3.2 Mecánica e interpretación de los mínimos cuadrados ordinarios

A continuación se resumen algunas características algebraicas y de cálculo del método de mínimos cuadrados ordinarios aplicado a un conjunto de datos determinado. También se analiza cómo interpretar la ecuación estimada.

Obtención de las estimaciones de MCO

Primero se considera la estimación del modelo con dos variables independientes. La ecuación estimada de MCO se escribe de manera similar a la de regresión simple:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2, \tag{3.9}$$

donde

$\hat{\beta}_0$ = la estimación de β_0.

$\hat{\beta}_1$ = la estimación de β_1.

$\hat{\beta}_2$ = la estimación de β_2.

Pero, ¿cómo se obtienen $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$? El método de **mínimos cuadrados ordinarios** elige las estimaciones que minimizan la suma de los residuales cuadrados. Es decir, dadas n observaciones sobre y, x_1 y x_2, $\{(x_{i1}, x_{i2}, y_i): i = 1, 2, \ldots, n\}$, las estimaciones $\hat{\beta}_0$, $\hat{\beta}_1$ y $\hat{\beta}_2$ se eligen de manera simultánea para que

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2 \tag{3.10}$$

sea tan pequeña como sea posible.

Para entender lo que hacen los MCO, es importante dominar el significado de los índices de las variables independientes en la ecuación (3.10). Aquí las variables independientes tienen dos subíndices, i seguido ya sea de un 1 o de un 2. El subíndice i se refiere al número de la observación. De manera que la suma en la ecuación (3.10) corresponde a todas las observaciones, desde $i = 1$ hasta n. El segundo índice es un método que sirve para distinguir entre las diferentes variables independientes. En el ejemplo en que se relacionan *wage* (salario) con *educ* (educación) y *exper* (experiencia), $x_{i1} = educ_i$ es la educación de la persona i de la muestra, y $x_{i2} = exper_i$ es la experiencia de la persona i. En la ecuación (3.10) la suma de los residuales cuadrados es $\sum_{i=1}^{n} (wage_i - \hat{\beta}_0 - \hat{\beta}_1 educ_i - \hat{\beta}_2 exper_i)^2$. En lo que se presenta a continuación, el subíndice i se reserva como índice del número de la observación. Si se escribe x_{ij}, esto significa la observación i-ésima sobre la variable independiente j-ésima. (Algunos autores prefieren intercambiar el orden del número de observación y el número de variable, de manera que x_{1i} es la observación i de la variable uno. Pero esto es sólo cuestión de gustos respecto a la notación.)

En el caso general, con k variables independientes, se buscan las estimaciones $\hat{\beta}_0$, $\hat{\beta}_1$, …, $\hat{\beta}_k$ en la ecuación

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k. \tag{3.11}$$

Estas $k + 1$ estimaciones de MCO se eligen de manera que minimicen la suma de los residuales cuadrados:

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik})^2. \tag{3.12}$$

Este problema de minimización se resuelve empleando cálculo multivariable (vea el apéndice 3A). Esto lleva a $k + 1$ ecuaciones lineales en $k + 1$ incógnitas $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$:

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\
\sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\
\sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\
&\ \ \vdots \\
\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0.
\end{aligned} \tag{3.13}$$

A estas ecuaciones se les suele llamar las **condiciones de primer orden** de MCO. Como en el caso del modelo de regresión simple de la sección 2.2, las condiciones de primer orden de MCO se obtienen con el método de los momentos: bajo el supuesto (3.8), $\mathrm{E}(u) = 0$ y $\mathrm{E}(x_j u) = 0$, donde $j = 1, 2, \dots, k$. Las ecuaciones en (3.13) son las contrapartes muestrales de estos momentos poblacionales, aunque se ha omitido la división entre el tamaño n de la muestra.

Resolver a mano las ecuaciones en (3.13) es tedioso aun cuando n y k sean de tamaño moderado. Sin embargo, las computadoras modernas que cuenten con software para estadística y econometría resuelven estas ecuaciones muy rápido, incluso con n y k grandes.

Sólo una pequeña advertencia: se debe suponer que las ecuaciones en (3.13) tienen soluciones *únicas* para las $\hat{\beta}_j$. Por ahora, sólo se supondrá esto, como suele ser el caso en los modelos bien definidos. En la sección 3.3 se establece el supuesto necesario para que existan las estimaciones únicas de MCO (vea el supuesto RLM.3).

