

CAPÍTULO 3-2

LA INDUCCIÓN ESTADÍSTICA APLICADA A LA REGRESIÓN MÚLTIPLE

Hasta el momento vimos cómo estimar los parámetros de un modelo teórico formalizado con una relación lineal. El método de estimación que decidimos usar, a saber el método de los menores cuadrados, consiste en escoger el valor de los parámetros para minimizar los errores de predicción que se cometen al momento de aplicar el modelo a las observaciones, mismas que sirvieron para la estimación.

Si no se fuera más allá, el análisis de regresión no sería más que un modo de resumir las relaciones que se observan en los datos entre las variables, es decir, el análisis de regresión sería entonces una técnica de estadística descriptiva (es, de por sí, uno de los usos legítimos del análisis de regresión). Sin embargo, por lo general, el análisis de los datos tiene como fin descubrir la relación subyacente cuyos parámetros son desconocidos pero que sirve para relacionar las variables entre sí.

Recordemos el contexto en el cual situamos el análisis de regresión. De principio, admitimos que el modelo (determinista) del cual pretendemos estimar los parámetros no es más que una aproximación de la realidad; subsiste una diferencia entre las predicciones del modelo y las observaciones. Se representa este “error” no sistemático con el término aleatorio

de la relación. El objetivo que se pretende es reconocer el valor de los parámetros (coeficientes) de la relación. Pero no es posible observar directamente estos parámetros. Se revelan indirectamente a través de los valores observados de las variables que el modelo relaciona. El problema es que, en las observaciones, la relación se ve afectada a causa del término aleatorio cuyo valor no es más observable que el valor de los parámetros de la relación.

En estas condiciones, sólo los métodos de la inducción estadística pueden permitirnos limitar la incertidumbre que afecta la estimación de los parámetros. Asimismo, la aplicación de estos métodos exige que completemos el modelo aleatorio con asociar al modelo determinista un modelo de muestreo, el cual será un modelo de relación aleatoria entre la muestra y la población. ¿Qué constituye la población y qué constituye la muestra en la regresión múltiple? En este contexto, se considera que el valor (inobservable) del término aleatorio asociado a cada observación es una muestra de tamaño 1 que se sorteó del conjunto infinito de valores posibles que podría tomar el término aleatorio en este caso; este conjunto infinito de valores posibles constituye una población. Por lo tanto, con n observaciones, existen n muestras de tamaño 1 que se sortea de n poblaciones. No se excluye, al principio, que estas n poblaciones sean diferentes entre sí; no se excluye tampoco que sean idénticas (lo que equivale a decir que los valores de los términos aleatorios se sortean dentro de una misma población). ¿Suponemos que las n poblaciones son idénticas? La respuesta forma parte del modelo de muestreo. Sin embargo, el modelo de muestreo se constituye sobre la base de hipótesis sobre las distribuciones de probabilidad de los términos aleatorios. Emitir la hipótesis que estas distribuciones de probabilidad son idénticas equivale a emitir la hipótesis de que las poblaciones son idénticas.

Un poco más tarde examinaremos nuevamente las hipótesis que constituyen el modelo de muestreo de la regresión li-

neal clásica. Para estudiar los procedimientos de inducción estadística, sólo basta, de manera provisoria, saber que en el marco de este modelo de muestreo es posible estimar los parámetros de las distribuciones de muestreo de los estimadores de los parámetros.

1. El estimador b_j del parámetro β_j es no sesgado. En otras palabras, el estimador de los menores cuadrados posee una distribución de muestreo cuyo promedio es igual al valor del parámetro.
2. Existe también un estimador no sesgado de la varianza de muestreo $\sigma_{b_j}^2$ de cada uno de los coeficientes estimados b_j , y de la covarianza de muestra de cada par de coeficientes estimados, $\sigma_{b_j b_h}$.

Simbología:

$$s_{b_j}^2 = \text{valor estimado de la varianza de muestreo } \sigma_{b_j}^2.$$

Esos valores estimados se entregan en los paquetes de estadística.

3-2.1 UNOS EJEMPLOS DE PRUEBAS DE HIPÓTESIS

3-2.1.1 Test bilateral de una hipótesis simple sobre el valor de un coeficiente (test de Student)

Queremos probar una hipótesis simple del tipo:

$$H_0 : \beta_j = c$$

Por ejemplo, Lemelin y Polèse (1995) estimaron los parámetros del modelo siguiente:

$$\ln PURB = \beta_1 + \beta_2 \ln PTOT + \beta_3 \ln GNPC + \beta_4 (\ln GNPC)^2$$

Este modelo equivale a

$$\ln\left(\frac{PURB}{PTOT}\right) = \ln PURB - \ln PTOT$$

$$\ln\left(\frac{PURB}{PTOT}\right) = \beta_1 + (\beta_2 - 1)\ln PTOT$$

$$+ \beta_3 \ln GNPC + \beta_4 (\ln GNPC)^2$$

Queremos probar la hipótesis de que el coeficiente β_2 es igual a uno:

$$H_0 : \beta_2 = 1$$

¿Porqué esta hipótesis? Porque, en caso que esta hipótesis sea verdadera, esto significa que el grado de urbanización $\frac{PURB}{PTOT}$ es independiente de la población total; en otras palabras, poco importa el tamaño de la población de un país, se determina la fracción de la población que vive en zona urbana con el PIB por cápita. Con un ejemplo concreto, si $\beta_2 = 1$, el modelo predice que el grado de urbanización de China con 1.1 billones de habitantes en 1990 es el mismo que el grado de urbanización de Kenia, que cuenta con 24 millones de habitantes, porque los dos países poseen un PIB por capita de US\$370.00 (vea la tabla 1 de Lemelin y Polèse, 1995). ¿Podemos rechazar esta hipótesis?

Recordemos los pasos a seguir para efectuar un test de probabilidad crítico (sin umbral de significación predefinido – p-value test):

1. Escoger una variable test.
2. Verificar que el modelo de muestreo asociado a esta variable test sea aceptable.
3. Calcular el valor de la variable test.
4. Determinar el valor de la probabilidad crítica correspondiente;

5. Tomar la decisión de rechazar o no la hipótesis según juzguemos que esta probabilidad crítica es lo suficiente pequeña o no (cuanto más la probabilidad crítica es pequeña, menos las observaciones son compatibles con la hipótesis y más el rechazo puede ser categórico).

Vamos a aplicar el test de Student que, en caso de un test de hipótesis simple sobre un coeficiente de regresión, usa la variable test siguiente:

$$t_{n-k} = \frac{b_j - c}{s_{b_j}}$$

donde b_j es el valor estimado del parámetro β_j y s_{b_j} es el valor estimado de la diferencia type de muestreo de b_j . Se puede observar una analogía evidente entre esta variable test y la variable test que se usa para el test de una hipótesis simple sobre un promedio:

$$t_{n-1} = \frac{m_x - \gamma}{\left(\frac{s_x}{\sqrt{n}} \right)}$$

donde el denominador $\left(\frac{s_x}{\sqrt{n}} \right)$ es la desviación estándar de muestreo del promedio.

La selección de esta variable test se justifica porque, en las condiciones del modelo clásico de la regresión lineal normal, la variable $\frac{b_j - \beta_j}{s_{b_j}}$ posee una distribución de Stu-

dent con $n - k$ grados de libertad, cuando n es el número de observaciones y k , el número de variables independientes del modelo (incluyendo la constante).

