Modelado Estocástico

Clase 3

Introducción al Modelo de Regresión lineal

Vamos a comenzar nuestra discusión de análisis de regresión lineal considerando el modelo más simple posible: un modelo lineal con dos variables. Más adelante extenderemos este modelo a más variables. En el caso de dos variables, nuestro objetivo será explicar la variación de “y” basándonos en la variación de “x”. Para ello, es esencial distinguir entre lo que observamos y lo que no es observable por el analista.

Observados: $y_i, x_i \qquad i = 1, \ldots, n$

Supuesto 1: Nuestro modelo que suponemos es:

$$E(y_i) = \alpha + \beta x_i \qquad (1)$$

De modo que:

$$y_i = \alpha + \beta x_i + (y_i - E(y_i))$$

Que lo reescribiremos como

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

Note que u_i es una manera resumida de llamar al término $y_i - E(y_i)$

No observados: α, β, u_i

Como α y β las suponemos idénticas para las distintas observaciones $i = 1, \dots, n$, podemos tener la esperanza de acumular suficiente información como para estimarlas. α y β se llaman **parámetros**. No son variables aleatorias. Son números (poblacionales) desconocidos. Nuestro objetivo consistirá en estimarlos. Pero para ello, necesitamos establecer un método de estimación. El método que vamos a usar se llama "Método de Mínimos Cuadrados Ordinarios" (MCO) y a continuación mencionaremos en qué consiste.

Definición 1: [ERROR] Los u_i se denominan errores. Recordemos que

$$u_i = y_i - E(y_i)$$

Como acabamos de mencionar, los errores no son observables. Los errores son variables aleatorias, pero por el momento no vamos a especificar qué tipo de distribución tienen. Tampoco pretenderemos estimarlos. El error, entonces, es la distancia vertical entre la observación y la recta poblacional que nosotros modelamos en la ecuación (1) (o supuesto 1), que tampoco es observable.

Resumiendo, nuestro objetivo será estimar α y β y llamaremos a los estimadores por MCO de α y β por $\hat{\alpha}$ y $\hat{\beta}$ respectivamente. El valor que predeciremos de y_i dado x_i (en inglés "*predicted value*") a partir de $\hat{\alpha}$ y $\hat{\beta}$ será llamado $\hat{y}_i = \hat{\alpha} + \hat{\beta}\, x_i$ y la diferencia entre y_i y el valor predicho, $\hat{y}_i$, será llamado residuo.

Definición 2: [RESIDUOS] $e_i = y_i - \hat{y}_i$. Los residuos son la diferencia entre el valor de

y_i observado y el predicho, $\hat{y}_i$ (este último es el estimado dado x_i). Gráficamente, si medimos X en el eje horizontal e Y en el de ordenadas, los residuos son las distancias verticales entre el valor observado de y, y_i, y el predicho, $\hat{y}_i$.

Método de Mínimos Cuadrados Ordinarios" (MCO)

Este método consiste en minimizar la suma de residuos al cuadrado, es decir, la suma de las distancias verticales entre y_i e $\hat{y}_i$.

¿Por qué elevamos al cuadrado? ¿Por qué tomamos sumatoria? ¿Daría la misma respuesta si minimizáramos las distancias horizontales? ¿Da lo mismo qué variable tomamos como "y" qué como "x"?

Sea $S(a,b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2$

S es la suma del cuadrado de las distancias verticales como función de los estimadores "a" y "b" de los parámetros α y β. Las condiciones de primer orden (CPO) son:

$$\frac{\partial S}{\partial a} = (-2)\sum_{i=1}^{n}(y_i - a - bx_i)$$

$$\frac{\partial S}{\partial b} = (-2)\sum_{i=1}^{n}(y_i - a - bx_i)\, x_i$$

Igualando las CPO a cero obtenemos las **ecuaciones normales:**

I. $\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = 0$

II. $\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) x_i = 0$

donde $\hat{\alpha}$ y $\hat{\beta}$ son los estimadores por MCO.