Como en el análisis de regresión simple, a la ecuación (3.11) se le llama **línea de regresión de MCO** o **función de regresión muestral (FRM)**. A $\hat{\beta}_0$ se le llama **estimación del intercepto de MCO** y a $\hat{\beta}_1, \dots, \hat{\beta}_k$, **estimaciones de las pendientes de MCO** (correspondientes a las variables independientes $x_1, x_2, \dots, x_k$).

Para indicar que se ha realizado una regresión por MCO, se escribe la ecuación (3.11) reemplazando la y y las $x_1, \dots, x_k$ por los nombres de las variables (por ejemplo, *wage*, *educ* y *exper*), o se dice que "se realizó una regresión por MCO de y sobre $x_1, x_2, \dots, x_k$" o que "se ha regresado y sobre $x_1, x_2, \dots, x_k$". Estas son maneras abreviadas de decir que se ha usado el método de mínimos cuadrados ordinarios para obtener la ecuación (3.11) de MCO. A menos que de manera explícita se diga otra cosa, aquí siempre se estimará el intercepto junto con las pendientes.

Interpretación de la ecuación de regresión de MCO

Más importante que los detalles relacionados con el cálculo de los $\hat{\beta}_j$ es la *interpretación* de la ecuación estimada. Para empezar se verá el caso en que se tienen dos variables independientes:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \tag{3.14}$$

En la ecuación (3.14) el intercepto $\hat{\beta}_0$ es el valor predicho para y cuando $x_1 = 0$ y $x_2 = 0$. Algunas veces es interesante hacer $x_1 = 0$ y $x_2 = 0$, otras no tiene sentido. Sin embargo, el intercepto siempre es necesario para obtener una predicción de y mediante la línea de regresión de MCO, como se indica en la ecuación (3.14).

Las estimaciones $\hat{\beta}_1$ y $\hat{\beta}_2$ se interpretan como **efectos parciales** o **efectos *ceteris paribus***. De acuerdo con la ecuación (3.14), se tiene

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1 + \hat{\beta}_2\Delta x_2,$$

de manera que dados los cambios en x_1 y x_2. se puede obtener el cambio predicho para y. (Observe cómo el intercepto no tiene nada que ver con el cambio en y.) En particular, cuando x_2 se mantiene constante, de manera que $\Delta x_2 = 0$, entonces

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1,$$

con x_2 constante. El punto clave es que, al incluir x_2 en el modelo, se obtiene un coeficiente para x_1 con una interpretación *ceteris paribus*. A esto se debe que el análisis de regresión múltiple sea tan útil. De manera similar,

$$\Delta\hat{y} = \hat{\beta}_2\Delta x_2,$$

x_1 constante.

Ejemplo 3.1

[Determinantes del promedio en la universidad]

Las variables en la base de datos GPA1.RAW incluyen el promedio general de calificaciones en la universidad (*colGPA*), el promedio general de calificaciones en el bachillerato (*hsGPA*) y la puntuación en el examen de admisión a la universidad (*ACT*) para una muestra de 141 estudiantes de una universidad grande; los promedios generales de calificaciones tanto del bachillerato como de la universidad se dan en una escala de cuatro puntos. Para predecir el promedio general de calificaciones en la universidad, a partir del promedio general de calificaciones en el bachillerato y de la calificación en el examen de admisión se obtiene la siguiente línea de regresión de MCO:

$$\widehat{colGPA} = 1.29 + .453\ hsGPA + .0094\ ACT. \tag{3.15}$$

¿Cómo se interpreta esta ecuación? Primero, el intercepto 1.29 es la predicción del promedio general de calificaciones en la universidad si *hsGPA y ACT* son ambos cero. Dado que ninguna persona que asista a la universidad tiene cero como promedio general de calificaciones de bachillerato ni cero en el examen de admisión a la universidad, el intercepto, en este caso, no tiene en sí ningún significado.

Estimaciones más interesantes son las de los coeficientes de pendiente de *hsGPA* y *ACT*. Como era de esperarse, existe una relación parcial positiva entre *colGPA* y *hsGPA*: con *ACT* constante, cada punto más en *hsGPA* se relaciona con .453 adicional en el promedio general de la universidad, es decir, casi medio punto. En otras palabras, si se eligen dos estudiantes, A y B, y éstos tienen la misma puntuación en el examen de admisión (ACT), pero el promedio general en el bachillerato del estudiante A es un punto superior al del estudiante B, entonces se predice que en la universidad el estudiante A tendrá un promedio general de calificaciones .453 más alto que el estudiante B. (Esto no dice nada acerca de dos personas reales, es sólo la mejor predicción.)