En el caso que nos interesa, el valor de la variable test se define con:¹⁵⁸

$$t_{n-k} = \frac{0.971663-1}{0.0279321} = -1.0145$$

El número de observaciones n es igual a 64 (Lemelin y Polèse, 1995, tabla 2); el número de variables independientes k es igual a 4. Tenemos, por lo tanto, 60 grados de libertad. La probabilidad crítica asociada a este valor para el test bilateral es de 0.314 o 31.4% (esta probabilidad crítica se calculó con la función TDIST del programa Excel;¹⁵⁹ Lemelin y Polèse (1995, p. 322) presentan los resultados de un test equivalente).

Al menos de escoger un umbral de significación muy alto (superior a 0.314), no se puede rechazar la hipótesis que $\beta_2 = 1$. Por tanto, la hipótesis no rechazada tampoco es probada. Sin embargo, es legítimo mantenerla.

3-2.1.2 Test de hipótesis de un coeficiente nulo

Sucede, a menudo, que queramos probar la hipótesis

$$H_0: \beta_j = 0$$

Por ejemplo, en el modelo

$$\ln PURB = \beta_1 + \beta_2 \ln PTOT \\ + \beta_3 \ln GNPC + \beta_4 (\ln GNPC)^2$$

el valor estimado del parámetro β_4 que presentan Lemelin y Polèse (1995, tabla 2) es de -0.045 . Este valor parece peque-

¹⁵⁸ Los valores que se emplean para el cálculo que sigue provienen directamente de los resultados de computadoras, los cuales se reproducen en el anexo 3-A. Sin embargo, no se encuentran en su totalidad en Lemelin y Polèse (1995).

¹⁵⁹ Note que el valor de la estadística t sirve de argumento en la función TDIST no puede ser negativo. El usuario debe, por lo tanto, tomar en cuenta que la distribución de Student es simétrica.

ño sin embargo ¿es, “de manera significativa, diferente de cero”? En otras palabras, ¿es posible rechazar la hipótesis de que este coeficiente sea nulo y se pudiera entonces quitar la variable correspondiente?

Para probar este tipo de hipótesis, sólo basta aplicar el test que describimos con anterioridad, pero con $c = 0$. La variable-test toma, entonces, la forma siguiente:

$$t_{n-k} = \frac{b_j - c}{s_{b_j}} = \frac{b_j}{s_{b_j}} \text{ cuando } c = 0$$

En nuestro ejemplo,¹⁶⁰

$$t_{64-4} = \frac{-0.0453}{0.01345} = -3.368$$

La probabilidad crítica asociada a este valor en un test bilateral con 60 grados de libertad es de 0.0013 o 0.13% (esta probabilidad crítica se calculó con la función TDIST del programa Excel). Con una probabilidad crítica tan pequeña, es casi imposible no rechazar la hipótesis. Diremos que el coeficiente es, de manera significativa, diferente de cero con un umbral de significación de menos de 1%, con más precisión de 0.0013, lo que representa un tanto más que 0.1%.

Es tan común probar este tipo de hipótesis que los paquetes de aplicación lo efectúan automáticamente: es el “*t*” que reportan los logiciales de aplicación estadística. Los logiciales procuran, también, el valor crítico correspondiente.

Las tablas de resultados de los artículos científicos presentan también, una evaluación del grado de significación de cada coeficiente. Lemelin y Polèse (1995) reportan la pro-

¹⁶⁰ Los valores que se emplean para el cálculo que sigue provienen directamente de los resultados de computadoras, los cuales se reproducen en el anexo 3-A. Sin embargo, no se encuentran en su totalidad en Lemelin y Polèse (1995).

babilidad crítica. Richardson et al. (1990) dan el valor del t de Student y Heikkikala et al. (1989) presentan los dos.

Algunos autores dan la desviación estándar de cada coeficiente estimado (de por sí, al conocer el valor estimado del coeficiente, es posible calcular su desviación estándar a partir de su estadística t y viceversa, puesto que

$t_{n-k} = \frac{b_j}{s_{b_j}}$). Cuando se presenta únicamente la desviación

estándar o el valor t de Student, se identifica, por lo general, por medio de llamadas, los coeficientes que son, de manera significativa, diferentes de cero con un umbral de significación de 1%, de 5% o de 10%.

3-2.1.3 Test unilateral de una hipótesis simple sobre el valor de un coeficiente (test de Student)

En algunas ocasiones es pertinente aplicar un test unilateral (vea capítulo 2-3). Por ejemplo, el modelo

$$PLAR_i = K PURB_i^h$$

predice que la ciudad más grande de un país crece más rápido o menos rápido que el resto de la población urbana dependiendo de si el valor del exponente, el parámetro h , es superior o inferior a 1 (con la intención de facilitar las referencias al artículo, guardamos, en este momento, la misma simbología). En particular, si $h < 1$, el modelo predice que el peso relativo de la ciudad más grande disminuye a medida que crece la población urbana. ¿Es posible rechazar esta hipótesis de que $h \geq 1$?

Lemelin y Polèse (1995, tabla 2) estimaron los parámetros del modelo una vez después de haber sufrido una transformación lineal

$$\ln PLAR_i = \ln K + h \ln PURB_i$$

El valor estimado de h es 0.636. Se efectúa el test unilateral de la hipótesis $H_0: h = 1$ con, como hipótesis complementaria, $H_A: h < 1$.

La zona de rechazo se sitúa, por lo tanto, a la izquierda, es decir que, para rechazar H_0 y aceptar $1 < h$, es necesario que la diferencia $1 - h$ sea lo suficiente grande como para juzgar muy improbable que $h \geq 1$. La variable test es, nuevamente, el t de Student.¹⁶¹

$$t_{64-2} = \frac{h-1}{s_h} = \frac{0.636-1}{0.0426} = -8.54$$

Con $n - k = 62$, la probabilidad crítica unilateral asociada a un valor absoluto tan grande del t de Student es menor que 0.0001.¹⁶² Podemos, pues, decidir con toda confianza rechazar H_0 y aceptar la hipótesis de que la importancia relativa de la ciudad más grande disminuye a medida que la población urbana crece (lo que sorprenderá a más de uno...).

3-2.1.4 Intervalos de confianza y márgenes de error

Es obvio que, como en el caso del promedio, la variable test t_{n-k} puede emplearse también para definir intervalos de confianza del tipo

$$b_j - s_{b_j} \theta_{n-k}(\alpha) < \beta_j < b_j + s_{b_j} \theta_{n-k}(\alpha)$$

¹⁶¹ Los valores que se emplean para el cálculo que sigue provienen directamente de las salidas de computadoras, las cuales se reproducen en el anexo 3-A. Sin embargo, no se encuentran en su totalidad en Lemelin y Polèse (1995).