Proposición: $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

donde
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Además,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Prueba: $\hat{\alpha}$ es fácil de demostrar de la primera ecuación normal, ya que

$$\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = 0$$
$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\alpha} - \sum_{i=1}^{n} \hat{\beta}x_i = 0$$
$$n\bar{y} - n\hat{\alpha} - \hat{\beta}n\bar{x} = 0$$

De modo que,

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Para hallar $\hat{\beta}$ sustituimos $\hat{\alpha}$ en la segunda ecuación normal:

$$\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)x_i = 0$$

$$\sum_{i=1}^{n} y_i x_i - (\bar{y} - \hat{\beta}\bar{x}\,)\sum_{i=1}^{n} x_i - \hat{\beta}\sum_{i=1}^{n} {x_i}^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i - (\bar{y} - \hat{\beta}\bar{x}\,)n\,\bar{x} - \hat{\beta}\sum_{i=1}^{n} {x_i}^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y} + \hat{\beta}n\,\bar{x}^2 - \hat{\beta}\sum_{i=1}^{n} {x_i}^2 = 0$$

$$\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y} = \hat{\beta}\left[\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2\right]$$

Despejando $\hat{\beta}$ obtenemos que

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

La última igualdad se desprende del hecho que ambos numeradores son iguales y ambos denominadores también son iguales, dado que

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

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

$$= \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i x_i - n\bar{x}\bar{y}.$$

La prueba de que ambos denominadores son iguales es similar a esta última.

■

Notemos que:

1) usando la definición de correlación:

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Si elevamos el coeficiente de correlación al cuadrado:

$$\hat{\rho}_{XY}{}^2 = \left(\frac{\widehat{Cov(X,Y)}}{\hat{\sigma}_X \hat{\sigma}_Y}\right)^2 = \frac{\widehat{Cov(X,Y)}^2}{\widehat{Var(X)}\widehat{Var(Y)}} = \hat{\beta}\,\frac{\widehat{Cov(X,Y)}}{\widehat{Var(Y)}}$$

ya que

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\widehat{Cov(X,Y)}}{\widehat{Var(X)}}$$

donde estas covarianzas y varianzas se refieren a las covarianzas y varianzas muestrales y no a las poblaciones (solemos ponerles un sombrerito a veces para distinguirlas).

2) La recta estimada pasa por el punto de medias $(\bar{x}, \bar{y})$.

3) $\sum_{i=1}^{n} e_i = 0$ y $\sum_{i=1}^{n} e_i \, x_i = 0$. Esto surge claramente de las ecuaciones normales.

4) El signo de $\hat{\beta}$ es igual al signo de la $Cov(X, Y)$.

Ahora bien, hasta acá solo derivamos e igualamos la derivada a cero. ¿Estamos frente a un mínimo o a un máximo?

Proposición: la solución obtenida para los estimadores de MCO efectivamente minimiza la suma de residuos al cuadrado.

Prueba:

Considere los estimadores alternativos a* y b*:

$$S(a^*, b^*) =$$

$$= \sum_{i=1}^{n} (y_i - a^* - b^* x_i)^2$$

$$= \sum_{i=1}^{n} \left[\left(y_i - \hat{\alpha} - \hat{\beta} x_i \right) + (\hat{\alpha} - a^*) + \left(\hat{\beta} - b^* \right) x_i \right]^2$$

$$= \sum_{i=1}^{n} \left[e_i + (\hat{\alpha} - a^*) + \left(\hat{\beta} - b^* \right) x_i \right]^2 =$$

$$= \sum_{i=1}^{n} {e_i}^2 + \sum_{i=1}^{n} 2\,(\hat{\alpha} - a^*) e_i + \sum_{i=1}^{n} 2\left(\hat{\beta} - b^* \right) e_i x_i$$

$$+ \sum_{i=1}^{n} \left[(\hat{\alpha} - a^*) + \left(\hat{\beta} - b^* \right) x_i \right]^2 \geq \sum_{i=1}^{n} {e_i}^2$$

(¿Por qué?) ■

Descomposición de la suma de cuadrados

Se puede descomponer la suma de residuos al cuadrado de la siguiente manera:

$$\sum_{i=1}^{n} e_i{}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 - \hat{\beta}^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$RSS = \quad TSS \quad - \quad ESS$$

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

Note que

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

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

$$ESS = \hat{\beta}^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

En otras palabras, podemos reescribir esta última expresión como TSS = ESS + RSS.