El signo de *ACT* implica que, si *hsGPA* permanece constante, un cambio de 10 puntos en el examen de admisión (ACT) —un cambio muy grande, ya que en la muestra la puntuación promedio es de 24 con una desviación estándar menor a tres— tendrá un efecto sobre *colGPA* de menos de una décima de punto. Este es un efecto pequeño que indica que, una vez que se ha tomado en cuenta el promedio general del bachillerato, la puntuación en el examen de admisión (ACT) no es un fuerte predictor del promedio general en la universidad. (Naturalmente, hay muchos otros factores que contribuyen al promedio general de calificaciones en la universidad, pero aquí nos concentramos en los estadísticos disponibles para los estudiantes de bachillerato.) Más adelante, después de que se analice la inferencia estadística, se mostrará que el coeficiente de ACT no sólo es pequeño para fines prácticos, sino que es estadísticamente insignificante.

Centrándose en el análisis de regresión simple que relaciona *colGPA* sólo con *ACT* se obtiene

$$\widehat{colGPA} = 2.40 + .0271\ ACT;$$

de manera que el coeficiente de *ACT* es casi el triple del estimado en (3.15). Pero esta ecuación *no* permite comparar dos personas con el mismo promedio general en el bachillerato; esta ecuación corresponde a otro experimento. Después se comentará más acerca de las diferencias entre la regresión múltiple y la simple.

El caso con más de dos variables independientes es similar. La línea de regresión de MCO es

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k. \tag{3.16}$$

Expresada en términos de cambios,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \ldots + \hat{\beta}_k \Delta x_k. \tag{3.17}$$

El coeficiente de x_1 mide el cambio en $\hat{y}$ por un aumento de x_1 de una unidad, manteniendo constantes todas las demás variables independientes. Es decir,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1, \tag{3.18}$$

manteniendo constantes $x_2, x_3, \ldots, x_k$. Por tanto, las variables $x_2, x_3, \ldots, x_k$ han sido *controladas* al estimar el efecto de x_1 sobre y. Los demás coeficientes tienen una interpretación similar.

A continuación se presenta un ejemplo con tres variables independientes.

Ejemplo 3.2

[Ecuación para el salario por hora]

Empleando las 526 observaciones sobre trabajadores en la base de datos WAGE1.RAW, las variables *educ* (años de educación), *exper* (años de experiencia en el mercado laboral) y *tenure* (años de antigüedad en el empleo actual) se incluyen en una ecuación para explicar log(*wage*). La ecuación estimada es

$$\widehat{\log(wage)} = .284 + .092\ educ + .0041\ exper + .022\ tenure. \tag{3.19}$$

Como en el caso de la regresión simple, los coeficientes tienen una interpretación porcentual. Aquí la única diferencia es que también tienen una interpretación *ceteris paribus*. El coeficiente .092 significa que manteniendo *exper* y *tenure* constantes, se predice que un año más de educación incrementa log(*wage*) en .092, lo que se traduce en un aumento aproximado de 9.2% [100(.092)] en *wage*. Es decir, si se toman dos personas con los mismos niveles de experiencia y antigüedad laboral, el coeficiente de *educ* es la diferencia proporcional con el salario predicho cuando en sus niveles de educación hay una diferencia de un año.

Esta medida del rendimiento de la educación al menos mantiene constantes dos factores importantes de la productividad; para saber si ésta es un buen estimado del rendimiento *ceteris paribus* de un año más de educación es necesario estudiar las propiedades estadísticas de los estimadores de MCO (vea la sección 3.3).

El significado de "mantener todos los demás factores constantes" en la regresión múltiple

La interpretación del efecto parcial de los coeficientes de pendiente en el análisis de regresión múltiple puede causar cierta confusión, por lo que a continuación se presenta un análisis más amplio.

En el ejemplo 3.1, se observó que el coeficiente de *ACT* mide la diferencia que se predice para *colGPA* cuando *hsGPA* se mantiene constante. El poder del análisis de regresión múltiple es que proporciona esta interpretación *ceteris paribus* incluso cuando los datos *no* hayan sido recolectados de manera *ceteris paribus*. Al darle al coeficiente de *ACT* una interpretación de efecto parcial, puede parecer que se salió y se muestrearon personas con el mismo promedio general en el bachillerato pero con puntuaciones diferentes en el examen de admisión (ACT). Este no es el caso. Los datos son una muestra aleatoria tomada de una universidad grande: para obtener los datos no se pusieron restricciones sobre los valores muestrales de *hsGPA* o de *ACT*. Es muy raro que al obtener una muestra pueda uno darse el lujo de mantener constantes ciertas variables. *Si* se pudiera obtener una muestra de individuos con un mismo promedio general en el bachillerato, entonces se podría realizar un análisis de regresión simple relacionando *colGPA* con *ACT*. La regresión múltiple permite imitar esta situación sin restringir los valores de ninguna de las variables independientes.

