y por otro lado, $\hat{\gamma}_0 + \hat{\gamma}_1 = \bar{y}_{hombres}$, de modo que el parámetro que premultiplica la dummy nos queda, $\hat{\gamma}_1 = \bar{y}_{hombres} - \bar{y}_{mujeres}$.

En tercer lugar, planteamos el siguiente modelo (Modelo 3):

$$y_i = \delta_0 M_i + \delta_1 H_i + u_i$$

Este modelo se puede reescribir como

$$y_i = \begin{cases} \delta_0 + u_i & si\ mujer \\ \delta_1 + u_i & si\ hombre \end{cases}$$

Consecuentemente, podemos afirmar que $\hat{\delta}_0 = \bar{y}_{mujeres}$ y por otro lado, $\hat{\delta}_1 = \bar{y}_{hombres}$, de modo que nosotros perfectamente podemos trabajar con los modelos 1 ó 2 ó con el 3. Todo lo que debemos fijarnos es cómo está especificado el modelo para poder interpretar los estimadores de los parámetros correctamente. En este último caso, $\hat{\delta}_1 - \hat{\delta}_0 = \bar{y}_{hombres} - \bar{y}_{mujeres}$.

Por último, consideremos el modelo 4:

$$y_i = \theta_0 + \theta_1 M_i + \theta_2 H_i + u_i$$

Este modelo se puede reescribir como

$$y_i = \begin{cases} \theta_0 + \theta_1 + u_i & si\ \ mujer \\ \theta_0 + \theta_2 + u_i & si\ hombre \end{cases}$$

Siguiendo la misma lógica anterior, podríamos afirmar que

$$\hat{\theta}_0 + \hat{\theta}_1 = \bar{y}_{mujeres}$$

y por otro lado,

$$\hat{\theta}_0 + \hat{\theta}_2 = \bar{y}_{hombres}$$

Pero acá hay un problema: tenemos 3 incógnitas $(\hat{\theta}_0, \hat{\theta}_1\ y\ \hat{\theta}_2)$ y solo dos ecuaciones. De modo que hay infinitas soluciones. Este problema ya lo vimos: acá incurrimos en el problema de multicolinealidad perfecta, toda vez que $M_i + H_i = 1$ guarda una relación lineal perfecta con la variable explicativa que multiplica a θ_0 (que no se ve, pero es como si el θ_0 tuviera una "x" multiplicándola, que toma siempre el valor "1" y es precisamente ésta la que guarda la relación lineal perfecta con la suma de las dos *dummies*).

En este ejemplo es bastante obvio, pero en finanzas muchas veces trabajamos con *dummies* y no es para nada obvio que podemos incurrir en un problema de multicolinealidad.

El modelo 4 está mal especificado.

Transformaciones de variables

Si el modelo que realmente nos interesa para la relación entre Y y X fuera exponencial, es decir, de la forma

$$y_i = A{x_i}^{\beta}u_i$$

no podemos usar MCO directamente porque el modelo no es lineal (en parámetros y variables). Sin embargo, se pueden realizar transformaciones de variables que permiten usar MCO en el modelo transformado. La idea es convertir al modelo en lineal: si tomamos logaritmos, sean

$$\begin{aligned} {y_i}^* &= \log{(y_i)} \\ {x_i}^* &= \log{(x_i)} \end{aligned}$$

$$y_i = A{x_i}^{\beta}u_i \Rightarrow \log(y_i) = A + \beta\log(x_i) + u_i$$

O sea, ${y_i}^* = A + \beta{x_i}^* + u_i$

En este caso, podemos correr una regresión de $y_i{}^*$ en una constante y en $x_i{}^*$ para estimar la constante A y la pendiente β.

La pendiente β, en este ejemplo, en donde la variable dependiente es un logaritmo y la independiente también lo es, se llama "elasticidad". La interpretación del estimador de esta pendiente es diferente a cuando tenemos la regresión simple de la clase 1. El $\hat{\beta}$ se interpreta como en qué porcentaje cambia y_i cuando x_i aumenta en un 1 %. Esto se usa mucho en marketing, para medir el impacto de cambios en precios o de campañas publicitarias.

También está los modelos semilogarítmicos: cuando alguna de las variables aparece en logaritmos, su interpretación es en términos porcentuales, mientras que si aparece en unidades, su interpretación es en unidades.

Ejemplo: sea *wage* el salario y sea *education* los años de educación de una persona; si obtuviéramos la ecuación estimada:

$$\ln(wage) = 0.6 + 0.081\ education$$

El beta sombrero se interpreta de esta manera: un año más de educación aumenta el salario en 8.1%.

El Test F de la salida de los software

Todos los *software* presentan, en general próximo al R^2, un estadístico llamado “F” y un *p-value* asociado a éste. Este test F es el estadístico F tal cual vimos más arriba; pero en particular éste, que reportan todos los software, es el estadístico F asociado a la hipótesis nula que todos los beta son iguales a cero simultáneamente (es decir, todas las pendientes son iguales a cero simultáneamente), contra la hipótesis alternativa, que al menos una de las pendientes (betas) sea distinta de cero. Este estadístico F o su *p-value* asociado es uno de los primeros números que debemos mirar al correr una regresión, ya que si el *p-value* es alto (mayor a 0,10) esto es una indicación de que nuestro modelo estimado no sirve (es decir, no es posible rechazar que todos los betas son cero, lo cual es equivalente a decir que la variable dependiente solo es explicada por la constante alfa más el error).