Notemos que la TSS es "n" veces la varianza muestral de la variable "Y", que es la variable que buscamos explicar.

Esta descomposición nos dice que el TSS se puede explicar a través de la ESS (esta es la parte del TSS que es explicada por las X's, es decir, por nuestro modelo de estimación) más otra parte que no puede explicarse con las X's que es la parte residual, es decir, el RSS.

Si dividimos la expresión

$$RSS = \quad TSS \quad - \quad ESS$$

por TSS, obtenemos que

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

$$RSS/TSS = 1 - ESS/TSS$$

Definición: el **coeficiente de determinación** $\mathbf{R^2}$ se define como $\mathbf{R^2 = \frac{ESS}{TSS}}$. Se puede demostrar que, en el caso de regresión simple, el R^2 es igual a la correlación entre X e Y al cuadrado.

En la medida que exista un intercepto en el modelo lineal, el coeficiente de determinación tomará un valor entre 0 y 1 y se lo usa como un coeficiente de bondad de ajuste del modelo estimado. Por ejemplo, si $R^2 = 0{,}3$, podemos decir que la variable X explica el 30% de la variabilidad de la variable dependiente Y.

INSESGADEZ: Decimos que un estimador $\hat{\gamma}$ del parámetro γ es insesgado si la esperanza del estimador es igual al parámetro. Es decir, $\hat{\gamma}$ es insesgado si $E(\hat{\gamma}) = \gamma$. Los estimadores por el Método de Mínimos Cuadrados Ordinarios son insesgados. O sea,

$$E(\hat{\beta}) = \beta$$
$$E(\hat{\alpha}) = \alpha.$$

Para mostrar esto, note que

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

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

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

Recordemos esta descomposición:

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

A veces aparece mencionada como la descomposición del estimador en el parámetro y el error muestral del estimador (*estimator´s* *sampling error*, en inglés). De más está decir que esta descomposición es teórica. No es algo que podamos calcular o corroborar.

Al tomar esperanza,

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

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

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

El primer paso usa la descomposición recién mencionada. El segundo usa el hecho que X viene dada para nosotros, y consecuentemente, el denominador puede ser tratado como una constante. El tercer paso usa esto mismo y el último, el supuesto 1. Queda entonces demostrado que $E(\hat{\beta}) = \beta$.

Se puede demostrar insesgadez del estimador $\hat{\alpha}$ de un modo similar, haciendo uso de la descomposición:

$$\hat{\alpha} - \alpha = \left(\beta - \hat{\beta}\right)\bar{x} + \bar{u},$$

donde $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$.

Introducimos ahora el supuesto 2:

<u>Supuesto 2:</u>

$$\begin{cases} (2a) \qquad Var(u_i) = \sigma^2 \quad \forall i = 1,2,\dots,n \\ \qquad\qquad (Homocedasticidad) \\ (2b) \qquad Cov(u_i, u_j) = 0 \qquad \forall i \neq j \\ \qquad\qquad (No\ autocorrelación) \end{cases}$$

Entonces, podemos calcular la varianza de $\hat{\beta}$:

$$Var(\hat{\beta}) = E\left[\hat{\beta} - E(\hat{\beta})\right]^2 = E\left[\hat{\beta} - \beta\right]^2 =$$

$$= E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^2 =$$

$$= \frac{E[(x_1 - \bar{x})^2 {u_1}^2 + (x_2 - \bar{x})^2 {u_2}^2 + \cdots + (x_n - \bar{x})^2 {u_n}^2]}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

También podemos calcular la varianza de $\hat{\alpha}$:

Notemos que $\hat{\alpha} - \alpha = \left(\beta - \hat{\beta}\right)\bar{x} + \bar{u}$.