El poder del análisis de regresión múltiple es que permite hacer en un ambiente no experimental, lo que en las ciencias naturales puede hacerse con experimentos controlados de laboratorio: mantener constantes otros factores.

Cambiar de manera simultánea más de una variable independiente

Algunas veces se desea cambiar a la vez más de una variable independiente para determinar el efecto resultante sobre la variable dependiente. Esto es fácil de hacer usando la ecuación (3.17). Por ejemplo, en la ecuación (3.19) se puede obtener el efecto estimado sobre *wage* cuando una persona permanece un año más en una misma empresa: tanto *exper* (experiencia general en la fuerza laboral) como *tenure* (antigüedad en el empleo actual) aumentan en un año. El efecto total (manteniendo *educ* constante) es

$$\Delta\widehat{\log(wage)} = .0041\ \Delta exper + .022\ \Delta tenure = .0041 + .022 = .0261,$$

es decir, aproximadamente 2.6%. Dado que tanto *exper* como *tenure* aumentan un año, simplemente se suman los coeficientes de *exper* y *tenure* y se multiplica por 100 para convertir el efecto en un porcentaje.

Valores ajustados y residuales de MCO

Después de obtener la línea de regresión de MCO (3.11), para cada observación puede obtenerse el *valor ajustado* o *valor predicho*. Para la observación *i*, el valor ajustado es

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik}, \tag{3.20}$$

que es el valor predicho que se obtiene sustituyendo en la ecuación (3.11) los valores de las variables independientes de la observación i. Al obtener los valores ajustados, no debe olvidarse el intercepto; de no ser así, el resultado puede ser erróneo. Por ejemplo, si en la ecuación (3.15) $hsGPA_i = 3.5$ y $ACT_i = 24$, $\widehat{colGPA}_i = 1.29 + .453(3.5) + .0094(24) = 3.101$ (redondeado a tres cifras decimales).

En general, el valor real de y_i para cualquier observación i no será igual al valor predicho, $\hat{y}_i$: el método de MCO minimiza el *promedio* del error de predicción cuadrado, lo cual no dice nada acerca del error de predicción de cualquier observación particular. El **residual** de la observación i está definido como en el caso de la regresión simple,

$$\hat{u}_i = y_i - \hat{y}_i. \tag{3.21}$$

Hay un residual para cada observación. Si $\hat{u}_i > 0$, entonces $\hat{y}_i$ es menor que y_i, lo que significa que, para esta observación, el valor predicho para y_i es menor al valor de y_i. Si $\hat{u}_i < 0$, entonces $y_i < \hat{y}_i$ y el valor predicho para y_i es mayor al valor de y_i.

Los valores ajustados y los residuales de MCO tienen algunas propiedades que son extensiones inmediatas del caso de una variable:

1. El promedio muestral de los residuales es cero y de esta manera $\bar{y} = \bar{\hat{y}}$.
2. La covarianza muestral entre cada una de las variables independientes y los residuales de MCO es cero. Por consiguiente, la covarianza muestral entre los valores ajustados de MCO y los residuales de MCO es cero.
3. El punto $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k, \bar{y})$ se encuentra siempre sobre la línea de regresión de MCO: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \ldots + \hat{\beta}_k\bar{x}_k$.

Pregunta 3.2

En el ejemplo 3.1, la línea ajustada de MCO que explica *colGPA* en términos de *hsGPA* y de la puntuación en *ACT* es

$$\widehat{colGPA} = 1.29 + .453\ hsGPA + .0094\ ACT.$$

Si el promedio de *hsGPA* es 3.4 y el promedio *ACT* es aproximadamente 24.2, ¿cuál es el promedio muestral de *colGPA*?

Las dos primeras propiedades son consecuencia inmediata del conjunto de ecuaciones empleadas para obtener las estimaciones de MCO. La primera ecuación en (3.13) dice que la suma de los residuales es cero. Las ecuaciones restantes son de la forma $\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0$, lo que significa que la covarianza entre cada variable independiente y $\hat{u}_i$ es cero. La propiedad (3) es consecuencia inmediata de la propiedad (1).