¹⁶² Para $t = 4$, la función *TDIST* del logicial Excel da una probabilidad crítica unilateral de 0.0000857. Abajo de esta probabilidad, la función *TINV* empieza a arrojar resultados aberrantes. Nada nos garantiza que la función *TDIST* nos dé resultados válidos para valores de t superiores a 4. Por lo tanto, es preferible y además suficiente, cerciorarse de que la probabilidad crítica asociada a $t = 8.54$ sea inferior a 0.0001.

con un nivel de confianza de $(1-\alpha)$. El margen de error correspondiente, con el mismo nivel de confianza, es igual a

$$\pm s_{b_j} \theta_{n-k}(\alpha)$$

Por ejemplo, calculemos un intervalo de confianza del parámetro β_4 en el modelo

$$\ln PURB = \beta_1 + \beta_2 \ln PTOT \\ + \beta_3 \ln GNPC + \beta_4 (\ln GNPC)^2$$

Con $n - k$ grados de libertad y un nivel de confianza de 0.99 (99%), los valores críticos del t de Student son -2.66 y $+2.66$ [$\theta_{64}(0.01) = 2.66$; valores que se calcularon con la función *TINV* del logicial Excel. Dado que $b_j = -0.0453$ y $s_{b_j} = 0.01345$ el intervalo de confianza de β_4 con 99% se define con:

$$-0,0453 - (0,01345 \times 2,66) < \beta_4 < -0,0453 + (0,01345 \times 2,66) \\ -0,0811 < \beta_4 < -0,0095$$

y el margen de error con un nivel de confianza de 99% es igual a

$$\pm 0,01345 \times 2,66 = \pm 0,358$$

En este momento, se deducen los intervalos de confianza igual como en el caso de un test de hipótesis simple sobre un promedio; el conjunto de hipótesis que no se rechazarían con un nivel de significación de α se define con

$$-\theta_{n-k}(\alpha) < \frac{b_j - c}{s_{b_j}} < +\theta_{n-k}(\alpha) \\ -\theta_{n-k}(\alpha) s_{b_j} < (b_j - c) < +\theta_{n-k}(\alpha) s_{b_j} \\ -b_j - \theta_{n-k}(\alpha) s_{b_j} < -c < -b_j + \theta_{n-k}(\alpha) s_{b_j} \\ +b_j + \theta_{n-k}(\alpha) s_{b_j} > +c > +b_j - \theta_{n-k}(\alpha) s_{b_j} \\ b_j - \theta_{n-k}(\alpha) s_{b_j} < c < b_j + \theta_{n-k}(\alpha) s_{b_j}$$

3-2.1.5 Test de una o varias relaciones lineales entre coeficientes (Test F de Fisher)

El test de Student permite examinar una sola hipótesis al mismo tiempo; además esta hipótesis se emite únicamente con relación a un solo coeficiente. El test de Fisher es mucho más polivalente; de hecho, permite examinar varias hipótesis al mismo tiempo y permite examinar hipótesis que asocian más de un coeficiente. Como en el caso del test de Student, el test de Fisher exige que se respeten las condiciones del modelo clásico de la regresión lineal normal. No detallaremos, en este momento, la mecánica del test de Fisher, sin embargo daremos a continuación algunos ejemplos de su uso.

No obstante, recordemos que la variable test de Fisher se calcula por medio de la fórmula siguiente:

$$F_{p,n-k} = \frac{\left(\frac{SSR_H}{p} \right)}{\left(\frac{SSR}{n-k} \right)}$$

donde p es el número de las restricciones lineales simultáneas que constituyen la hipótesis y SSR_H es la suma de los cuadrados de los residuos que se obtuvieron bajo las restricciones (es decir, cuando estimamos los parámetros del modelo forzándolo a respetar la hipótesis).

De esta manera, Lemelin y Polèse (1995) estimaron los parámetros de los modelos siguientes (para facilitar las referencias al artículo, usamos en este momento la misma simbología):

$$\ln PLAR = p' + q' \ln PTOT \\ + r' \ln GNPC + t' (\ln GNPC)^2 + s \ln PURB$$

$$\ln PLAR = \ln K + h \ln PURB$$

La hipótesis que queremos probar que es aceptable (o sea no rechazado) dejar de lado las variables que están ausentes en la segunda ecuación. De hecho, esta hipótesis la constituyen tres hipótesis simples:

$$H_1: q' = 0$$

$$H_2: r' = 0$$

$$H_3: t' = 0$$

Las probabilidades críticas que podemos encontrar en la tabla 2 de Lemelin y Polèse (1995) permiten concluir que, consideradas una por una, ninguna de estas hipótesis se puede rechazar: para H_1 , la probabilidad crítica es de 0.484; para H_2 , es de 0.189; para H_3 , es de 0.173. Pero, cada uno de estos tres test de hipótesis se basa en un modelo de muestreo donde aparecen las otras dos variables: en el primer caso, por ejemplo, dado que el modelo contiene las variables $\ln GNPC$ y $(\ln GNPC)^2$, sólo es posible rechazar la hipótesis que $q' = 0$. ¿Qué sucede, entonces, con la hipótesis que los tres coeficientes sean nulos al mismo tiempo? Es justamente este tipo de hipótesis de que el test de Fisher permite examinar. Lemelin y Polèse (1995, p. 323) reportan que la aplicación de este test arroja una probabilidad crítica de 0.53 para la hipótesis que q' , r' , y t' y son nulos de manera simultánea, por lo tanto no es posible rechazar esta hipótesis.

Consideremos, ahora, la función de producción Cobb-Douglas:

$$Y = A K^B T^C$$

donde

Y es la cantidad producida;

K es la cantidad de capital empleada;

T es la cantidad de mano de obra empleada;

A , B y C son los parámetros;

Al momento de aplicar la transformación logarítmica, el modelo se vuelve lineal:

$$\log Y = \log A + B \log K + C \log T$$

Una de las hipótesis que queremos examinar en este modelo es la siguiente:

$$H_0: B + C = 1$$

Esta hipótesis tiene un especial interés porque, si $B + C = 1$, se trata de una función de producción con rendimientos constantes a la escala. Esto significa que, si aumentamos (o disminuimos) todos los factores de producción de manera proporcional, entonces la producción aumenta (o disminuye) en la misma proporción.

No es tan complicado demostrar esta propiedad. Teniendo K_0 , T_0 y Y_0 , los valores iniciales de K , T y Y , y teniendo λ la proporción según la cual aumentamos las cantidades de los factores, entonces

$$Y = A (\lambda K_0)^B (\lambda T_0)^C = \lambda^{B+C} A K_0^B T_0^C = \lambda^{B+C} Y_0$$

y si $B + C = 1$,

$$Y = \lambda Y_0$$

No es posible aplicar el test de Student a la hipótesis “ $H_0: B + C = 1$ ”, porque no se trata de una hipótesis con un parámetro sino, más bien, de una hipótesis sobre una relación entre dos parámetros. Sin embargo, el test de Fisher permite examinar este tipo de hipótesis.

Generalizando y de manera más formal, el test de Fisher permite probar cualquier hipótesis que podamos expresar con una o varias restricciones lineales con relación a los coeficientes. Una restricción lineal con relación a los coeficientes $\beta_1, \beta_2, \beta_3$, etc. se escribe como sigue:

$$\sum_{j=1}^k w_j \beta_j = w_1 \beta_1 + w_2 \beta_2 + \dots + w_k \beta_k = c$$

donde c y los w_j son constantes que el usuario definió en función de la restricción que desea representar.

La hipótesis que más se acostumbra probar con el test de Fisher (pero no forzosamente la más interesante) es:

$$H_0: \beta_2 = \beta_3 = \dots \beta_k = 0$$

Es la hipótesis de que todos los coeficientes de la regresión, menos la constante β_1 , sean nulos; se constituye, por lo tanto, esta hipótesis con $(k-1)$ hipótesis simples. Dicho de otra manera, es la hipótesis de que el “verdadero” valor del coeficiente de determinación múltiple R^2 es cero y que el R^2 calculado no es más que la correlación fortuita que los términos aleatorios causaron. Los logicales de aplicación estadística proporcionan de manera automática el valor de la variable-test asociada a esta hipótesis.