Entonces,

$$\begin{aligned} Var(\hat{\alpha}) &= E[\hat{\alpha} - E(\hat{\alpha})]^2 = E(\hat{\alpha} - \alpha)^2 \\ &= E\left[\left(\beta - \hat{\beta}\right)\bar{x} + \bar{u}\right]^2 \\ &= \bar{x}^2 Var\left(\hat{\beta}\right) + \frac{\sigma^2}{n} \end{aligned}$$

Consecuentemente,

$$Var(\hat{\alpha}) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

Finalmente, como $\hat{\alpha}$ y $\hat{\beta}$ son variables aleatorias (ambos dependen de las u_i´s), se puede calcular la covarianza entre los dos estimadores:

$$\begin{aligned} Cov(\hat{\alpha}, \hat{\beta}) &= E\{[\hat{\alpha} - E(\hat{\alpha})][\hat{\beta} - E(\hat{\beta})]\} \\ &= E\{[(\beta - \hat{\beta})\bar{x} + \bar{u}][\hat{\beta} - \beta]\} \\ &= -\bar{x}E[\hat{\beta} - \beta]^2 + E\{[\hat{\beta} - \beta]\bar{u}\} \\ &= -\frac{\bar{x}\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{aligned}$$

El último paso usa que $E\{[\hat{\beta} - \beta]\bar{u}\} = 0$. ¿Por qué?

¿Cómo se interpretan las fórmulas de las varianzas y las covarianzas?

Teorema de Gauss Markov

Los estimadores de mínimos cuadrados ordinarios son los mejores estimadores lineales insesgados (MELI) o BLUE, por sus siglas en inglés (*best linear unbiased estimator*). A continuación mostraremos la prueba para el estimador $\hat{\beta}$.

Prueba: Definamos

$$w_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

de modo que $\hat{\beta} = \sum w_i y_i$

Considere un estimador lineal insesgado alternativo al que llamaremos: $\tilde{\beta} = \sum c_i y_i$

Sin ninguna pérdida de generalidad en esta prueba, escribamos $c_i = w_i + d_i$

Note que al ser insesgado, $\mathrm{E}(\tilde{\beta}) = \beta$, y note que

$$\begin{aligned} E(\tilde{\beta}) &= E\left(\sum c_i y_i\right) \\ &= E\left(\sum c_i(\alpha + \beta x_i + u_i)\right) \\ &= \alpha \sum c_i + \beta \sum c_i x_i \end{aligned}$$

De modo que insesgadez implica que $\sum c_i = 0$ y $\sum c_i x_i = 1$

Notar además que los w_i satisfacen que $\sum w_i = 0$ y que $\sum w_i x_i = 1$ de modo que también debe cumplirse que $\sum d_i = 0$ y que $\sum d_i x_i = 0$.

Entonces,

$$Var(\tilde{\beta}) = E(\sum c_i u_i)^2$$
$$= \sigma^2 \sum c_i{}^2 = \sigma^2 \sum (w_i + d_i)^2 =$$
$$= \sigma^2 \sum (w_i{}^2 + d_i{}^2 + 2w_i d_i)$$

Note también que

$$Var(\hat{\beta}) = E(\sum w_i u_i)^2 = \sigma^2 \sum w_i{}^2$$

Por lo tanto,

$$Var(\tilde{\beta}) - Var(\hat{\beta}) = \sigma^2 \sum \left(d_i{}^2 + 2w_i d_i\right) = \sigma^2 \sum d_i{}^2$$

porque $\sum w_i d_i = 0$

De modo que demostramos que

$$Var(\tilde{\beta}) - Var(\hat{\beta}) \geq 0$$
$$Var(\tilde{\beta}) \geq Var(\hat{\beta}) \blacksquare$$

Un argumento similar se aplica a $\hat{\alpha}$ y a cualquier combinación lineal de $\hat{\alpha}$ y $\hat{\beta}$.