En otras palabras, este es el test de hipótesis:

$H_0: \beta_1 = \beta_2 = \cdots = 0$
$H_A: al\ menos\ un\ beta\ distinto\ de\ cero$

Bajo la hipótesis nula, el modelo que quedaría consiste en que la variable dependiente es explicada solamente por una constante (el intercepto) más el error. En este caso, el RRSS coincide con el TSS y el estadístico F

$$F = \frac{(RRSS - URSS)/q}{URSS/(n-k)}$$

tendrá q grados de libertad en el numerador, donde q va a ser el número de pendientes (betas) que hay en el modelo original, y $n -$

k es el número de observaciones menos el número de parámetros que hay que estimar en el modelo (número de betas más uno, ya que debemos contar al intercepto dentro del "*k*").

Al mirar el *p-value* ya no es necesario preocuparse por la cantidad de grados de libertad.

Variables relevantes omitidas e irrelevantes incluidas

Recordemos que si nuestro modelo fuera

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

(regresión múltiple), el estimador de MCO de β_1 es:

$$\hat{\beta}_1 = \frac{S_{y1}S_{22} - S_{y2}S_{12}}{S_{11}S_{22} - {S_{12}}^2}$$

y su varianza,

$$Var(\hat{\beta}_1) = \frac{\sigma^2 S_{22}}{S_{11}S_{22} - {S_{12}}^2} = \frac{\sigma^2}{S_{11}(1 - {r_{12}}^2)}$$

Mientras que si el modelo fuera el de regresión simple,

$$y_i = \alpha + \beta_1 x_{1i} + u_i$$

$\hat{\beta}_1 = \frac{S_{y1}}{S_{11}}$ con varianza $Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{11}}$.

Variables relevantes omitidas:

Supongamos que nuestro modelo verdadero (o modelo correcto) es el de regresión múltiple,

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

pero al estimar el modelo omitimos la variable x_2. El estimador que obtendremos de β_1 será, consecuentemente, el de regresión simple: $\hat{\beta}_1 = \frac{S_{y1}}{S_{11}}$. Un poco de álgebra muestra que:

$$\hat{\beta}_1 = \frac{S_{y1}}{S_{11}} = \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2} =$$

$$= \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i)}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

$$= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)x_{2i}}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2} + \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

$$= \beta_1 + \beta_2 \frac{S_{12}}{S_{11}} + \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

Tomando esperanza obtenemos,

$$\mathrm{E}(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{S_{12}}{S_{11}}$$

En otras palabras, el estimador por MCO de β_1 es sesgado cuando omitimos variables relevantes. El sesgo consiste en $\beta_2 \frac{S_{12}}{S_{11}} = \beta_2 \frac{\mathrm{Cov}(x_1, x_2)}{\mathrm{Var}(x_1)}$

Ejemplo: supongamos que tenemos un modelo de ingresos:

$$y_i = \alpha + \beta_1 S_i + \beta_2 a_i + u_i$$

Donde y_i son ingresos, que son explicados por años de educación (*schooling*), S_i, y por habilidad (*ability*), a_i. Suponga que la habilidad no es observada y consecuentemente omitida. Es intuitivo pensar que la habilidad está positivamente correlacionada con años de educación y también que $\beta_2 > 0$.

Por lo anterior, sabemos que

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\mathrm{Cov}(S_i, a_i)}{\mathrm{Var}(S_i)}$$

Y el sesgo será positivo, de modo que al omitir habilidad el valor esperado de nuestro estimador por MCO estará sobreestimando al verdadero parámetro.

Note también que $Var(\hat{\beta}_1) = \frac{\sigma^2}{\mathrm{Var}(S_i)}$ es menor a la varianza del mismo estimador si no se hubiera omitido habilidad, $Var(\hat{\beta}_1) = \frac{\sigma^2}{\mathrm{Var}(S_i)(1-r_{12}{}^2)}$

En conclusión, si omitiéramos una variable relevante, el estimador que obtenemos es sesgado pero tiene menor varianza. Note que esto no contradice Gauss-Markov (ya que no se cumple que ambos estimadores sean insesgados). Cuando los estimadores son sesgados, suele hablarse del *Mean Square Error* del estimador (MSE), en castellano "error cuadrático medio", que se define como la varianza más el sesgo al cuadrado del estimador, cuando nos referimos al concepto de "mejor" estimador, es decir, como criterio para la "eficiencia" del estimador. En otras palabras, con estimadores sesgados comparamos los MSE y no simplemente varianzas.