3-2.2 ESPECIFICACIÓN DE UN MODELO DE MUESTREO: LAS CONDICIONES DEL MODELO CLÁSICO DE REGRESIÓN LINEAL NORMAL

Hicimos hincapié de que la validez de los test de hipótesis que acabamos de describir depende de la validez del modelo de muestreo sobre el cual se fundamentan. Por lo tanto les vamos a echar un ojo a continuación.

3-2.2.1 El modelo clásico de la regresión lineal*

El modelo de muestreo clásico se constituye de cuatro hipótesis. Estas hipótesis son bastante generales, en el sentido de que imponen pocas restricciones para la forma general de la distribución de probabilidad del término aleatorio.

Las dos primeras hipótesis tratan con los parámetros de las distribuciones de probabilidad de los términos aleatorios:

H1: Para cada observación, el valor del término aleatorio es sorteado de una población teórica de promedio nulo; por consiguiente: $E(u_i) = 0$ para todo i .

H2a): Para todas las observaciones, las poblaciones teóricas de donde se sortean los valores de los términos aleatorios tienen la misma varianza:¹⁶³
 $\sigma_i^2 = \sigma^2$ para todo i .

H2b): Para cada observación, el valor del término aleatorio es estadísticamente independiente de los valores de los términos aleatorios de las demás observaciones:¹⁶⁴
 $\sigma_{ij} = 0$ para todas las combinaciones i, j cuando $i \neq j$.

Veremos, con más detalle, lo que significan estas condiciones al momento de examinar lo que sucede cuando no se respetan. La tercera hipótesis circunscribe el papel de lo aleatorio en el modelo:

H3: Las variables independientes x_{ij} son no aleatorias.

* Referencias: Kennedy (1992, pp. 43-45).

¹⁶³ Esta propiedad se llama “homoscedasticidad” (de la palabra griega ομοσ, igual y σκεδασις, dispersión). Lo contrario es “heteroscedasticidad” (de la palabra griega ετερος, otro).

¹⁶⁴ Esta propiedad se llama la ausencia de autocorrelación.

La hipótesis H3 exige, en particular, que se midan los valores de las variables independientes sin error. Entre las otras situaciones que no son compatibles con esta condición, mencionemos la presencia de los valores atrasados de la variable dependiente del lado de las variables independientes, como sucede, por ejemplo, en un modelo donde la tasa de desempleo de cada mes depende, entre otras cosas, de la tasa de desempleo del mes anterior (un modelo del tipo $C_t = a + bC_{t-1} + \dots$).

Mencionemos también los sistemas de ecuaciones simultáneas donde la variable dependiente de una ecuación aparece entre las variables independientes de otra (por ejemplo, un modelo donde el PIB depende del consumo y los demás componentes de la demanda mientras que el consumo depende del PIB: $Y = C + X$ y $C = a + bY$). En estas circunstancias, se deben adaptar los métodos con el fin de tomar en cuenta el no respeto de H3.

Finalmente, la cuarta hipótesis se refiere a las relaciones entre las variables independientes y al número de observaciones.

H4: Existe menos parámetros para estimar que observaciones y no hay redundancias entre las variables independientes.¹⁶⁵

“No hay redundancias” significa, aquí, que ninguna variable independiente puede representarse como una combinación de las demás; es decir, las variables independientes son linealmente independientes entre sí. H4 no es tanto una hipótesis como una condición de aplicación, puesto que se puede determinar con un análisis de datos si esta condición se respeta.

¹⁶⁵ Técnicamente, esto se traduce por la condición que el rango de la matriz X de orden $n \times k$, sea igual a $k < n$.

Al momento de combinar las hipótesis H1 hasta H4 con el modelo lineal general que se expuso en el apartado 1, definen un modelo aleatorio que acostumbramos designar como el “modelo clásico de la regresión lineal”. La especificación de este modelo de muestreo es, no obstante, incompleta puesto que el tipo de la distribución de probabilidad de los términos aleatorios no se define. Sin embargo, cuando las hipótesis H1 hasta H4 se respetan, el estimador de los menores cuadrados ordinarios posee varias propiedades deseables. Se demuestran estas propiedades en el teorema de Gauss-Markov.

3-2.2.2 Propiedades del estimador de los menores cuadrados bajo el modelo clásico de la regresión lineal: el teorema de Gauss-Markov

En este momento, nos contentaremos de enunciar, sin demostrarlas, las principales conclusiones del teorema de Gauss-Markov¹⁶⁶. Estas condiciones se refieren, en particular, a los parámetros de la distribución de muestreo de los coeficientes estimados b_j . Este teorema establece, por lo tanto, los fundamentos de los tests de hipótesis aplicables a los coeficientes estimados b_j .

Cuando las hipótesis H1 hasta H4 se respetan, entonces los resultados del método de los menores cuadrados tienen las propiedades siguientes:

1. El método de los menores cuadrados produce estimadores no sesgados de los parámetros β_j . En otras palabras, cada uno de los coeficientes estimados b_j posee una distribución de muestreo cuyo promedio (la espe-

¹⁶⁶ El matemático Carl Friedrich Gauss (1777-1855) es el inventor de la distribución normal y del método de los menores cuadrados (en 1794 Gauss aplicó el método por primera vez para la estimación en 1801 de la trayectoria del asteroide Ceres); Andreï Andrelevitch Markov (1856-1922) es, en particular, el autor de un teorema límite central.

ranza matemática) es igual al “verdadero” valor del coeficiente β_j .

2. El método de los mínimos cuadrados produce, también, un estimador no sesgado de σ^2 el cual corresponde a la varianza común de los términos aleatorios. La variable aleatoria

$$\frac{1}{n-k} \sum_i (y_i - \hat{y}_i)^2 = \frac{SSR}{n-k}$$

es un estimador no sesgado de σ^2 .

3. El método de los menores cuadrados produce, también, un estimador no sesgado de la varianza de muestreo $\sigma_{b_j}^2$ de cada uno de los coeficientes estimados b_j ¹⁶⁷ y de la covarianza de muestreo de cada par de coeficientes estimados $\sigma_{b_j b_h}$ ¹⁶⁸.
4. En el conjunto de todos los estimadores de los β_j que son lineales y que no son sesgados, el estimador de los mínimos cuadrados b_j es el “mejor”, o sea posee la más alta eficacia relativa (con relación a las propiedades deseables de los estimadores, vea el capítulo 2-2). En otras palabras, es el estimador cuya distribución de muestreo posee la más pequeña varianza; es cuando decimos que el estimador de los mínimos cuadrados es BLUE (“Best Linear Unblased Estimate”).

3-2.2.3 El modelo clásico de la regresión lineal normal

Vimos que el modelo de muestreo que se define combinando el modelo lineal general con las hipótesis H1 hasta H4 es in-

¹⁶⁷ Los logicales de aplicación estadística procuran automáticamente estos valores estimados.

¹⁶⁸ La matriz estimada de las varianzas-covarianzas se define con $(X'X)^{-1} SSR / (n-k)$

completo. Así, antes de efectuar cualquier test de hipótesis es necesario complementar la especificación del modelo de muestreo; esto es, tenemos que especificar la forma de la distribución de los términos aleatorios u_i .

H5: Cada uno de los términos aleatorios posee una distribución normal.

Puesto que la distribución normal posee únicamente dos parámetros (el promedio y la varianza), al combinar H1, H2 y H5, se alcanza una especificación casi completa de la distribución de los términos aleatorios:

Los términos aleatorios poseen distribuciones normales idénticas con un promedio nulo; además, son independientes entre sí, es decir, el único parámetro desconocido es su varianza común σ^2 .