Estimación de la varianza

Es natural usar la RSS para estimar σ^2 (la varianza de los errores)

$$\begin{aligned} e_i = y_i - \hat{y}_i &= y_i - \hat{\alpha} - \hat{\beta} x_i \\ &= (y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x}) \\ &= -(\hat{\beta} - \beta)(x_i - \bar{x}) + (u_i - \bar{u}) \end{aligned}$$

Esto último hace uso de que

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

y consecuentemente, $\bar{y} = \alpha + \beta \bar{x} + \bar{u}$

Vamos a elevar al cuadrado la expresión anterior y sumar de $i = 1 \ldots, n$

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left[-\left(\hat{\beta} - \beta\right)(x_i - \bar{x}) + (u_i - \bar{u})\right]^2$$

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

Tomemos esperanza de esta expresión y para facilitar las cuentas, separemos el lado derecho en 3 términos:

Primer término:

$$E\left[\left(\hat{\beta}-\beta\right)^2\sum_{i=1}^{n}(x_i-\bar{x})^2\right]=$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})^2 E\left[\left(\hat{\beta}-\beta\right)^2\right]$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})^2 Var(\hat{\beta})$$
$$=\sum_{i=1}^{n}(x_i-\bar{x})^2\ \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\sigma^2$$

Segundo Término:

$$E\left(\sum_{i=1}^{n}(u_i-\bar{u})^2\right)=$$
$$=E[(u_1-\bar{u})^2+(u_2-\bar{u})^2$$
$$+\cdots(u_n-\bar{u})^2]$$
$$=E\left[\sum u_i^2+n\bar{u}^2-2\bar{u}\sum u_i\right]$$
$$=n\sigma^2+\sigma^2-2\sigma^2=(n-1)\sigma^2$$

Tercer Término:

$$2E\left[\left(\hat{\beta}-\beta\right)\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\right]$$
$$=2E\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})u_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\right]$$
$$=2E\left\{\frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})u_i\right]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}=2\sigma^2$$

Sumando los términos,

$$E\left(\sum_{i=1}^{n}{e_i}^2\right)=\sigma^2+(n-1)\sigma^2-2\sigma^2=(n-2)\sigma^2$$

Residuos

Es decir, la esperanza del RSS es $(n-2)\sigma^2$

Esto sugiere usar el siguiente estimador, s^2, como estimador de la varianza σ^2:

$$s^2 = \frac{\sum_{i=1}^{n} {e_i}^2}{n-2}$$

de parámetros

Note que $E(s^2) = \sigma^2$. Es decir, s^2 es un estimador insesgado de σ^2. Note también que s^2 es cuadrático en las y_i.

Por ahora, esto es todo lo que podemos decir sin introducir más supuestos.

Introducimos el Supuesto 3:

$$u_i \sim N(0, \sigma^2)\,, \quad i.i.d.$$

Es decir, todos los errores tienen la misma distribución, que suponemos ahora normal, con media 0 y varianza σ^2. Nótese que el supuesto 3 no contradice en nada a nuestro supuesto 1 ($E(u_i) = 0$, la esperanza de los errores sigue siendo cero) ni al supuesto 2 (los errores siguen siendo homocedásticos y no autocorrelacionados). Lo que agrega el supuesto 3 es la distribución de los errores, que ahora suponemos siguen una distribución normal.

Introducir el supuesto 3 tiene los siguientes corolarios:

Como $y_i = \alpha + \beta x_i + u_i$, una consecuencia es que $y_i \sim N(\alpha + \beta x_i, \sigma^2)$ y consecuentemente, como los estimadores por MCO $\hat{\alpha}$ y $\hat{\beta}$ son combinaciones lineales de los y_i y cada uno de los y_i sigue una distribución normal, los estimadores por MCO $\hat{\alpha}$ y $\hat{\beta}$ también van a seguir una distribución normal con la misma media y la misma varianza encontradas previamente.

Más específicamente,

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$\hat{\alpha} \sim N\left(\alpha, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

Distribución del s^2:

Ya vimos que $s^2 = \frac{RSS}{n-2} = \frac{\sum_{i=1}^{n} {e_i}^2}{n-2}$ es un estimador insesgado de σ^2.