Inclusión de variables irrelevantes:

Supongamos ahora que nuestro modelo verdadero fuera

$$y_i = \alpha + \beta_1 x_{1i} + u_i$$

pero decidimos incluir una variable irrelevante, x_2 . Por lo tanto, estaríamos en el caso de regresión múltiple y el estimador de β_1 que obtendríamos es

$$\hat{\beta}_1 = \frac{S_{y1}S_{22} - S_{y2}S_{12}}{S_{11}S_{22} - {S_{12}}^2}$$

A partir de esta última expresión, en la clase 4 habíamos desarrollado otra, la de la descomposición del estimador en el parámetro más el error muestral del estimador (*sampling error*):

$$\hat{\beta}_1 = \beta_1 + \frac{S_{22}\sum(x_{1i} - \bar{x}_1)\,u_i - S_{12}\sum(x_{2i} - \bar{x}_2)\,u_i}{S_{11}S_{22} - {S_{12}}^2}$$

Tomando esperanza, podemos ver que $\hat{\beta}_1$ sigue siendo un estimador insesgado de β_1, es decir, $E\left(\hat{\beta}_1\right) = \beta_1$. Sin embargo, $Var(\hat{\beta}_{1\,múltiple}) \geq Var(\hat{\beta}_{1\,simple})$. Es decir, incluir variables irrelevantes arroja estimadores insesgados pero conlleva una pérdida de eficiencia de los estimadores (varianzas más altas).

Notación Matricial:

Sea $\boldsymbol{y}$ un vector de dimensión n, $\boldsymbol{X}$ una matriz de dimensión n x k, sea $\boldsymbol{\beta}$ un vector de parámetros de dimensión k y el error, $\boldsymbol{u}$, un vector de dimensión n.

Nuestro modelo de regresión lineal puede representarse en forma matricial de la siguiente manera:

$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{u}$$

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{k-1,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{k-1,2} \\ 1 & x_{1,3} & x_{2,3} & \cdots & x_{k-1,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & x_{2,n} & \cdots & x_{k-1,n} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{pmatrix}, \qquad \boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix}$$

El objetivo es estimar beta y llamaremos $\widehat{\boldsymbol{\beta}}$ al estimador de $\boldsymbol{\beta}$**.** El valor predicho de $\boldsymbol{y}$ dado $\boldsymbol{X}$ es $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ Los residuos se definen, al igual que antes, como $\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$**.**

El método de Mínimos Cuadrados Ordinarios busca minimizar la suma de residuos al cuadrado, $\boldsymbol{e}'\boldsymbol{e}$.

$$\boldsymbol{e} = \begin{pmatrix} \mathrm{e}_1 \\ \mathrm{e}_2 \\ \mathrm{e}_3 \\ \vdots \\ \mathrm{e}_\mathrm{n} \end{pmatrix}, \ \boldsymbol{e}' = (\mathrm{e}_1 \quad \mathrm{e}_2 \quad \mathrm{e}_3 \quad \cdots \quad \mathrm{e}_\mathrm{n}),$$

$$\boldsymbol{e}'\boldsymbol{e} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \mathrm{e}_{\boldsymbol{i}}^{\mathbf{2}}$$

Minimizar $\boldsymbol{e}'\boldsymbol{e} = \left(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})$ arroja como estimador:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y}$$

Si además suponemos que los supuestos de homocedasticidad y no autocorrelación se cumplen, es decir, $\boldsymbol{\Omega} = E(\boldsymbol{u}\boldsymbol{u}') = \sigma^2 \boldsymbol{I}$, se cumple que

$$Var(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}$$

Ejemplo: supongamos que tenemos que

$$\boldsymbol{y} = \begin{pmatrix} 4 \\ 7 \\ 3 \\ 9 \\ 17 \end{pmatrix},\ \boldsymbol{X} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \\ 1 & 9 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

β_0 es el **parámetro** del intercepto y β_1 el de la pendiente (que previamente los llamamos α y β respectivamente).

Los estimadores que obtemos haciendo las cuentas son:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 1 \\ 1{,}75 \end{pmatrix}$$

El valor predicho de y es

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \\ 1 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 1{,}75 \end{pmatrix} = \begin{pmatrix} 4{,}5 \\ 6{,}25 \\ 2{,}75 \\ 9{,}75 \\ 16{,}75 \end{pmatrix}$$

Y los residuos,

$$\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \begin{pmatrix} 4 \\ 7 \\ 3 \\ 9 \\ 17 \end{pmatrix} - \begin{pmatrix} 4{,}5 \\ 6{,}25 \\ 2{,}75 \\ 9{,}75 \\ 16{,}75 \end{pmatrix} = \begin{pmatrix} -0{,}5 \\ 0{,}75 \\ 0{,}25 \\ -0{,}75 \\ -0{,}25 \end{pmatrix}$$

Note que

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{y} =$$

Note que $X = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \\ 1 & 9 \end{pmatrix}$,

de modo que $X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 5 & 9 \end{pmatrix}$

y consecuentemente, $\boldsymbol{X}'\boldsymbol{X} =$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 5 & 9 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \\ 1 & 9 \end{pmatrix} =$$

$$\begin{pmatrix} 1+1+1+1+1 & 2+3+1+5+9 \\ 2+3+1+5+9 & 4+9+1+25+81 \end{pmatrix} =$$

$$\begin{pmatrix} 5 & 20 \\ 20 & 120 \end{pmatrix}$$

La inversa de esta matriz es

$$(\boldsymbol{X'X})^{-1} = \frac{\mathbf{1}}{\mathbf{det}(\boldsymbol{X'X})}\begin{pmatrix}120 & -20\\ -20 & 5\end{pmatrix} =$$
$$= \frac{\mathbf{1}}{\mathbf{200}}\begin{pmatrix}120 & -20\\ -20 & 5\end{pmatrix} = \begin{pmatrix}0,6 & -0,1\\ -0,1 & 0,025\end{pmatrix}$$