Basándose en las hipótesis H1 hasta H5, deducimos que se obtienen los valores de los términos aleatorios de una misma población.

Claro está, la selección de la distribución normal es muy cómoda. Para empezar, en estas condiciones, los estimadores poseen también una distribución normal,¹⁶⁹ puesto que una combinación lineal de variables normales es también una variable normal y que los estimadores de los menores cuadrados son lineales con relación a los y_i (y, por lo tanto, con relación a los términos aleatorios u_i). Luego, la distribución normal es de relativo fácil manejo por encerrar únicamente dos parámetros.

Pero, ¿en qué nos basamos al momento de pretender que la distribución de términos aleatorios es, efectivamente, normal? Esta pretensión es, a veces, solamente una implicación del modelo teórico o de la naturaleza del fenómeno estudia-

¹⁶⁹ Es esta propiedad, en particular, la que permite aplicar el test t de Student.

do. Sin embargo, la justificación de más rigor se refiere al teorema límite central. Basándose en algunas variantes de este teorema, si el término aleatorio de la regresión representa la influencia combinada de un gran número de variables faltantes en el modelo, entonces se puede considerar que la distribución normal es una aproximación razonable de la distribución de la influencia combinada de las variables faltantes.¹⁷⁰

La suma de la hipótesis H5 completa la especificación del modelo de muestreo, conocido también como “modelo clásico de regresión lineal normal” (es decir, el modelo clásico de regresión lineal, más la hipótesis de normalidad de los términos aleatorios). Cuando se respetan las hipótesis H1 hasta H5, la estadística facilita al investigador toda una gama de variables-tests que permiten efectuar diversos tests de hipótesis de diferentes tipos. En particular, ya vimos cómo se podía aplicar el test *t* de Student y el test *F* de Fisher.

No obstante, es responsabilidad del investigador decidir si las condiciones H1 hasta H5 que constituyen el modelo de muestreo son aceptables y si, por consiguiente, los tests que dependen del modelo son válidos.¹⁷¹ En efecto, al tratar con los tests clásicos, como el test de Student del cual dimos algunos ejemplos, el mismo modelo de muestreo no se discute. La decisión de aceptar o no las hipótesis H1 hasta H5 puede basarse en consideraciones a priori. Sin embargo, es posible validarlas luego con, para empezar, un examen visual de los residuos de la regresión y luego, con más rigor, con la aplicación de tests de diagnóstico. Estos tests de hipótesis son, por

¹⁷⁰ Gujarati (1992, p. 93); Theil (1971, p. 368-370), Freund (1962, pp. 185-188); Malinvaud (1969, pp. 268-271).

¹⁷¹ Es la razón de ser de la observación que podemos leer al pie de la p. 19 de Lemelin y Polèse (1995): “Stricly speaking, however, the classical hypotheses under which the tests are exact are not fully realized [...]”.

así decirlo, tests de “nivel superior” que se aplican, justamente, a ciertos aspectos del modelo de muestreo.¹⁷²

Resumen: especificación de un modelo aleatorio

Condiciones del modelo clásico de regresión lineal:

H1: Para cada observación, el valor del término aleatorio es sorteado de una población teórica de promedio nulo:

$$E(u_i) = 0 \text{ para todo } i.$$

H2a: Para todas las observaciones, las poblaciones teóricas de donde se sortean los valores de los términos aleatorios tienen la misma varianza:

$$\sigma_i^2 = \sigma^2 \text{ para todo } i.$$

H2b: Para cada observación, el valor del término aleatorio es estadísticamente independiente de los valores de los términos aleatorios de las demás observaciones:

$$\sigma_{ij} = 0 \text{ para todas las combinaciones } i,j \text{ cuando } i \neq j.$$

H3: Las variables independientes x_{ij} son no aleatorias (en particular sus mediciones se efectúan sin errores).

H4: Existen menos parámetros para estimar que observaciones y no hay redundancias entre las variables independientes.

Propiedades del estimador de los menores cuadrados en el modelo clásico de la regresión lineal: el teorema de Gauss-Markov

1. El estimador de los menores cuadrados de β_j es no sesgado:
el promedio de la distribución de muestreo de b_j es igual a β_j .

¹⁷² Por ejemplo, Heikkila *et al.* (1989) examinan con la ayuda de tests de diagnósticos el problema de la multicolinealidad espacial entre diversas variables de distancia (p. 228).

2. $\frac{1}{n-k} \sum_i (y_i - \hat{y}_i)^2 = \frac{SSR}{n-k}$ es un estimador no sesgado de σ^2 .
3. El método de los menores cuadrados produce, también, un estimador no sesgado de la varianza de muestreo $\sigma_{b_j}^2$ de cada uno de los coeficientes estimados b_j , y de la covarianza de muestreo de cada par de coeficientes estimados $\sigma_{b_j b_h}$.

El estimador de los menores cuadrados es el estimador que posee la más grande eficacia relativa (la más pequeña varianza muestral): es cuando decimos que el estimador de los menores cuadrados es BLUE (“Best Linear Unblased Estimate”).

*Condición suplementaria del modelo clásico
de regresión lineal normal*

H4: Existen menos parámetros para estimar qué observaciones y no hay redundancias entre las variables independientes.

3-2.3 ¿SE RESPETAN LAS HIPÓTESIS DEL MODELO
DE MUESTREO? ¿Y EN CASO CONTRARIO, QUÉ SUCEDE?*

El momento de efectuar los tests de hipótesis que requieren de variables-tests, como es el caso del t Student o del f de Fisher, nunca se cuestiona el modelo de muestreo en sí (vea 3-2.1). Sin embargo, es importante que el analista lo ponga en tela de juicio.

Existen procedimientos estadísticos formales para probar algunos aspectos del modelo de muestreo. No obstante, en muchos de los casos, es posible establecer un diagnóstico preliminar con un simple examen visual del gráfico de los residuos. Por lo general, cualquier motivo geométrico, cual-

* Referencias: Wonnacott y Wonnacott (1992, pp. 524-527)

quier apariencia de organización debería llamarnos la atención.

Los principales problemas que puede revelar un examen de los residuos son:

4. Una inadecuada especificación del modelo teórico.
5. La autocorrelación de los términos aleatorios.
6. La heteroescadasticidad.
7. Observaciones excéntricas.

Examinemos concretamente en qué consiste cada uno de estos cuatro problemas. Luego comentaremos algo sobre la multicolinealidad.

3-2.3.1 Error de especificación del modelo teórico

Especificar un modelo consiste en construir una lista de las variables independientes y definir la forma de la relación entre éstas y la variable dependiente. El error de especificación más frecuente se comete al omitir una de las variables independientes que debería aparecer en el modelo. De la misma manera, se comete un error de especificación en cuanto a la forma, al instituir una relación lineal entre las variables cuando una relación lineal entre sus logaritmos se imponía. Es evidente que al cometer estos tipos de errores de especificación, se compromete, en su totalidad, a las hipótesis del modelo clásico de regresión lineal, empezando por la misma relación lineal.

Existe una multitud de posibilidades de cometer errores de especificación de un modelo. Nos contentaremos con dar una ilustración de lo dicho. Supongamos que el “verdadero” modelo se defina con

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + u_i$$

La variable dependiente y_i es una función de x_i y del cuadrado de x_i . Contando la constante, tenemos tres variables independientes. Aunque no sea lineal, es posible “linealizarlo”

fácilmente con solo considerar x_i y x_i^2 como dos variables diferentes.