Al introducir el supuesto 3, ahora se puede demostrar (no lo vamos a probar) que s^2/σ^2 puede escribirse como la suma del cuadrado de "$n-2$" variables aleatorias normales estándar independientes.

Esto nos permite afirmar que

$$\frac{(n-2)s^2}{\sigma^2} \sim \mathrm{X}^2_{\mathrm{n}-2}$$

Es decir, al introducir el supuesto 3 ahora $\frac{(n-2)s^2}{\sigma^2}$ tiene una distribución chi-cuadrado con $n-2$ grados de libertad.

Una propiedad de las variables aleatorias que tienen una distribución chi-cuadrado es que su esperanza es el número de grados de libertad y su varianza es dos veces el número de grados de libertad.

Es decir,

$$E\left(\frac{(n-2)s^2}{\sigma^2}\right) = n-2$$

$$Var\left(\frac{(n-2)s^2}{\sigma^2}\right) = 2(n-2)$$

La esperanza no nos sorprende porque cancelando los "$n-2$" corroboramos que $E(s^2)=\sigma^2$.

En relación a la varianza, notemos que

$$Var\left(\frac{(n-2)s^2}{\sigma^2}\right)=(n-2)^2 Var\left(\frac{s^2}{\sigma^2}\right)$$
$$=(n-2)^2\frac{1}{(\sigma^2)^2}Var(s^2)$$
$$=\frac{(n-2)^2}{\sigma^4}Var(s^2)$$

De modo que, usando la propiedad de que la varianza de una variable aleatoria chi-cuadrado es 2 veces su número de grados de libertad, obtenemos que

$$\frac{(n-2)^2}{\sigma^4} Var(s^2) = 2(n-2).$$

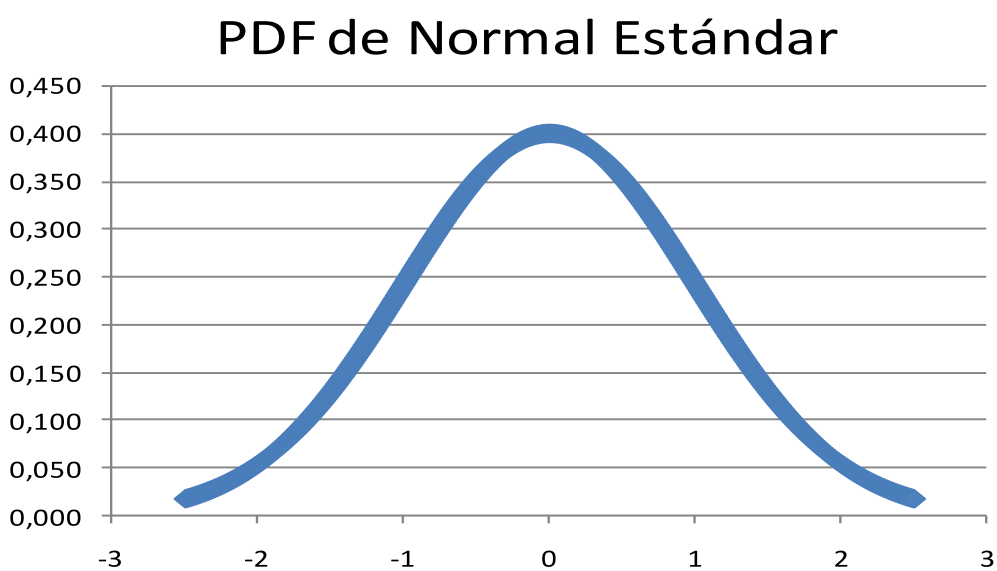
O sea, $Var(s^2) = \frac{2\sigma^4}{n-2}$

Esta es la varianza de nuestro estimador de la varianza, s^2.