Por otro lado, $\boldsymbol{X'y} =$

$$\begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 2 & 3 & 1 & 5 & 9\end{pmatrix}\begin{pmatrix}4\\ 7\\ 3\\ 9\\ 17\end{pmatrix} =$$
$$= (4+7+3+9+17 \quad 8+21+3+45+153)$$
$$= (40 \quad 230)$$

De modo que,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} =$$
$$\begin{pmatrix}0,6 & -0,1\\ -0,1 & 0,025\end{pmatrix}(40 \quad 230) = \begin{pmatrix}1\\ 1,75\end{pmatrix}$$

Interacciones y No linealidades en las variables

Cambiamos de tema. Dejamos ahora de lado la notación matricial. Supongamos que queremos estimar el siguiente modelo:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

Y la ecuación estimada es:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$$

Notemos que la interpretación de $\hat{\beta}_1$ es la siguiente: por cada unidad que aumente la variable x_{1i}, $\hat{y}_i$ va a aumentar en $\hat{\beta}_1$ unidades, *ceteris paribus*. Matemáticamente, $\hat{\beta}_1$ es la derivada parcial de $\hat{y}_i$ con respecto a x_{1i}, es decir, $\frac{\partial \hat{y}_i}{\partial x_{1i}} = \hat{\beta}_1$.

Sin embargo, a veces encontramos modelos donde una de las variables explicativas aparece "elevada a la uno" pero también aparece "elevada al cuadrado". Por ejemplo, el modelo

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 (x_{2i})^2 + u_i$$

La ecuación estimada será:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 \, x_{1i} + \hat{\beta}_2 \, x_{2i} + \hat{\beta}_3 \, (x_{2i})^2$$

La pregunta que nos formulamos es cómo se interpretan $\hat{\beta}_2$ y $\hat{\beta}_3$. El punto que deben notar es que si x_{2i} cambia, $\hat{y}_i$ se va a ver afectado tanto por el término lineal como por el término cuadrático. La manera de pensar cómo lo afecta es en términos de la derivada parcial, que es:

$$\frac{\partial \, \hat{y}_i}{\partial x_{2i}} = \hat{\beta}_2 + 2\hat{\beta}_3 x_{2i}$$

De manera que en estos modelos, $\hat{\beta}_2$ y $\hat{\beta}_3$ no pueden ni deben interpretarse por separado. Deben interpretarse juntos como la anterior derivada parcial indica. Nótese que en esa derivada parcial debemos establecer un valor de x_{2i} para obtener un número de esa derivada parcial. En general se usa la media o la mediana. Pero perfectamente también podría usarse otro valor si estamos interesados en un valor de x_{2i} en particular. Este es un caso de no linealidades en las variables explicativas. Siempre que hayan no linealidades en las variables explicativas, deben interpretar los coeficientes que están en una misma derivada parcial de manera conjunta.

Interacciones: esta misma idea, de mirar la derivada parcial, debe utilizarse para interpretar modelos en donde aparecen interacciones entre variables explicativas.

Por ejemplo, en el ejemplo casa, suponga que usted quiere saber si un pie cuadrado más de LOTE se valora de la misma manera si la casa queda en un vecindario agradable (NBHD = 1) que cuando no está en un vecindario agradable (NBHD = 0). Uno puede crear una nueva variable explicativa que sea el producto o interacción entre LOTE y NBHD. E incorporarla como otra variable explicativa. Esto es lo que obtendríamos:

Dependent Variable: PRECIO
Method: Least Squares
Sample: 1 546
Included observations: 546

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2899.316	3469.795	-0.835587	0.4038
LOTE	3.181576	0.411037	7.740354	0.0000
CUARTOS	1892.736	1045.809	1.809829	0.0709
BANOS	14390.84	1487.706	9.673172	0.0000
PISOS	6587.293	923.8656	7.130142	0.0000
ENTRADA	7220.913	2065.985	3.495144	0.0005
REC	4664.056	1898.831	2.456277	0.0144
SOTANO	5266.002	1589.118	3.313789	0.0010
CALEF	12975.44	3213.169	4.038208	0.0001
AIRE	12590.64	1552.535	8.109728	0.0000
GARAGE	4142.113	841.2940	4.923502	0.0000
NBHD	2601.531	4341.065	0.599284	0.5492
LOTE_X_NBHD	1.179356	0.698514	1.688380	0.0919

R-squared	0.674863	Mean dependent var	68121.60
Adjusted R-squared	0.667542	S.D. dependent var	26702.67
S.E. of regression	15396.53	Akaike info criterion	22.14519
Sum squared resid	1.26E+11	Schwarz criterion	22.24764
Log likelihood	-6032.638	Hannan-Quinn criter.	22.18524
F-statistic	92.19222	Durbin-Watson stat	1.601231
Prob(F-statistic)	0.000000		

Note que la ecuación estimada es:

$$
\begin{aligned}
\widehat{precio} = -2899 + \ 3{,}18 * LOTE \\
+ \ 1892{,}74 * Cuartos + \cdots \\
+ 2601{,}53 * \ NBHD + 1{,}18 \\
* LOTE * NBHD
\end{aligned}
$$