Supongamos ahora que estimamos el modelo incompleto:

$$y_i = \alpha_0 + \alpha_1 x_i + u_i$$

El término faltante $\alpha_2 x_i^2$ se manifestará entonces en los residuos. Al momento de examinar el gráfico de la relación entre los residuos y la variable independiente x_i , será posible detectar entre ellos una relación sistemática que tomará la forma de una curva. Se ilustra esta situación con las figuras 6 a 8.

Ilustración geométrica de un error de especificación

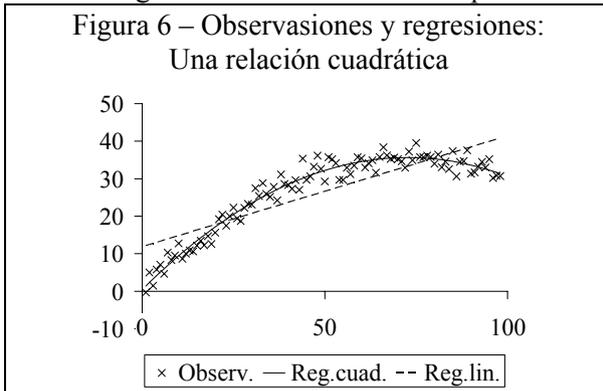


Figura 7 – Residuos de la regresión cuadrática
Residuos sin error de especificación

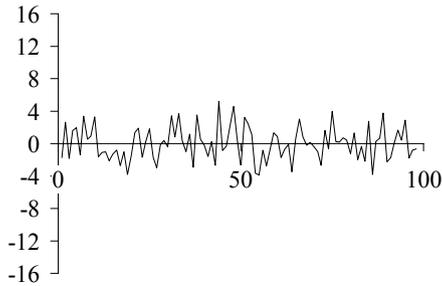
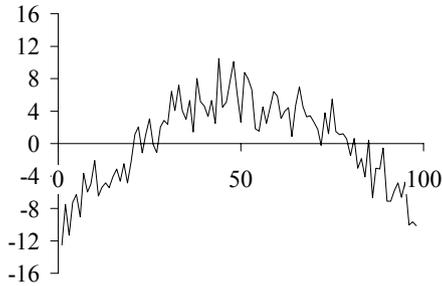


Figura 8 – Residuos de la regresión lineal
Residuos con error de especificación



3-2.3.2 Autocorrelación de los términos aleatorios

Surge una contradicción entre la autocorrelación de los términos aleatorios y la hipótesis H2b del modelo clásico de la regresión lineal, la cual nos indica que los términos aleatorios de las diferentes observaciones son independientes entre sí ($\sigma_{ij} = 0$ para todas las combinaciones i, j donde $i \neq j$).

La auto-correlación es frecuente cuando las observaciones se efectúan en momentos sucesivos en el tiempo. Datos de

esta naturaleza se llaman series temporales o series cronológicas. No obstante, con las series temporales, sucede a menudo que, a causa de cierta inercia, las desviaciones aleatorias necesiten algún tiempo para desaparecer¹⁷³. Así, si u_t , valor del término aleatorio en el periodo t , es positivo entonces el promedio de u_{t+1} teniendo $u_t > 0$ (la esperanza matemática condicional) no será nula, pero positiva. En estas condiciones, tenemos

$$\sigma_{t,t-1} \neq 0$$

lo que contradice la hipótesis H2b del modelo clásico de la regresión lineal.

A menudo, es posible detectar cierta auto-correlación en las series temporales con examinar la gráfica de los residuos en función del tiempo. Podremos, entonces, observar que los errores sucesivos parecen encadenarse los unos con los otros al lugar de efectuar saltos desordenados. Las figuras 9 y 10 ilustran este fenómeno.

Se generaron los datos subyacentes a los residuos de regresiones que se presentan en las figuras 8 y 10 con la ecuación

$$y_t = x_t + 10 \eta_t$$

donde se generó η_t gracias a un proceso autorregresivo del tipo

$$\eta_t = (1 - \alpha) \varepsilon_t + \alpha \eta_{t-1}$$

¹⁷³ Consideremos, por ejemplo, el fenómeno de la evolución de los precios: existe en el funcionamiento de la economía cierta “rigidez institucional” (tal es el caso con los contratos a largo plazo como las convenciones colectivas) la cual provoca que los precios no reaccionen inmediatamente a los cambios en los factores fundamentales. En la relación entre los precios y los factores fundamentales, esto se traduce con la aparición de auto-correlación.

teniendo ε_t una distribución normal. Se repiten las figuras tres veces, cada una con valores diferentes del parámetro α : 0.9, 0.6 y 0.

Es posible encontrar también autocorrelación al manejar datos espaciales. Su detección se complica aún más; en efecto, cuando el tiempo no posee más que una dimensión, el espacio posee dos dimensiones, lo que impide trazar gráficas como se efectuó arriba.

Consecuencias

Los estimadores de los mínimos cuadrados siguen insesgados, sin embargo su varianza es fuerte (los estimadores son menos precisos). Además, las fórmulas que se exhibieron en el caso anterior para estimar la varianza de los estimadores subestiman la verdadera varianza (o sea que nos deja creer que hay más precisión); de aquí que los tests estadísticos ya no sean válidos.

Test de detección

Existe un test de detección de la autocorrelación temporal conocido como el test de Durbin-Watson (vea también Kennedy, 1992, p. 128).

Remedios

En cuanto exista autocorrelación, es necesario aplicar el remedio que consiste en completar el modelo emitiendo hipótesis sobre el mecanismo de autocorrelación, lo que permite usar un método llamado método de los mínimos cuadrados generalizados.

Ilustración geométrica de la autocorrelación
(autocorrelación fuerte)

Figura 9a – Observaciones y regresión:
Autocorrelación de los términos aleatorios

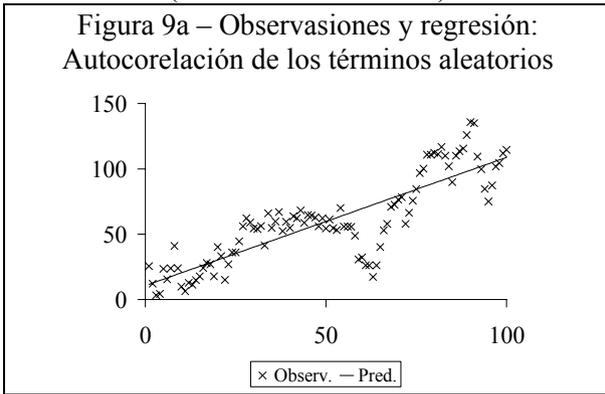
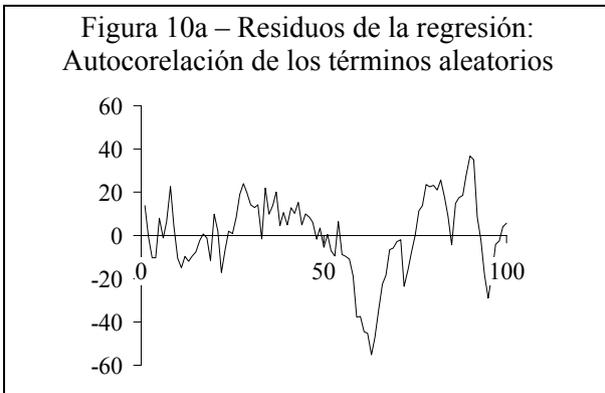


Figura 10a – Residuos de la regresión:
Autocorrelación de los términos aleatorios