Repaso: tabla de la distribución normal estándar. Sea $Z \sim N(0,1)$. Calcule usando la tabla:

- $P(Z < 1{,}96) =$
- $P(Z > 1{,}96) =$
- $P(Z < 1{,}645) =$
- $P(Z < -1{,}645) =$
- $P(-2{,}576 < Z < 2{,}576) =$

La información contenida en la tabla de la normal estándar son los valores que corresponden a la función de distribución acumulada (cdf). Es útil ver un gráfico de la función de densidad (pdf):



La cdf es el área bajo la pdf a la izquierda de cada valor del soporte de la variable aleatoria donde se evalúe.

Inferencia

Como

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$\hat{\alpha} \sim N\left(\alpha, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right),$$

estandarizando, podemos afirmar que:

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim N(0,1)$$

$$\frac{\hat{\alpha} - \alpha}{\sqrt{\sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}} \sim N(0,1)$$

Problema: σ^2 es desconocido. Como no es observable, vamos a tener que estimar la varianza de los errores con nuestro estimador s^2.

Lema:

Sea Z una variable aleatoria normal estándar y sea Y una variable aleatoria chi-cuadrado con v grados de libertad. Si Z e Y son independientes, entonces,

$$t_v = \frac{Z}{\sqrt{Y_v/v}}$$

tiene una distribución t de Student con v grados de libertad.

Proposición:

\# parametros

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

donde t_{n-2} denota una variable aleatoria con distribución t de Student con n-2 grados de libertad.

Prueba: sabemos que $\frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim N(0,1)$ y que

$$\frac{(n-2)s^2}{\sigma^2} \sim \mathrm{X}^2_{\mathrm{n}-2}\ .$$

Por lo tanto, si la normal estándar y la chi cuadrado son independientes, podemos usar el lema, y concluir que

$$\frac{\dfrac{\hat{\beta}-\beta}{\sqrt{\dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}}}{\sqrt{\dfrac{\dfrac{(n-2)s^2}{\sigma^2}}{n-2}}} = \frac{\dfrac{\hat{\beta}-\beta}{\sqrt{\dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}}}{\sqrt{\dfrac{s^2}{\sigma^2}}}$$

$$= \frac{\hat{\beta}-\beta}{\sqrt{\dfrac{s^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}} \sim t_{n-2}$$

Resta demostrar que $\hat{\beta}$ y s^2 son independientes.

Como $\hat{\beta}$ y cada uno de los residuos, digamos el e_j tienen una distribución normal, si podemos mostrar que la covarianza entre ellos es cero, por normalidad, también son independientes. Si $\hat{\beta}$ y e_j son independientes, también lo son $\hat{\beta}$

y e_j^2, y consecuentemente también $\hat{\beta}$ y $s^2 = \frac{\sum_{i=1}^{n} e_i{}^2}{n-2}$.

Ejercicio: mostrar que $Cov(\hat{\beta}, e_j) = 0$.

Ayuda:

$$
\begin{aligned}
Cov(\hat{\beta}, e_j) &= E\left[(\hat{\beta} - \beta)\left(e_j - E(e_j)\right)\right] \\
&= E\left[(\hat{\beta} - \beta)e_j\right] \\
&= E\left[(\hat{\beta} - \beta)(e_j - \bar{e})\right]
\end{aligned}
$$

Y antes de tomar esperanza, haga uso de que

$$e_j - \bar{e} = (\beta - \hat{\beta})(x_j - \bar{x}) + (u_j - \bar{u})$$

Nota al pie: a la expresión $\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$, que es la raíz cuadrada de la varianza estimada del $\hat{\beta}$ se la llama **error estándar** del $\hat{\beta}$ y suele denotarse por $se(\hat{\beta})$.

En general, la raíz cuadrada de la varianza estimada de los estimadores se llama error estándar o *standard error* en inglés y se denota con las letras "se", como por ejemplo, $se(\hat{\alpha})$.

El lema puede usarse de la misma manera para demostrar que

$$\frac{\hat{\alpha}-\alpha}{\sqrt{s^2\left[\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]}} \sim t_{n-2}$$

Resumiendo, sabemos que

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}} \sim N(0,1)$$

Pero como σ^2 es desconocido, usamos s^2 como estimador de σ^2 y sabemos que

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}} \sim t_{n-2}$$

El gráfico de la pdf de la t de Student tiene la misma forma (simétrica) que la de la normal estándar pero tiene colas más pesadas. Cuando la cantidad de grados de libertad tiende a infinito, la t de Student tiende a la normal estándar.