De modo que el efecto sobre el precio de que el tamaño del lote de la casa aumente en un pie cuadrado será (su derivada parcial):

$$\frac{\partial \widehat{precio}}{\partial LOTE} = 3{,}18 + 1{,}18 * NBHD$$

De modo que si NBHD = 1, $\frac{\partial \widehat{precio}}{\partial LOTE} = 4{,}36$ mientras que si NBHD = 0, $\frac{\partial \widehat{precio}}{\partial LOTE} = 3{,}18$. Lo mismo aplica para la interpretación de cómo cambia el precio predicho de la casa si la casa está en un vecindario agradable en comparación a cuando no lo está. Como NBHD es una variable dummy. En este caso, $\frac{\Delta \widehat{precio}}{\Delta NBHD} = 2601{,}53 + 1{,}18 * LOTE$, si bien el primer estimador no es estadísticamente significativo.

HETEROCEDASTICIDAD

Supongamos que nuestro modelo fuera el de regresión simple,

$$y_i = \alpha + \beta x_i + u_i$$

con $E(u_i) = 0$ pero ahora $Var(u_i) = \sigma_i^2$.

Antes suponíamos que $Var(u_i) = \sigma^2$. Este supuesto (supuesto 2.A) se llama homocedasticidad de los errores (todos los errores tienen la misma varianza). Ahora vamos a suponer que $Var(u_i) = \sigma_i^2$, es decir, se viola el supuesto por el cual los errores son homocedásticos y en este caso decimos que los errores son heterocedásticos.

Sabemos que los estimadores por MCO siguen siendo insesgados, pero ahora no se cumple Gauss-Markov. Es decir, los estimadores por el método de MCO no nos va a dar estimadores que sean MELI (Mejor Estimador Lineal Insesgado). En otras palabras, podríamos encontrar estimadores lineales e insesgados con menor varianza, y esto es importante al momento de realizar tests de hipótesis, por ejemplo, porque uno podría estar llegando a una conclusión equivocada (los errores estándar de los estimadores se ven afectados, y consecuentemente el t-test y el p-value).

Cuando conocemos la varianza de los errores, la manera de resolver el problema de heterocedasticidad es dividir cada una de las variables por el desvío estándar de los errores, σ_i. O, equivalentemente, multiplicando cada término por $\frac{1}{\sigma_i}$.

Entonces, el modelo transformado nos queda

$$y_i \frac{1}{\sigma_i} = \alpha \frac{1}{\sigma_i} + \beta x_i \frac{1}{\sigma_i} + u_i \frac{1}{\sigma_i}$$

Sean:

$$y_i^* = y_i \frac{1}{\sigma_i}$$

$$x_i^* = x_i \frac{1}{\sigma_i}$$

$$u_i^* = u_i \frac{1}{\sigma_i}$$

$$w_i = \frac{1}{\sigma_i}$$

El modelo transformado es:

$$y_i^* = \alpha w_{\ i} + \beta x_i^* + u_i^*$$

¿Por qué hacemos esto? Porque notemos que ahora los errores en el modelo transformado, u_i^*, tienen varianza igual a uno (note que $Var(u_i^*) = Var\left(u_i \frac{1}{\sigma_i}\right) =$

$\left(\frac{1}{\sigma_i}\right)^2 Var(u_i) = 1)$ y consecuentemente son homocedásticos y por lo tanto, podemos usar el método de Mínimos Cuadrados Ordinarios (MCO) en el modelo transformado, y obtendremos estimadores que son MELI. Usar MCO en el modelo transformado es lo que se llama en el contexto de heterocedasticidad "Mínimos Cuadrados Ponderados" (MCP) y en inglés "*Weighted Least Squares*" (WLS). Se llama "ponderado" porque ahora cada observación recibe un peso o ponderación diferente a la hora de calcular el estimador. Cuanto más alto sea el desvío estándar (o varianza) del error de una observación, menor será la ponderación de esa observación.

Cabe mencionar que a veces a este mismo estimador se lo llama estimador por el "Método de Mínimos Cuadrados

Generalizados" (MCG) y en inglés "*Generalized Least Squares*" (GLS). Este término, GLS, es muy amplio. Se usa también para referirse al método por el cual estamos aplicando MCO a un modelo transformado, donde es necesario transformar el modelo para que se satisfagan los supuestos clásicos que introdujimos en las primeras dos clases (supuestos 1, 2(a), 2(b)).

Para encontrar el estimador por mínimos cuadrados ponderados debemos usar, como recién mencionamos, MCO en el modelo transformado: este modelo es

$$y_i{}^* = \alpha w_{\ i} + \beta x_i{}^* + u_i{}^*$$

Note que este modelo difiere del de la clase 2 bajo regresión múltiple, porque si bien hay dos variables explicativas, no hay constante. Haciendo las cuentas, se obtiene (hay un atajo: al resolver el problema de minimización de la clase 2 sin constante, obtiene expresiones muy parecidas, que solamente difieren en que en $S_{y1}, S_{y2}, S_{11}, S_{12}$ y S_{22} debe sacarle la media de cada variable que aparece restando, es decir, en la fórmula reemplace cuando aparece $S_{y1} = \sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(y_i - \bar{y})$ por la expresión $\sum_{i=1}^{n} x_{1i} y_i$.)

