Sesión 16 (B) — Condicionar en $\boldsymbol{X}$: la misma fórmula, dos experimentos

Marcos Bujosa

Descripción de la práctica

Qué es condicionar en $\boldsymbol{X}$ ya está definido con precisión: la lección 9 presentó $E[Z \mid \boldsymbol{X}]$ como la proyección ortogonal de Z sobre las funciones de $\boldsymbol{X}$, y la lección 11 trabajó con esa definición. Esta práctica busca otra cosa: la *consecuencia observable* de condicionar, es decir, qué cambia en los números cuando se condiciona y cuando no. Quien practica la econometría lee errores estándar condicionales a diario, y rara vez ve qué distingue esa cifra de la que saldría sin condicionar.

En la práctica anterior mantuvimos el regresor `rooms` fijo a lo largo de las 2000 réplicas, y solo dejamos cambiar la perturbación. Lo justificamos con una frase que conviene ahora tomarse en serio: *"eso es lo que significa condicionar en $\boldsymbol{X}$"*.

Esta práctica pone esa frase a prueba haciendo lo contrario. Vamos a repetir el mismo experimento de Montecarlo dos veces:

- *Experimento A*: el regresor se fija una vez y no cambia; solo cambia la perturbación.
- *Experimento B*: en cada réplica se sortea *también* un regresor nuevo.

Si la distinción fuera irrelevante, ambos experimentos darían lo mismo. Veremos que no.
Objetivos

1. Comprobar que la fórmula $Var[\hat{\beta}_2 \mid \boldsymbol{X}] = \sigma^2 / \sum_i (X_i - \overline{X})^2$ describe el experimento A, no el B.
2. Ver qué consecuencia numérica concreta tiene la barra vertical de $Var[\cdot \mid \boldsymbol{X}]$, que las lecciones 9 y 11 ya definieron con precisión.
3. Ver que la insesgadez, en cambio, se cumple en *ambos* experimentos —y entender por qué las dos propiedades se comportan de forma distinta.
4. Apreciar por qué el resultado que reportan los programas estadísticos es siempre el condicional.

Advertencia de diseño. Para que la diferencia se vea con claridad, trabajaremos con una muestra deliberadamente *pequeña*: doce viviendas. Cuanto mayor es la muestra, menos importa la distinción que queremos ilustrar, y con las 506 observaciones de la práctica anterior costaría apreciarla. En la última actividad repetiremos el experimento con una muestra grande para comprobar precisamente eso.

Actividad 1 - El mismo mundo, pero con doce viviendas

Reutilizamos el mundo artificial de la práctica A —mismos β_1, β_2 y σ—, pero nos quedamos solo con doce viviendas, elegidas al azar. Además, anotamos la media y la desviación típica de `rooms` en el conjunto completo: harán de "población" de la que se sortean regresores nuevos en el experimento B.

o bien teclee en línea de comandos:

```
open hprice2.gdt --quiet

ols price const rooms --quiet
scalar b1_true = $coeff(const)
scalar b2_true = $coeff(rooms)
scalar sigma   = sqrt($ess/$df)

scalar mu_r = mean(rooms)              # "poblacion" del regresor
scalar sd_r = sd(rooms)
```

```
printf "Poblacion del regresor: media = %.4f   desv. tipica = %.4f\n", mu_r, sd_r

set seed 160
smpl 12 --random                       # doce viviendas elegidas AL AZAR
scalar Sxx_obs  = sum((rooms-mean(rooms))^2)
scalar var_cond = sigma^2 / Sxx_obs

printf "\nSubmuestra observada (n = %d):\n", $nobs
printf "  media(rooms) = %.4f    desv. tipica(rooms) = %.4f\n", mean(rooms), sd(rooms)
printf "  Sxx = suma de cuadrados en desviaciones = %.4f\n", Sxx_obs
printf "  Sxx tipico de una muestra de este tamano = %.4f\n", ($nobs-1)*sd_r^2
printf "\nVar[b2 | X] que predice la leccion 11 = %.1f   (desv. tipica %.1f)\n", var_cond, sqrt(var_cond)
```

```
Poblacion del regresor: media = 6,2841   desv. tipica = 0,7026

Submuestra observada (n = 12):
  media(rooms) = 6,3917    desv. tipica(rooms) = 0,5450
  Sxx = suma de cuadrados en desviaciones = 3,2670
  Sxx tipico de una muestra de este tamano = 5,4300

Var[b2 | X] que predice la leccion 11 = 13417907,5   (desv. tipica 3663,0)
```