Test de Hipótesis

Supongamos que nuestro modelo es

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

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

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

Este tipo de test de hipótesis se llama a dos colas, porque la hipótesis nula se va a rechazar siempre que encontremos suficiente evidencia estadística para sostener que β es negativo o bien que β es positivo.

Un test de hipótesis a una cola sería, por ejemplo,

$$H_0: \beta \geq 0$$
$$H_A: \beta < 0$$

Hay dos posibles errores que podemos cometer:

	H_0 es verdadera	**H_0 es falsa**
Rechazo H_0	Error Tipo I	✓
No rechazo H_0	✓	Error Tipo II

El error tipo I consiste en rechazar la hipótesis nula cuando la hipótesis nula es verdadera.

El error tipo II consiste en no rechazar la hipótesis nula cuando la hipótesis nula es falsa.

En otras palabras, si el test de hipótesis fuera

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

el error tipo I sería rechazar H_0 cuando $\beta = 0$ y el error tipo II sería no rechazar H_0 cuando $\beta \neq 0$.

El test de hipótesis $H_0: \beta = 0$ nos va a interesar particularmente porque si no rechazamos H_0 estamos diciendo que no hay evidencia para sostener que beta sea distinto de cero y consecuentemente, la variable X no explica a Y.

Pasos para realizar un test de hipótesis:

1) Fijar el nivel de significancia o α (probabilidad de error tipo I con la que va a trabajar)
2) En la tabla que corresponda, encontrar el o los valores críticos (CV), que son los que determinan la región de rechazo/no rechazo de H_0 dado el nivel de significancia (del paso 1)
3) Calcular el estadístico bajo la hipótesis nula (donde β_0 es el valor de β bajo H_0), es decir, calcule $\frac{\widehat{\beta}-\beta_0}{se(\widehat{\beta})}$
4) Concluir si se rechaza o no H_0 en base a si el estadístico obtenido en el paso (2) cae en la región donde se rechaza H_0 o en la región donde no se rechaza H_0 que se obtuvo en el paso (3).

Todos los *software* presentan en la tercera columna el estadístico "t", que básicamente consiste en dividir la primera columna por la segunda, $\frac{\hat{\beta}}{se(\hat{\beta})}$. Éste es, justamente, el estadístico que corresponde utilizar cuando la hipótesis nula es

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

La cuarta columna en las salidas de todos los software es el *p-value* o valor p.

P- Value

El *p-value* es un número que está asociado a un test de hipótesis y se define como la mínima probabilidad de cometer un error tipo I tal que rechazo H_0. Cada test de hipótesis tiene asociado un *p-value*. El *p-value* es siempre un número entre 0 y 1.

La utilidad del *p-value* es que si el *p-value* es menor a la probabilidad de error tipo I con la que decidí trabajar, entonces, sin tener que mirar tabla alguna, puedo decir que rechazo H_0. Caso contrario, no puedo rechazar H_0.

Rechazo H_0 si	p-value $\leq \alpha$
No Rechazo H_0 si	p-value $> \alpha$

Intervalos de Confianza

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

$= \bar{X} \pm T_{n-k,\alpha/2}\ se(\hat{\gamma})$

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

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

Esto se deduce de (para "n" grande),

$$P\left(-1{,}96 < \frac{\hat{\beta} - \beta}{se(\hat{\beta})} < 1{,}96\right) = 0{,}95$$

$$P\left(-1{,}96\ se(\hat{\beta}) < \hat{\beta} - \beta < 1{,}96\ se(\hat{\beta})\right) = 0{,}95$$

$$P\left(-\hat{\beta} - 1{,}96\ se(\hat{\beta}) < -\beta < -\hat{\beta} + 1{,}96\ se(\hat{\beta})\right) = 0{,}95$$

Los valores inferior y superior del intervalo de confianza se reportan en las salidas de los diferentes software.