Teniendo en cuenta que ahora $y_i^* = y_i w_i$, x_{1i} es w_i y x_{2i} es $x_i^* = x_i w_i$ donde $w_i = \frac{1}{\sigma_i}$, el estimador que obtendrá será:

$$\hat{\beta}_{WLS} = \frac{\sum_{i=1}^{n} y_i x_i \, w_i{}^2 \sum_{i=1}^{n} w_i{}^2 - \sum_{i=1}^{n} y_i \, w_i{}^2 \sum_{i=1}^{n} x_i w_i{}^2}{\sum_{i=1}^{n} w_i{}^2 \sum_{i=1}^{n} x_i{}^2 \, w_i{}^2 - \left(\sum_{i=1}^{n} x_i{}^2 \, w_i{}^2\right)^2}$$

La varianza de este estimador es:

$$Var(\hat{\beta}_{WLS}) = \frac{\sum_{i=1}^{n} w_i{}^2}{\sum_{i=1}^{n} w_i{}^2 \sum_{i=1}^{n} x_i{}^2 w_i{}^2 - \left(\sum_{i=1}^{n} x_i{}^2 w_i{}^2\right)^2}$$

Note que $\hat{\beta}_{WLS}$ es MELI. Sin embargo, $\hat{\beta}_{OLS}$ (donde OLS significa *Ordinary Least Squares* o Mínimos Cuadrados Ordinarios) ¿será insesgado? ¿Será MELI? ¿Por qué? Al no cumplirse el supuesto 2.a, de homocedasticidad, la fórmula de $\hat{\beta}_{OLS}$ sigue siendo la misma, $\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, pero la varianza ya no es la misma que la obtenida cuando se cumple el supuesto 2A, ya que no corresponde hacer los últimos dos pasos en ese cálculo, que suponía que todos los errores tenían la misma varianza y por esa razón sacábamos el sigma cuadrado afuera en el numerador. Ahora,

$$Var(\hat{\beta}_{OLS}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

Tests para detectar heterocedasticidad

1) **Test de White**: primero regresá Y en las X's por MCO y calculá los residuos. Segundo, regresá estos residuos al cuadrado en una constante, las X's, sus cuadrados y sus productos cruzados. Computá el R^2 de esta regresión.

La hipótesis nula es homocedasticidad de los errores y la alternativa es heterocedasticidad. El test de White consiste en computar el estadístico nR^2 donde "n" es el número de observaciones.

Bajo la hipótesis nula, este estadístico tiene asintóticamente una distribución chi-cuadrado con "q" grados de libertad, donde "q" es el número de parámetros en la segunda regresión excluyendo la constante.

$$nR^2 \xrightarrow{a} \mathrm{X}^2_{\mathrm{q}}$$

2) **Test de Goldfeld y Quandt**: supongamos que σ_i^2 estuviera positivamente correlacionada con alguno de los regresores (alguna de las X's). Entonces, primero, reordene las X's por valores, de menor a mayor. Segundo, omití "c" observaciones centrales según ese regresor. Tercero, corra una regresión de Y en las X's para las $(n - c)/2$ observaciones con menores valores según este regresor por el cual las ordenaste. Compute la RSS y llámela RSS_1. Corra otra regresión de Y en las X's para las $(n - c)/2$ observaciones con mayores valores de ese regresor. Compute la RSS y llamela RSS_2 Cuarto, compute el estadístico que consiste en el ratio

$$\frac{RSS_2}{RSS_1}$$

La hipótesis nula es homocedasticidad de los errores y la alternativa es heterocedasticidad.

Bajo la hipótesis nula,

$$\frac{RSS_2}{RSS_1} \sim F_{(n-c-2k)/2,(n-c-2k)/2}$$

esto es, tiene una distribución F con $\frac{n-c-2k}{2}$ grados de libertad en el numerador y $\frac{n-c-2k}{2}$ grados de libertad en el denominador, donde k es el número de variables explicativas en cada regresión.

Generalmente, como "c" se toma $n/3$.

Pregunta: ¿en qué caso sería $\frac{RSS_2}{RSS_1}$ cercano a 1?

3) **Test de Breusch-Pagan**: primero regresá Y en las X's por MCO, computá los residuos y calculá $\tilde{\sigma}^2 = \frac{RSS}{n}$.

Segundo, regresá los residuos al cuadrado divididos por $\tilde{\sigma}^2$ en una constante y en todas las variables que sospeches que generan heterocedasticidad. Computá el ESS (*explained sum of squares*) de esta regresión.

La hipótesis nula es homocedasticidad de los errores y la alternativa es heterocedasticidad.

Bajo la hipótesis nula,

$$\frac{1}{2} ESS \xrightarrow{a} \mathrm{X}^2_{\mathrm{q}}$$

donde "q" es el número de parámetros en la segunda regresión excluyendo la constante.