Recuerde que `Sxx` es la abreviatura que introdujimos en la práctica A para $\sum_i (X_i - \overline{X})^2$ —el cuadrado de la norma euclídea del vector en desviaciones del regresor, o equivalentemente $n\,S^2$—. No es notación del curso, solo una comodidad para los guiones.

Dos detalles del código merecen comentario. El primero es `smpl 12 --random`: elegimos las doce viviendas *al azar*, en lugar de tomar las doce primeras. Es importante, porque `hprice2` viene ordenado por distritos censales y las primeras viviendas se parecen mucho entre sí; tomarlas sin más nos daría una submuestra rara por un motivo ajeno a lo que queremos estudiar. El segundo es `set seed`, que fija *qué* doce viviendas salen, para que el resultado sea reproducible.

Compare ahora las dos últimas cifras: el Sxx de nuestras doce viviendas y el Sxx que cabría esperar de doce viviendas cualesquiera. Guarde esa comparación: será la clave de todo lo que sigue.

Actividad 2 - Experimento A: el regresor se queda quieto

Es el experimento de la práctica anterior, sin más cambio que el tamaño muestral.

o bien teclee en línea de comandos:

```
scalar R = 4000

set seed 160
loop R --progressive --quiet
    series u = normal(0, sigma)
    series y = b1_true + b2_true*rooms + u      # rooms NO cambia
    ols y const rooms --quiet
    scalar b_fijo = $coeff(rooms)
    store "@workdir/repFijo.gdt" b_fijo
endloop
```

Actividad 3 - Experimento B: el regresor también se sortea

Ahora añadimos una línea dentro del bucle: antes de construir el regresando, sorteamos un regresor nuevo. Conceptualmente estamos diciendo "no solo podría haber sido otra la perturbación; también podrían haber sido otras las doce viviendas".

o bien teclee en línea de comandos:

```
set seed 160
loop R --progressive --quiet
    series xnew = mu_r + sd_r*normal(0,1)       # doce viviendas NUEVAS
    series u    = normal(0, sigma)
    series y    = b1_true + b2_true*xnew + u
    ols y const xnew --quiet
    scalar b_var = $coeff(xnew)
    scalar Sxx_i = sum((xnew-mean(xnew))^2)     # el Sxx de ESTA replica
    store "@workdir/repVar.gdt" b_var Sxx_i
endloop
```

Guardamos también `Sxx_i`, la suma de cuadrados en desviaciones del regresor *de cada réplica*. En el experimento A ese número era una constante; aquí es una variable aleatoria más, y en ella está la explicación de todo.

Actividad 4 - Las dos varianzas, cara a cara

o bien teclee en línea de comandos:

```
printf "=== LO QUE PREDICE LA LECCION 11 (condicional en la X observada) ===\n"
printf "beta_2 verdadero  = %.2f\n", b2_true
printf "Var[b2 | X]       = %.1f\n\n", var_cond
```

```
printf "=== EXPERIMENTO A: regresor FIJO ===\n"
printf "media de b2hat    = %.2f\n", mean(b_fijo)
printf "varianza de b2hat = %.1f\n\n", sum((b_fijo-mean(b_fijo))^2)/$nobs
```

```
printf "=== EXPERIMENTO B: regresor REGENERADO en cada replica ===\n"
printf "media de b2hat    = %.2f\n", mean(b_var)
printf "varianza de b2hat = %.1f\n", sum((b_var-mean(b_var))^2)/$nobs
printf "\nSxx a lo largo de las replicas del experimento B:\n"
printf "  medio = %.4f    minimo = %.4f    maximo = %.4f\n", mean(Sxx_i), min(Sxx_i), max(Sxx_i)
```

```
=== LO QUE PREDICE LA LECCION 11 (condicional en la X observada) ===
beta_2 verdadero  = 9119,55
Var[b2 | X]       = 13417907,5

=== EXPERIMENTO A: regresor FIJO ===
media de b2hat    = 9061,70
varianza de b2hat = 13168607,8

=== EXPERIMENTO B: regresor REGENERADO en cada replica ===
media de b2hat    = 9125,59
varianza de b2hat = 9747173,3

Sxx a lo largo de las replicas del experimento B:
  medio = 5,4482    minimo = 0,7178    maximo = 19,0432
```