Una interpretación de descuento de efectos parciales de la regresión múltiple

Cuando se aplica el método de MCO, no es necesario conocer las fórmulas explícitas para obtener los $\hat{\beta}_j$ que resuelven el sistema de ecuaciones dado en (3.13). Sin embargo, para ciertas deducciones, sí se necesitan las fórmulas explícitas para los $\hat{\beta}_j$. Estas fórmulas aclaran, además, el funcionamiento de los MCO.

Considérese nuevamente el caso con $k = 2$ variables independientes, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1x_1 + \hat{\beta}_2x_2$. Para mayor concreción, la atención se concentrará en $\hat{\beta}_1$. Una manera de expresar $\hat{\beta}_1$ es

$$\hat{\beta}_1 = \left(\sum_{i=1}^{n} \hat{r}_{i1}y_i\right) \Bigg/ \left(\sum_{i=1}^{n} \hat{r}_{i1}^2\right), \tag{3.22}$$

donde los $\hat{r}_{i1}$ son los residuales de MCO de una regresión simple de x_1 sobre x_2, usando la muestra presente. Se regresa la primera variable independiente, x_1, sobre la segunda variable independiente, x_2, y después se obtienen los residuales (aquí y no interviene). La ecuación (3.22) muestra que después se puede hacer una regresión simple de y sobre $\hat{r}_1$ para obtener $\hat{\beta}_1$. (Observe que los residuales $\hat{r}_{i1}$ tienen media muestral cero y de esta manera $\hat{\beta}_1$ es la estimación habitual de la pendiente en la regresión simple.)

La representación de la ecuación (3.22) da otra demostración de la interpretación del efecto parcial de $\hat{\beta}_1$. Los residuales $\hat{r}_{i1}$ son la parte de x_{i1} que no está correlacionada con x_{i2}. Otra manera de decir esto es que $\hat{r}_{i1}$ es x_{i1} después de que los efectos parciales de x_{i2} han sido descontados o deducidos. De esta manera, $\hat{\beta}_1$ mide la relación muestral entre y y x_1 después descontar los efectos parciales de x_2.

En el análisis de regresión simple no se descuentan los efectos parciales de otras variables porque en la regresión no se incluyen otras variables. El ejercicio de computadora C3.5 lo llevará a través del proceso de deducción de los efectos parciales empleando los datos de salario del ejemplo 3.2. Para fines prácticos, lo importante es que en la ecuación $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, $\hat{\beta}_1$ mide el cambio en y por cada incremento de una unidad en x_1, manteniendo x_2 constante.

En el modelo general con k variables explicativas, $\hat{\beta}_1$ también puede escribirse como en la ecuación (3.22), pero los residuales $\hat{r}_{i1}$ provienen de la regresión de x_1 sobre $x_2, \ldots, x_k$. Por tanto, $\hat{\beta}_1$ mide el efecto de x_1 sobre y después de que los efectos parciales de $x_2, \ldots, x_k$ han sido descontados o deducidos.

Comparación entre las estimaciones de la regresión simple y de la regresión múltiple

Existen dos casos especiales en los que la regresión simple de y sobre x_1 produce la *misma* estimación de MCO para el coeficiente de x_1 que la regresión de y sobre x_1 y x_2. Para ser más precisos, escriba la regresión simple de y sobre x_1 como $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$, y la regresión múltiple como $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$. Se sabe que, en general, el coeficiente $\tilde{\beta}_1$ de la regresión simple no es igual al coeficiente $\hat{\beta}_1$ de la regresión múltiple. Resulta que existe una relación sencilla entre $\tilde{\beta}_1$ y $\hat{\beta}_1$, que permite hacer interesantes comparaciones entre la regresión simple y la múltiple:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1, \qquad \textbf{3.23}$$

donde $\tilde{\delta}_1$ es el coeficiente de pendiente de la regresión simple de x_{i2} sobre x_{i1}, $i = 1, \ldots, n$. Esta ecuación muestra cómo difiere $\tilde{\beta}_1$ del efecto parcial de x_1 sobre $\hat{y}$. El término de confusión es el efecto parcial de x_2 sobre $\hat{y}$ multiplicado por la pendiente de la regresión muestral de x_2 sobre x_1 (vea una verificación más general en la sección 3A.4 del apéndice de este capítulo).