Ejemplo de Johnston y DiNardo:

EXAMPLE 6.1. TESTS FOR HETEROSCEDASTICITY. The CPS88 data file on the diskette contains a random sample of 1000 observations from the *Current Population Survey*, 1988. The first 100 observations from the file were taken, and a conventional earnings equation was estimated. The results are shown in Table 6.2. The dependent variable is the log of wage (LNWAGE). Years of education are indicated by GRADE. Years of experience and its square are given by POTEXP and EXP2, and UNION is a zero/one dummy variable for membership in a union. The results conform with expectations. Education has a significant positive effect, experience has a quadratic effect, and the union dummy variable has a positive but not very significant coefficient.

To apply the White test for heteroscedasticity to this relation, we need first of all to square the regression residuals. The resultant series is denoted by RESSQ. Next we need to regress RESSQ on the original regressors and their squares and cross products. Taking account of the nature of the specific regressors, there are eight new regressors, noting that the square of the union dummy replicates the original dummy. The new variables are these:

TABLE 6.2
A conventional earnings equation

LS // Dependent Variable is LNWAGE
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	0.595106	0.283485	2.099248	0.0384
GRADE	0.083543	0.020093	4.157841	0.0001
POTEXP	0.050274	0.014137	3.556214	0.0006
EXP2	−0.000562	0.000288	−1.951412	0.0540
UNION	0.165929	0.124454	1.333248	0.1856

R-squared	0.371796	Mean dependent var	2.359213
Adjusted R-squared	0.34534	S.D. dependent var	0.580941
S.E. of regression	0.470043	Akaike info criterion	−1.461153
Sum squared resi	20.9893	Schwartz criterion	−1.330895
Log likelihood	−63.83618	F-statistic	14.05620
Durbin-Watson stat	2.161735	Prob(F-statistic)	0.000000

GRADE2 = GRADE^2 EXP4 = EXP2^2

EXP3 = POTEXP * EXP2 GX = GRADE * POTEXP

GX2 = GRADE * EXP2 GU = GRADE * UNION

XU = POTEXP * UNION XU2 = EXP2 * UNION

The White regression is shown in Table 6.3. The test statistic is $nR^2 = 10.79$ and $\chi^2_{.05}(12) = 21.03$. So the hypothesis of homoscedasticity is not rejected.

To apply the Breusch-Pagan/Godfrey test one must specify the variable or variables that one thinks influence the heteroscedasticity. Selecting GRADE, POTEXP, and UNION as possible candidates gives the regression shown in Table 6.4. From the table, and correcting for the scale factor, $\tilde{\sigma}^2 = 0.2099$, from Table 6.2

$$\frac{1}{2}\text{ESS} = \frac{1}{2}\frac{R^2}{1 - R^2}\text{RSS} = \frac{(0.0428)(10.5480)}{2(0.9572)(0.2099)^2} = 5.35$$

The relevant critical value is $\chi^2_{.05}(3) = 7.815$, so homoscedasticity is not rejected. The alternative test statistic is $nR^2 = 4.28$, which is likewise insignificant.

Finally we illustrate the Goldfeld-Quandt test on these data. We sort the data by POTEXP and take the first and last 35 observations. The ratio of the second RSS to the first RSS is R = 7.5069/7.2517 = 1.06, which is insignificant, since $F_{.05}(30, 30) = 1.84$.

TABLE 6.3
White auxiliary regression

LS // Dependent Variable is RESSQ
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	−0.077672	0.985804	−0.078790	0.9374
GRADE	−0.012200	0.125021	−0.097586	0.9225
POTEXP	0.077838	0.071880	1.082882	0.2819
EXP2	−0.003990	0.004095	−0.974433	0.3325
UNION	0.648787	0.861596	0.753006	0.4535
GRADE2	0.002196	0.004247	0.516939	0.6065
EXP4	−3.34E-07	1.51E-06	−0.220995	0.8256
EXP3	6.17E-05	0.000142	0.434796	0.6648
GX	−0.003752	0.004942	−0.759234	0.4498
GX2	0.000117	0.000111	1.052392	0.2955
GU	−0.051374	0.044304	−1.159596	0.2494
XU	0.001933	0.060614	0.031885	0.9746
XU2	−0.000222	0.001259	−0.176223	0.8605

R-squared	0.107881	Mean dependent var	0.209894
Adjusted R-squared	−0.015170	S.D. dependent var	0.333630
S.E. of regression	0.336151	Akaike info criterion	−2.059652
Sum squared resid	9.830776	Schwartz criterion	−1.720980
Log likelihood	−25.91123	F-statistic	0.876722
Durbin-Watson stat	1.807900	Prob(F-statistic)	0.573082