Lo que debe observar

Tres hechos, en este orden:

1. *El experimento A reproduce la fórmula.* La varianza empírica de las estimaciones con el regresor fijo queda cerca de σ^2/Sxx, calculado con el Sxx de *nuestras doce viviendas*. La fórmula de la lección 11 describe este experimento.

2. *El experimento B da otra cosa.* Con el regresor regenerado, la varianza es apreciablemente distinta. No es un error de programación ni ruido de Montecarlo: los dos experimentos responden a preguntas diferentes.

3. *Las medias, en cambio, coinciden.* En ambos experimentos la media de las estimaciones queda cerca de β_2. La insesgadez sobrevive a los dos diseños; la fórmula de la varianza, no.

De dónde viene la diferencia

Mire la última línea de la salida: el recorrido de `Sxx_i` a lo largo de las réplicas del experimento B. Verá que ese número varía muchísimo de una réplica a otra, y compárelo con el Sxx fijo de nuestras doce viviendas observadas.

Ahí está todo. En el experimento A la varianza de $\hat{\beta}_2$ es σ^2 dividido por *un número fijo*. En el experimento B es, en promedio, σ^2 multiplicado por el promedio de $1/\text{Sxx}$ —y el promedio de un cociente no es el cociente de los promedios—. Si nuestras doce viviendas concretas resultan tener valores de `rooms` más apretados de lo habitual, su Sxx será menor que el típico y, en consecuencia, condicionar en ellas dará una varianza *mayor* que la que resulta de promediar sobre todos los conjuntos posibles de doce viviendas. Si hubieran salido más dispersas, ocurriría lo contrario.

Esta es la razón por la que la lección 11 se detuvo en la fórmula condicional y solo mencionó de pasada la incondicional: al intentar calcular $\text{Var}(\hat{\beta}_2)$ aparecía $\text{E}\big(\sigma^2/(nS^2)\big)$, la esperanza de un cociente, que no se simplifica.

Actividad 5 - ¿Y si la muestra fuera grande?

Repita el experimento con cuatrocientas viviendas en lugar de doce, dejando todo lo demás igual.
o bien teclee en línea de comandos:

```
open hprice2.gdt --quiet
ols price const rooms --quiet
scalar b1_true = $coeff(const)
scalar b2_true = $coeff(rooms)
scalar sigma   = sqrt($ess/$df)
scalar mu_r = mean(rooms)
scalar sd_r = sd(rooms)

set seed 160
smpl 400 --random
scalar Sxx_obs  = sum((rooms-mean(rooms))^2)
scalar var_cond = sigma^2 / Sxx_obs

set seed 160
loop 40000 --progressive --quiet      # diez veces mas replicas: ver nota
    series u = normal(0, sigma)
    series y = b1_true + b2_true*rooms + u
    ols y const rooms --quiet
    scalar bg_fijo = $coeff(rooms)
    series xnew = mu_r + sd_r*normal(0,1)
    series u2   = normal(0, sigma)
    series y2   = b1_true + b2_true*xnew + u2
    ols y2 const xnew --quiet
    scalar bg_var = $coeff(xnew)
    store "@workdir/repGrande.gdt" bg_fijo bg_var
endloop
```

```
printf "Con n = 400:  Sxx observado = %.2f   (tipico: %.2f)\n", Sxx_obs, ($nobs-1)*sd_r^2
printf "             Var[b2|X] condicional teorica = %.2f\n", var_cond
printf "             cociente Sxx observado/tipico = %.3f  <- lo que cabe esperar de var_B/var_A\n",
↪  Sxx_obs/(($nobs-1)*sd_r^2)
```

```
printf "  regresor FIJO       : var = %.2f\n", sum((bg_fijo-mean(bg_fijo))^2)/$nobs
printf "  regresor REGENERADO : var = %.2f\n", sum((bg_var-mean(bg_var))^2)/$nobs
printf "  cociente observado  : %.3f\n", (sum((bg_var-mean(bg_var))^2))/(sum((bg_fijo-mean(bg_fijo))^2))
```

```
Con n = 400:  Sxx observado = 205,76   (tipico: 196,96)
             Var[b2|X] condicional teorica = 213046,34
             cociente Sxx observado/tipico = 1,045  <- lo que cabe esperar de var_B/var_A
  regresor FIJO       : var = 214203,96
  regresor REGENERADO : var = 224464,19
  cociente observado  : 1,048
```