La relación entre $\tilde{\beta}_1$ y $\hat{\beta}_1$ muestra, además, que hay dos casos en los que $\tilde{\beta}_1$ y $\hat{\beta}_1$ son iguales:

1. El efecto parcial de x_2 sobre $\hat{y}$ es cero en la muestra. Es decir, $\hat{\beta}_2 = 0$.
2. x_1 y x_2 no están correlacionadas en la muestra. Es decir, $\tilde{\delta}_1 = 0$.

Aun cuando las estimaciones de regresión simple y de regresión múltiple casi nunca son idénticas, la fórmula anterior puede emplearse para identificar por qué estas estimaciones pueden ser o muy diferentes o bastante similares. Por ejemplo, si $\hat{\beta}_2$ es pequeño, puede esperarse que las estimaciones de β_1 en la regresión simple y en la regresión múltiple sean parecidos. En el ejemplo 3.1, la correlación muestral entre *hsGPA* y *ACT* es aproximadamente .346, que es una correlación no trivial. Pero el coeficiente de *ACT* es bastante pequeño. No extrañará encontrar que la regresión simple de *colGPA* sobre *hsGPA* produzca una estimación para la pendiente de .482, que no difiere mucho de la estimación .453 en (3.15).

Ejemplo 3.3

[Participación en los planes de pensión 401(k)]

Se usará la base de datos 401K.RAW para estimar el efecto de la tasa de aportación de la empresa al plan (*mrate*) sobre la tasa de participación (*prate*) en ese plan de pensión 401(k). La tasa de aportación es el monto que la empresa aporta al plan de ahorro de un trabajador por cada dólar que éste invierte (hasta un cierto límite); así $mrate = .75$ significa que la empresa aporta 75¢ por cada dólar que invierte el trabajador. La tasa de participación es el porcentaje de trabajadores con derecho al plan que en efecto están inscritos en un plan de pensión 401(k). La variable *age* es la antigüedad del plan 401(k). En esta base de datos hay 1,534 planes, el *prate* promedio es 87.36, el *mrate* promedio es .732 y la *age* promedio es 13.2.

Regresando *prate* sobre *mrate* y *age* se obtiene

$$\widehat{prate} = 80.12 + 5.52\, mrate + .243\, age.$$

De manera que tanto *mrate* como *age* tienen el efecto esperado. ¿Qué ocurre si no se controla *age*? El efecto estimado de *age* no es trivial y, por tanto, puede esperarse una variación grande en el efecto estimado de *mrate* si se elimina *age* de la regresión. Sin embargo, la regresión simple de *prate* sobre *mrate* da $\widetilde{prate} = 83.08 + 5.86\, mrate$. En la regresión simple, el efecto estimado de *mrate* sobre *prate* es claramente diferente del que se obtiene en la regresión múltiple, pero la diferencia no es muy grande. (La estimación en la regresión simple es apenas cerca de 6.2% mayor que la estimación en la regresión múltiple.) Esto puede explicarse por el hecho de que la correlación muestral entre *mrate* y *age* es sólo .12.

En el caso con k variables independientes, la regresión simple de y sobre x_1 y la regresión múltiple de y sobre $x_1, x_2, \ldots, x_k$ se obtiene una estimación idéntica de x_1 sólo si 1) todos los coeficientes de MCO, desde x_2 hasta x_k son cero *o si* 2) x_1 *no* está correlacionada con *ninguna* de las $x_2, \ldots, x_k$. En la práctica ninguna de estas dos cosas es muy probable. Pero si los coeficientes de x_2 a x_k son pequeños o si las correlaciones muestrales entre x_1 y las demás variables independientes son insignificantes, entonces las estimaciones para el efecto de x_1 sobre y obtenidos mediante regresión simple y regresión múltiple pueden ser similares.

Bondad de ajuste

Como en el caso de la regresión simple, se definen la **suma total de cuadrados (STC)**, la **suma explicada de cuadrados (SEC)** y la **suma residual de cuadrados** o **suma de residuales cuadrados (SRC)** como

$$\text{STC} \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \textbf{3.24}$$

$$\text{SEC} \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad \textbf{3.25}$$

$$\text{SRC} \equiv \sum_{i=1}^{n} \hat{u}_i^2. \qquad \textbf{3.26}$$

Empleando el mismo argumento que en el caso de la regresión simple, se puede mostrar que

$$\text{STC} = \text{SEC} + \text{SRC}. \qquad \textbf{3.27}$$