TABLE 6.4
The Breusch-Pagan/Godfrey test

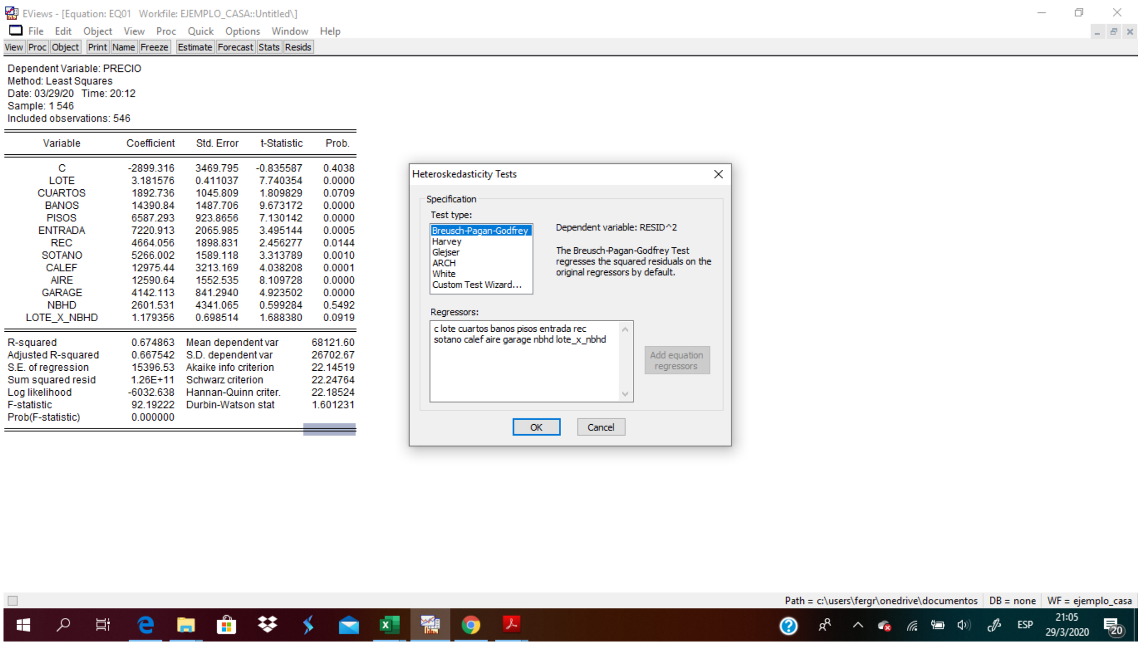
LS // Dependent Variable is RESSQ
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	−0.068446	0.199232	−0.343551	0.7319
GRADE	0.020768	0.013507	1.537566	0.1274
POTEXP	0.002089	0.002770	0.754211	0.4526
UNION	−0.122248	0.083188	−1.469547	0.1450

R-squared	0.042797	Mean dependent var	0.209894
Adjusted R-squared	0.012884	S.D. dependent var	0.333630
S.E. of regression	0.331474	Akaike info criterion	−2.169236
Sum squared resid	10.54798	Schwartz criterion	−2.065029
Log likelihood	−29.43206	F-statistic	1.430730
Durbin-Watson stat	1.791593	Prob(F-statistic)	0.238598

Basta que con uno de los 3 tests de heterocedasticidad rechacemos la hipótesis nula, para que debamos tener en cuenta la presencia de heterocedasticidad. En la práctica, se suele usar Mínimos Cuadrados Generalizados Factibles (*Feasible GLS*) donde el software estima las varianzas de los errores y con esa estimación, estima los parámetros del modelo. FGLS arroja estimadores consistentes. Esto vamos a verlo en las próximas clases.

En Eviews ustedes pueden correr estos tests de heterocedasticidad yendo (con la regresión abierta) a View/Residual Tests/ Heteroskedasticity Tests y ahí se les va a abrir una ventana con todas las opciones de tests que ofrece EViews.



Consistencia

La consistencia es otra propiedad de los estimadores. Decimos que un estimador $\hat{\beta}$ del parámetro β es un estimador consistente de β si el estimador tiende al parámetro cuando el número de observaciones, "n", tiende a infinito, es decir, si $\hat{\beta} \to \beta$ cuando $n \to \infty$. A veces solemos denotar esta propiedad como $plim(\hat{\beta}) = \beta$, donde "*plim*" denota "*probability limit*".

Veamos la definición formal de convergencia en probabilidad:

Definición: Una secuencia de variables aleatorias $\mathrm{X_1, X_2, \dots, X_n}$ converge en probabilidad a una variable aleatoria X si para cada $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

Vamos a mostrar que los estimadores por MCO son consistentes. Es decir, vamos a mostrar que para cada ε > 0,

$$\lim_{n\to\infty} P\left(\left|\hat{\beta} - \beta\right| < \varepsilon\right) = 1$$

En otras palabras, $plim\left(\hat{\beta}\right) = \beta$. Para esto, necesitamos hacer dos supuestos:

(a) $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \to \sigma_x^2$ cuando $n \to \infty$

(b) $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i \to 0$ cuando $n \to \infty$

El supuesto (a) dice que la varianza de la variable explicativa X tiende a una constante σ_x^2 donde $\sigma_x^2 < \infty$. El supuesto (b) indica que la variable explicativa X no está correlacionada en el límite con el término de error, u_i.

La demostración es muy fácil en el caso de regresión simple: usando la descomposición del estimador en el parámetro y el error muestral del estimador,

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\alpha+\beta x_i+u_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
&= \beta + \frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
&= \beta + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}
\end{aligned}$$

Usando la propiedad de plim,

$$plim(\hat{\beta}) = \beta + plim\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$= \beta + \frac{plim\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i\right)}{plim\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)}$$

$$= \beta + \frac{0}{\sigma_x^2} = \beta$$

Notemos que si el supuesto (b) no se cumpliera, es decir, si por ejemplo, $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i \to A \neq 0$ cuando $n \to \infty$, entonces $plim(\hat{\beta}) \neq \beta$ y consecuentemente, $\hat{\beta}$ ya no sería un estimador consistente de β.