Lo que debe observar

Lo primero que conviene mirar no es el tamaño de la diferencia, sino su *signo*.

Con doce viviendas, la varianza del experimento B salía **menor** que la del A. Con cuatrocientas sale **mayor**. Puede parecer que algo se ha estropeado, pero es justo lo contrario: es la confirmación más limpia del mecanismo que explicamos en la actividad anterior. Allí dijimos que todo dependía de si nuestra muestra concreta tenía un Sxx más pequeño o más grande que el típico, y añadimos: "si hubieran salido más dispersas, ocurriría lo contrario". Compruébelo en las salidas:

- *Doce viviendas*: Sxx observado = 3,27 frente a un típico de 5,43. Nuestra muestra estaba **apretada**, condicionar en ella daba una varianza mayor que el promedio, y por eso B quedaba por debajo de A.
- *Cuatrocientas viviendas*: Sxx observado = 205,76 frente a un típico de 196,96. Ahora la muestra está, ligeramente, **más dispersa** de lo normal; condicionar en ella da una varianza menor que el promedio, y B queda por encima de A.

El cambio de signo no es un accidente: es la predicción, cumplida. Y de paso desactiva una moraleja falsa que sería fácil sacar de la actividad anterior —"condicionar infla la varianza"—: condicionar no la infla ni la desinfla, la *ajusta a la muestra que tenemos*, y hacia qué lado la ajuste depende de cómo sea esa muestra.

En cuanto al tamaño, la diferencia *se reduce mucho* al pasar de doce a cuatrocientas observaciones, pero no llega a desaparecer. Conviene entender por qué, porque en esa distinción hay dos efectos superpuestos:

1. *Nuestra muestra deja de ser excepcional.* Con doce viviendas, el Sxx que nos tocó podía quedar muy lejos del típico; con cuatrocientas, cualquier muestra tiene un Sxx bastante parecido al típico. Este efecto se desvanece al crecer n, pero despacio.
2. *El promedio de un cociente no es el cociente de los promedios.* Aunque el Sxx medio coincidiera con el observado, promediar σ^2/Sxx sobre las réplicas no daría σ^2 dividido por el Sxx medio. Este segundo efecto también se atenúa con n, y más deprisa que el primero.

Con $n = 400$ el segundo efecto ya es prácticamente invisible, y todo lo que queda es el primero. Eso se puede comprobar con los números delante: el cociente de varianzas que observamos debería parecerse al cociente Sxx observado/típico, que la salida imprime, y en efecto se le parece.[1]

La conclusión práctica es que con muestras grandes la distinción tiene poca consecuencia numérica. La conclusión conceptual, en cambio, no cambia en absoluto: siguen siendo dos preguntas distintas, y la que responde la fórmula de la lección 11 —y la que responde su programa estadístico— es siempre la condicional.