Términos aleatorios η_t generados con la fórmula

$$\eta_t = (1 - \alpha) \varepsilon_t + \alpha \eta_{t-1}, \text{ con } \alpha = 0.9$$

Ilustración geométrica de la autocorrelación
(autocorrelación mediana)

Figura 9b – Observaciones y regresión:
Autocorrelación de los términos aleatorios

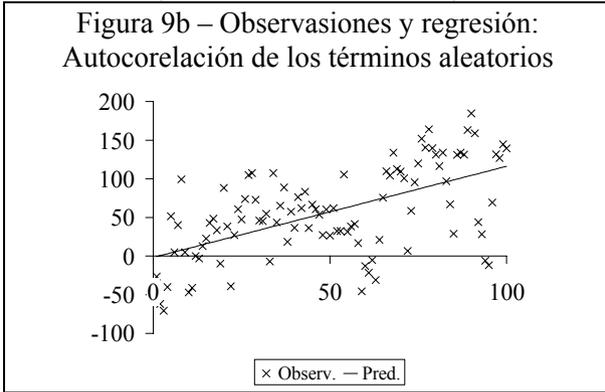
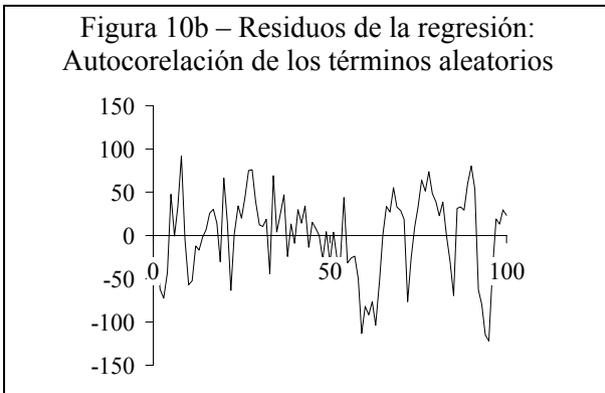


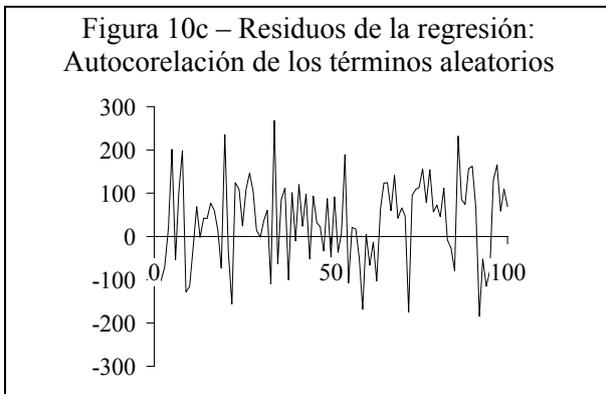
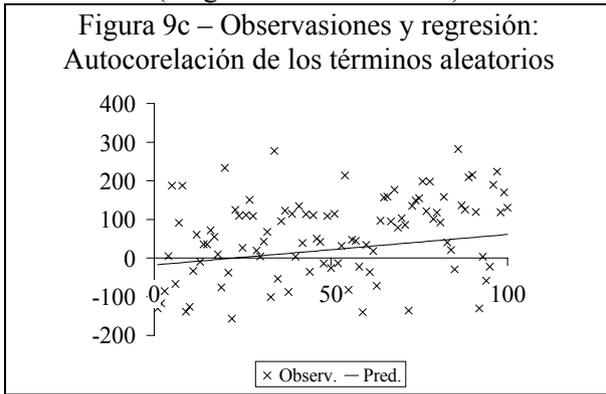
Figura 10b – Residuos de la regresión:
Autocorrelación de los términos aleatorios



Términos aleatorios η_t generados con la fórmula

$$\eta_t = (1 - \alpha) \varepsilon_t + \alpha \eta_{t-1}, \text{ con } \alpha = 0.6$$

Ilustración geométrica de la autocorrelación
(ninguna autocorrelación)



Términos aleatorios η_t generados con la fórmula

$$\eta_t = (1 - \alpha) \varepsilon_t + \alpha \eta_{t-1}, \text{ con } \alpha = 0$$

3-2.3.3 Heteroscedasticidad

La heteroscedasticidad es el contrario de la homoscedasticidad. La homoscedasticidad es la hipótesis H2a del modelo clásico de la regresión lineal, la cual nos indica que la varian-

za del error es la misma para todas las observaciones ($\sigma_i^2 = \sigma^2$ para todos los i)

Se encuentra la heteroscedasticidad a menudo, y particularmente en los estudios en corte transversal donde es posible que los términos aleatorios sean proporcionales al “tamaño” del sujeto observado.

Por ejemplo, en un estudio sobre los gastos de vivienda de los hogares, es posible que la variabilidad de éstos aumente al mismo tiempo que el ingreso, es decir, los hogares con bajos ingresos se ven obligados a limitar sus gastos de vivienda cuando, por lo contrario, entre los hogares con altos ingresos, algunos optan por usar gran parte de sus ingresos en una vivienda lujosa, y otros prefieren encontrar una vivienda confortable aunque sin grandes lujos, con el fin de gastar su dinero de otra manera. En este particular caso, las variaciones aleatorias debidas a los diferentes gustos¹⁷⁴ son mayores para los hogares más holgados.

En un gráfico de residuos, la heteroscedasticidad podría aparecer como un motivo en forma de trompeta o cono, siempre y cuando ordenemos las observaciones en orden creciente de la variable dependiente o de una de las variables independientes. En lugar de tener en abscisa solamente los números de orden de las observaciones, es posible construir también una gráfica donde se representen los residuos en función de los valores correspondientes de la variable dependiente o de una de las variables independientes en abscisa.

Se generaron los datos subyacentes a los residuos de regresiones que se presentan en las figuras 11 y 12 con la ecuación

$$y_i = x_i + 100 \eta_i$$

¹⁷⁴ Este es el ejemplo de un modelo al cual le faltan algunas variables inobservables (los gustos) cuyo efecto se representa con el término aleatorio.

donde se generó η_i con la ecuación

$$\eta_i = 0,1 \left(\varepsilon_i \sqrt{x_i} \right)$$

con ε_i la cual tiene una distribución normal.

Es importante observar que $x_i = i$ de tal manera que las observaciones son automáticamente ordenadas en orden creciente de la variable independiente x_i .

Consecuencias

La precisión del estimador no es tan buena y los tests de hipótesis no son válidos.

Test de detección

Test de Goldfield y Quandt (Theil, 1971, pp. 196-199; vea también Kennedy, 1992, p. 126).