[1] Aquí hemos subido el número de réplicas de 4.000 a 40.000, y merece la pena decir por qué: con 4.000 réplicas el ruido de Montecarlo sobre un cociente de varianzas es de un 3 % largo, del mismo orden que el efecto de un 4,5 % que queremos medir. Repitiendo el experimento con distintas semillas, el cociente oscilaba entre 1,00 y 1,08 —es decir, con algunas semillas el efecto *parecía* desaparecer del todo—. Con 40.000 réplicas el cociente se estabiliza por encima de uno y el experimento ilustra lo que pretende ilustrar. El coste es de unos pocos segundos. Es, de paso, una lección sobre el propio método de Montecarlo: un experimento simulado tiene su propia precisión, y conviene asegurarse de que es mejor que el efecto que se quiere ver.

Preguntas de interpretación para la clase

1. ¿Qué pregunta responde el experimento A y qué pregunta, distinta, responde el experimento B? Formule ambas con palabras, sin fórmulas.

2. ¿Por qué la insesgadez se cumple en los dos experimentos, mientras que la fórmula de la varianza solo describe el primero? (Pista: ¿en qué paso de la demostración de la lección 11 interviene $\boldsymbol{X}$ de forma esencial?)

3. Cuando Gretl le muestra el error estándar de un coeficiente, ¿está respondiendo a la pregunta del experimento A o a la del B? ¿Por qué cree que es así?

4. Imagine que un investigador *diseña* el experimento y puede elegir los valores del regresor (por ejemplo, las dosis en un ensayo). A la vista de la fórmula de la lección 11, ¿qué le convendría hacer con esas dosis?

5. En la Actividad 5, con $n = 400$, las dos varianzas casi coinciden. ¿Diría entonces que la distinción entre condicionar y no condicionar es irrelevante en la práctica? Argumente a favor y en contra.

Para profundizar:

- Lección 11 (lección 11), sección "Varianza", en particular el párrafo final sobre la varianza incondicional y la ley de la varianza total.
- Lección 9 (lección 9) para la esperanza condicional como proyección.
- Wooldridge, J. M. (2020). *Introductory Econometrics*, cap. 2, sección 2.5.

Código completo de la práctica

```
# ------------------------------------------------------------
# Copyright (C) 2026 Marcos Bujosa
# Licencia: GNU General Public License v3.0 o posterior
# Este código es software libre y puede ser redistribuido y/o modificado bajo los términos de la GPL.
# Ver el archivo LICENSE del repositorio para más detalles.
# ------------------------------------------------------------

# Los dos primeros comandos son necesarios para que Gretl guarde los resultados de la práctica en el
↪  directorio de trabajo
# al ejecutar lo siguiente desde un terminal (use los nombres y ruta que correspondan)
#
#    DIRECTORIO="Nombre_Directorio_trabajo" gretlcli -b ruta/nombre_fichero_de_la_practica.inp
#
# Si esto no le funciona en su sistema, comente las siguientes dos líneas y sitúese en el directorio de
↪  trabajo de gretl
# que corresponda (configure dicho directorio de trabajo desde la ventana principal de Gretl).

string directory = getenv("DIRECTORIO")
set workdir "@directory"

set verbose off

open hprice2.gdt --quiet

ols price const rooms --quiet
scalar b1_true = $coeff(const)
scalar b2_true = $coeff(rooms)
scalar sigma   = sqrt($ess/$df)

scalar mu_r = mean(rooms)               # "poblacion" del regresor
scalar sd_r = sd(rooms)
outfile --quiet preparacion.txt
    printf "Poblacion del regresor: media = %.4f   desv. tipica = %.4f\n", mu_r, sd_r

    set seed 160
    smpl 12 --random                        # doce viviendas elegidas AL AZAR
    scalar Sxx_obs  = sum((rooms-mean(rooms))^2)
    scalar var_cond = sigma^2 / Sxx_obs

    printf "\nSubmuestra observada (n = %d):\n", $nobs
    printf "  media(rooms) = %.4f    desv. tipica(rooms) = %.4f\n", mean(rooms), sd(rooms)
    printf "  Sxx = suma de cuadrados en desviaciones = %.4f\n", Sxx_obs
    printf "  Sxx tipico de una muestra de este tamano = %.4f\n", ($nobs-1)*sd_r^2
    printf "\nVar[b2 | X] que predice la leccion 11 = %.1f   (desv. tipica %.1f)\n", var_cond,
    ↪  sqrt(var_cond)
end outfile

scalar R = 4000

set seed 160
loop R --progressive --quiet
    series u = normal(0, sigma)
    series y = b1_true + b2_true*rooms + u      # rooms NO cambia
    ols y const rooms --quiet
    scalar b_fijo = $coeff(rooms)
    store "@workdir/repFijo.gdt" b_fijo
endloop

set seed 160
loop R --progressive --quiet
    series xnew = mu_r + sd_r*normal(0,1)       # doce viviendas NUEVAS
    series u    = normal(0, sigma)
    series y    = b1_true + b2_true*xnew + u
    ols y const xnew --quiet
```

```
    scalar b_var = $coeff(xnew)
    scalar Sxx_i = sum((xnew-mean(xnew))^2)     # el Sxx de ESTA replica
    store "@workdir/repVar.gdt" b_var Sxx_i
endloop

outfile --quiet comparacion.txt
    printf "=== LO QUE PREDICE LA LECCION 11 (condicional en la X observada) ===\n"
    printf "beta_2 verdadero  = %.2f\n", b2_true
    printf "Var[b2 | X]       = %.1f\n\n", var_cond
end outfile
open "@workdir/repFijo.gdt" --quiet
outfile --quiet --append comparacion.txt
    printf "=== EXPERIMENTO A: regresor FIJO ===\n"
    printf "media de b2hat    = %.2f\n", mean(b_fijo)
    printf "varianza de b2hat = %.1f\n\n", sum((b_fijo-mean(b_fijo))^2)/$nobs
end outfile
open "@workdir/repVar.gdt" --quiet
outfile --quiet --append comparacion.txt
    printf "=== EXPERIMENTO B: regresor REGENERADO en cada replica ===\n"
    printf "media de b2hat    = %.2f\n", mean(b_var)
    printf "varianza de b2hat = %.1f\n", sum((b_var-mean(b_var))^2)/$nobs
    printf "\nSxx a lo largo de las replicas del experimento B:\n"
    printf "  medio = %.4f    minimo = %.4f    maximo = %.4f\n", mean(Sxx_i), min(Sxx_i), max(Sxx_i)
end outfile

open hprice2.gdt --quiet
ols price const rooms --quiet
scalar b1_true = $coeff(const)
scalar b2_true = $coeff(rooms)
scalar sigma   = sqrt($ess/$df)
scalar mu_r = mean(rooms)
scalar sd_r = sd(rooms)

set seed 160
smpl 400 --random
scalar Sxx_obs  = sum((rooms-mean(rooms))^2)
scalar var_cond = sigma^2 / Sxx_obs

set seed 160
loop 40000 --progressive --quiet      # diez veces mas replicas: ver nota
    series u = normal(0, sigma)
    series y = b1_true + b2_true*rooms + u
    ols y const rooms --quiet
    scalar bg_fijo = $coeff(rooms)
    series xnew = mu_r + sd_r*normal(0,1)
    series u2   = normal(0, sigma)
    series y2   = b1_true + b2_true*xnew + u2
    ols y2 const xnew --quiet
    scalar bg_var = $coeff(xnew)
    store "@workdir/repGrande.gdt" bg_fijo bg_var
endloop

outfile --quiet muestraGrande.txt
    printf "Con n = 400:  Sxx observado = %.2f   (tipico: %.2f)\n", Sxx_obs, ($nobs-1)*sd_r^2
    printf "             Var[b2|X] condicional teorica = %.2f\n", var_cond
    printf "             cociente Sxx observado/tipico = %.3f  <- lo que cabe esperar de var_B/var_A\n",
    ↪  Sxx_obs/(($nobs-1)*sd_r^2)
end outfile
open "@workdir/repGrande.gdt" --quiet
outfile --quiet --append muestraGrande.txt
    printf "  regresor FIJO       : var = %.2f\n", sum((bg_fijo-mean(bg_fijo))^2)/$nobs
    printf "  regresor REGENERADO : var = %.2f\n", sum((bg_var-mean(bg_var))^2)/$nobs
    printf "  cociente observado  : %.3f\n",
    ↪  (sum((bg_var-mean(bg_var))^2))/(sum((bg_fijo-mean(bg_fijo))^2))
end outfile
```