Remedios

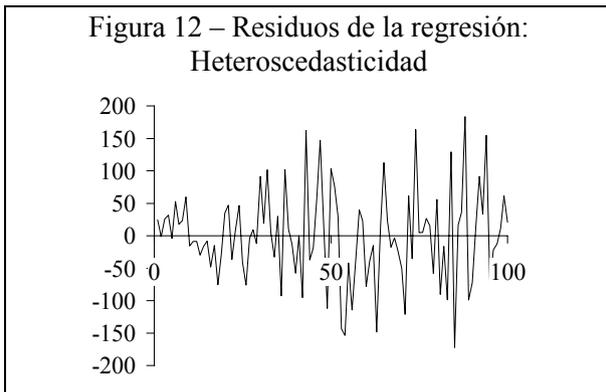
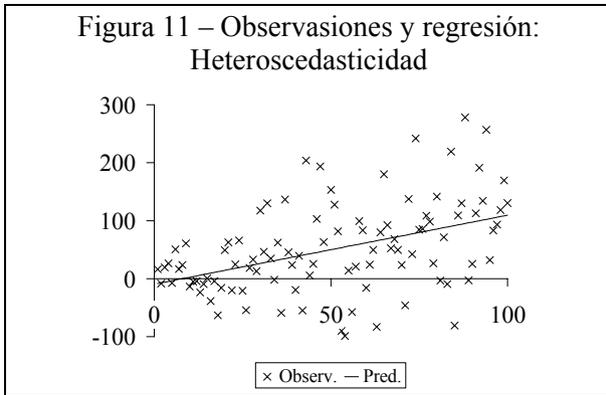
Se puede corregir el modelo de muestreo con la transformación de los datos. Por ejemplo, si nos damos cuenta que la varianza depende de una de las variables independientes, digamos x_{ik} , y que es posible representar de manera aproximativa esta relación con

$$\sigma_i^2 = x_{ik} \sigma^2$$

entonces es posible recrear la homoscedasticidad y las condiciones de Gauss-Markov con sólo aplicar a las variables la transformación que sigue:

$$y'_i = \frac{y_i}{\sqrt{x_{ik}}} \text{ y } x'_{ij} = \frac{x_{ij}}{\sqrt{x_{ik}}}$$

Ilustración geométrica de la heteroscedasticidad



Términos aleatorios η_i generados con la fórmula

$$\eta_i = 0,1 (\varepsilon_i \sqrt{x_i})$$

3-2.3.4 Observaciones excéntricas

Las observaciones excéntricas (outliers) provienen a veces de situaciones donde factores que no se tomaron en cuenta en el modelo intervinieron. Aunque las observaciones excéntricas no interfieren con las hipótesis del modelo clásico de regresión lineal, es posible que falseen los resultados de la regresión

sión. Un examen de los residuos permite detectar las observaciones excéntricas. Luego, es posible buscar si factores particulares son la causa de tales observaciones y decidir dejarlos a un lado. Sin embargo, es importante evitar eliminar observaciones ad hoc, por simple comodidad... Las figuras que mostramos a continuación nos dan un ejemplo visual de residuos con la presencia de observaciones excéntricas.

Ilustración geométrica de la presencia de observaciones excéntricas

Figura 13 – Observaciones y regresión:
Observaciones excéntricas (*outliers*)

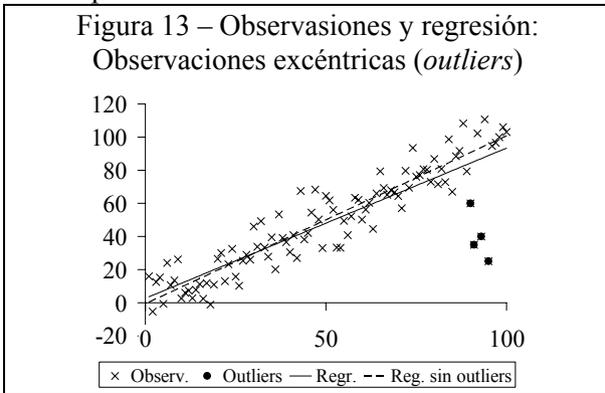


Figura 14 – Residuos de la regresión con observaciones excéntricas (*outliers*)

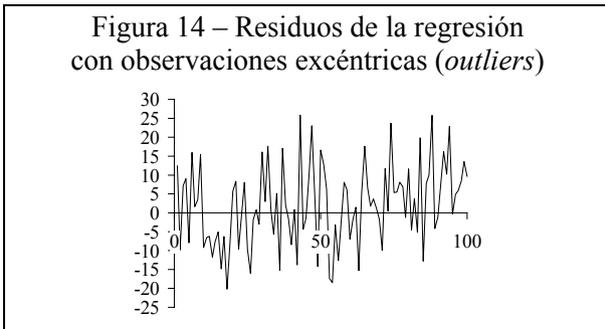
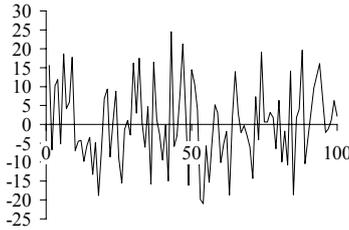


Figura 15 – Residuos de la regresión sin observaciones excéntricas (*outliers*)



3-2.3.5 Multicolinealidad¹⁷⁵

Definición

Se hace la distinción entre la multicolinealidad estricta y la multicolinealidad aproximativa.

Multicolinealidad estricta. Surge una contradicción entre la multicolinealidad estricta y la hipótesis H4 del modelo clásico de la regresión lineal, la cual nos indica que no debe existir redundancia entre las variables independientes¹⁷⁶. La multicolinealidad estricta es muy poco común con datos reales cuando puede ser el resultado de un error de especificación si el modelo contiene variables mudas (dummy variables). Trataremos nuevamente este problema cuando estudiaremos el análisis de varianza por medio de la regresión múltiple (capítulo 4-2).

La multicolinealidad estricta no representa ningún problema de detección, puesto que se diagnostica su presencia

¹⁷⁵ Wonnacott y Wonnacott (1992, pp. 568-572).

¹⁷⁶ Técnicamente, cuando existe multicolinealidad estricta, el rango de la matriz X es inferior a k , lo que implica que la matriz inversa $(X'X)^{-1}$ no exista y el estimador tampoco.

con el software de aplicación estadística (por la imposibilidad de los cálculos de estimación).

Multicolinealidad aproximativa. La multicolinealidad aproximativa es mucho más frecuente. Acontece cuando una de las variables independientes se correlaciona fuertemente con otra o una combinación lineal de las demás. Esta variable, aunque no estrictamente redundante, puede considerarse como “casi” redundante, es decir, su aportación de información no es relevante comparada a la información que aportan las otras variables.

Consecuencias

La precisión de los estimadores es baja, o sea que sus varianzas de muestreo $s_{b_j}^2$ son elevadas. No es posible diferenciar correctamente la influencia de las variables que están correlacionadas entre sí.

De manera concreta, en el caso de dos variables, esto puede manifestarse como sigue: ninguna de las dos variables posee un coeficiente significativamente diferente de cero, sin embargo, al quitar las dos variables, el modelo resultante no aprueba el test F.

Test de detección

En el caso de la correlación entre dos variables, es posible examinar los coeficientes de correlación simple de las variables independientes entre ellas. Pero la multicolinealidad es, a menudo, mucho más compleja y es necesario recurrir a tipos de análisis más sofisticados (vea Kennedy, 1992, p. 180, con relación a *condición index*).

Remedios

En algunos casos existe quizás una o varias variables independientes que sobran en el modelo. No obstante, casi siempre debemos aceptar “vivir con” y renunciar a separar la influencia de las variables correlacionadas. En algunos casos, descartar una variable por motivo de multicolinealidad podría hasta revelarse un grave error.

La figura que presentamos a continuación ilustra tal situación. Los factores inobservables A , B y C influyen la variable dependiente Y . Los factores inobservables no pueden figurar en el modelo. Para reemplazarlos, el modelo contiene dos variables independientes observables X_1 y X_2 : la primera es influenciada por A y B , y la segunda, por B y C . A causa de la influencia compartida de B , X_1 y X_2 son correlacionadas. Sin embargo, al descartar una de las dos variables, descartaríamos al mismo tiempo el factor subyacente A o C .

¿Descartar una variable, o no?