Enlace al guión: S16-Prct-B-condicionar.inp

Respuestas

1. *Las dos preguntas.* El experimento A pregunta: "dadas *estas* doce viviendas, con estos valores concretos de `rooms`, ¿cuánto variaría mi estimación si el azar hubiera repartido otras perturbaciones?". El experimento B pregunta algo más amplio: "¿cuánto variaría mi estimación entre todos los estudios posibles de doce viviendas, considerando que también podrían haberme tocado otras viviendas?". La primera toma la muestra de regresores como un dato del problema; la segunda la considera parte de lo que podría haber sido distinto.

2. *Por qué la insesgadez sobrevive.* Porque la demostración de la insesgadez pasa por $E[\hat{\beta}_2 \mid \boldsymbol{X}] = \beta_2\, 1$ y después aplica la ley de las esperanzas iteradas: el resultado condicional *no depende de* $\boldsymbol{X}$, es la misma constante sea cual sea la matriz de regresores, y por tanto su promedio sobre todas las $\boldsymbol{X}$ posibles vuelve a ser esa constante. Con la varianza ocurre lo contrario: $Var[\hat{\beta}_2 \mid \boldsymbol{X}] = \sigma^2/\text{Sxx}$ *sí depende* de $\boldsymbol{X}$, así que promediar sobre las $\boldsymbol{X}$ posibles da un número distinto del que corresponde a una $\boldsymbol{X}$ concreta.

3. *Lo que reporta Gretl.* Responde a la pregunta del experimento A: el error estándar se calcula con la matriz de datos efectivamente observada, sustituyendo σ^2 por su estimación $\mathfrak{s}^2$. Tiene sentido que sea así por dos motivos. Uno práctico: no conocemos la distribución de la que salieron los regresores, así que no podríamos promediar sobre ella aunque quisiéramos. Y otro conceptual, más importante: la pregunta que le interesa a quien analiza unos datos es cuánta confianza merece *su* estimación, con *su* muestra —no cuánta merecería en promedio un estudio hipotético que nunca hizo.

4. *Si se pueden elegir los valores del regresor.* La fórmula dice que la varianza es σ^2 dividido por $\sum_i (X_i - \overline{X})^2$. Conviene, por tanto, hacer ese denominador lo más grande posible: elegir dosis muy separadas entre sí, en los extremos del rango admisible, en lugar de concentrarlas en torno a un valor central. Es el mismo principio que hemos visto al revés en esta práctica: nuestras doce viviendas tenían valores de `rooms` poco dispersos, y eso encarecía la estimación. (Naturalmente, esto supone que el modelo lineal es válido en todo ese rango: separar mucho las dosis mejora la precisión, pero solo si la relación sigue siendo lineal allí.)

5. *¿Es irrelevante con n grande?* A favor de que lo es: numéricamente las dos varianzas se acercan mucho —la diferencia pasa de ser enorme con doce observaciones a ser modesta con cuatrocientas—, así que a efectos de calcular un error estándar la distinción cambia poco las cifras. En contra: primero, que "n grande" no siempre se da —con muestras pequeñas, que abundan en economía aplicada, la diferencia es la que hemos visto—; y segundo, que la distinción no es solo numérica, sino conceptual: aclara qué población de muestras estamos imaginando cuando decimos que un estimador "varía". Confundir ambas preguntas lleva a interpretar mal qué significa exactamente un error estándar.