

Contents

Resumen	1
1 ¿Qué es el ajuste de covariables?	1
2 Control de covariables en la etapa de diseño (bloques)	2
3 Cómo controlar las covariables en una regresión	2
4 Por qué hacer ajuste de covariables	3
5 ¿Qué tanto ayuda el ajuste de covariables?	5
6 Control de las covariables con poder de pronóstico independientemente de si presentan imbalances	8
7 Cuándo no debemos hacer ajuste de covariables	10
8 Preocupaciones sobre el sesgo en muestras pequeñas	11
9 Cómo hacer que las decisiones sobre el ajuste de covariables sean transparentes	11
10 Las covariables pueden ayudarlo a investigar la integridad de la asignación aleatoria	12
Lecturas complementarias	12

Resumen

Esta guía¹ le servirá para entender cuándo tiene sentido tratar de “controlar otras variables” al estimar los efectos del tratamiento utilizando datos experimentales. Nos centramos en las grandes ideas y proporcionamos ejemplos en R.

1 ¿Qué es el ajuste de covariables?

Las “covariables” son las características básicas de los sujetos experimentales. Cuando realizamos un experimento estamos principalmente interesados en recopilar datos sobre las variables de resultado que su intervención puede afectar, p. Ej. decisiones de gasto, actitudes hacia la democracia o contribuciones para un bien público en un experimento de laboratorio. Pero también es una buena idea recopilar datos sobre las características iniciales de los sujetos antes de que ocurra la asignación del tratamiento, p. Ej. género, nivel de educación o grupo étnico. Si hace esto, puede explorar cómo los efectos del tratamiento varían con estas características (consulte 10 cosas que debe saber sobre los efectos heterogéneos del tratamiento). Recolectar esta información también le permitirá realizar un ajuste de covariables.

El ajuste de covariables forma de llamar al control de variables pre-tratamiento al estimar los efectos del tratamiento. A menudo esto se hace para mejorar la precisión. Es probable que los resultados de los sujetos tengan alguna correlación con las variables que pueden medirse antes de la asignación aleatoria. Tener en cuenta variables como el género le permitirá dejar de lado la variación en los resultados que predicen estas variables de referencia inicial (baseline), de modo que pueda aislar el efecto del tratamiento en los resultados con mayor precisión y poder.

El ajuste de covariables puede ser una forma menos costosa para mejorar la precisión que aumentar el número de sujetos en el experimento. En parte por esa razón, los investigadores a menudo recopilan datos extensos

¹Autor: Lindsay Dolan. Revisiones: Don Green y Winston Lin, Nov 1, 2016. Esta guía es un documento en vivo y sujeto a actualización por parte de los miembros de EGAP en cualquier momento; los colaboradores mencionados no son responsables por las ediciones posteriores. Gracias a Macartan Humphreys y Diana Mutz por sus útiles discusiones.

sobre los valores de covariables antes de la asignación aleatoria del tratamiento. Ensayos previos (medidas que son análogas a la variable de resultado pero están restringidas a períodos de tiempo previos a la asignación aleatoria) pueden ser especialmente valiosos para predecir resultados y en las encuestas de referencia inicial puede preguntarle a los sujetos sobre otras características de contexto.

2 Control de covariables en la etapa de diseño (bloques)

La mejor manera de controlar las covariables es usar la aleatorización en bloques para hacerlo en la etapa de diseño, antes de iniciar el experimento. La aleatorización en bloques nos permite crear grupos de tratamiento y control que están balanceados en determinadas covariables. Por ejemplo, podría esperar que el género y los ingresos ayuden a predecir la variable de resultado. La aleatorización en bloques puede garantizar que los grupos de tratamiento y control tengan proporciones iguales de poblaciones de mujeres con ingresos altos, mujeres con ingresos bajos, hombres con ingresos altos y hombres con ingresos bajos. Cuando las variables usadas para definir los bloques ayudan a predecir los resultados, los bloques mejoran la precisión al evitar correlaciones al azar entre la asignación del tratamiento y las covariables pre-tratamiento.

Para obtener más información sobre cómo construir bloques y cómo implementarlos en R, consulte 10 cosas que debe saber sobre la aleatorización. Las ganancias de precisión de los bloques (en relación con el ajuste de covariables sin bloques) tienden a ser mayores cuando los tamaños de muestra son pequeños².

Cuando realizamos aleatorizaciones en bloques para mejorar la precisión, los errores estándar (EE) estimados deben tener en cuenta los bloques. (De lo contrario, los EE tenderán a ser conservadores porque no le darán crédito por la mejora de precisión que se obtuvo con los bloques). Un método simple y comúnmente utilizado es hacer una regresión de la variable de resultado en la variable ficticia de asignación de tratamiento, así como en las variables indicadoras de los bloques. Cuando la probabilidad de asignación al tratamiento es constante en todos los bloques, incluir las variables indicadoras de los bloques en la regresión no cambia el efecto estimado del tratamiento, pero tiende a dar una estimación más precisa del EE³.

Si la probabilidad de asignación al tratamiento varía de acuerdo a los bloques, entonces debemos controlar estas desigualdad en las probabilidades para obtener estimaciones no sesgadas de los efectos promedio del tratamiento. 10 cosas que debe saber sobre la aleatorización analiza las formas de hacer esto.

3 Cómo controlar las covariables en una regresión

Puede que a veces no podamos implementar un diseño experimental con bloques (por ejemplo, si nos vinculamos a un proyecto después de que ocurra una asignación aleatoria) o queramos simplificar el esquema de asignación al azar para reducir el riesgo de errores administrativos. Aun en este caso podemos ajustar las covariables en el “back-end” mediante el uso de regresión múltiple. Recuerde que en una regresión bivariada, es decir cuando regresamos una variable de resultado solo en el indicador de tratamiento, el coeficiente de tratamiento es solo una diferencia en las medias. Este simple método proporciona una estimación insesgada del efecto promedio del tratamiento (Average Treatment Effect, ATE). Cuando agregamos al modelo covariables pre-tratamiento que están correlacionadas con la variable de resultado, el coeficiente de tratamiento es una estimación aproximadamente insesgada del ATE que tiende a ser más precisa que la regresión bivariada.

Para ajustar las covariables mediante regresión múltiple, puede usar el siguiente modelo:

$$Y_i = \alpha + \beta Z_i + \gamma X_i + \epsilon_i$$

donde Y_i es la variable de resultado, Z_i es el indicador del tratamiento y X_i es un vector de una o más covariables. El término restante, ϵ_i , es un error, es decir el remanente del ruido no explicado.

²Miratrix, Luke W., Jasjeet S. Sekhon y Bin Yu (2013). “Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments.” *Journal of the Royal Statistical Society, Series B* 75: 369–396.

³ver p. Ej, páginas 217–219 de Miriam Bruhn y David McKenzie (2009), “In Pursuit of Balance: Randomization in Practice in Development Field Experiments,” *American Economic Journal: Applied Economics* 1 (4): 200–232.

Cuando los grupos de tratamiento y control tienen tamaños diferentes, lo que ganamos en precisión gracias al ajuste de covariables puede ser mayor si incluye interacciones entre el tratamiento y las covariables (puede ver esta publicación de blog para una discusión más detallada). Para facilitar la interpretación, puede centrar las covariables para que tengan una media cero:

$$Y_i = \alpha + \beta Z_i + \gamma W_i + \delta Z_i * W_i + \epsilon_i$$

donde $W_i = X_i - \bar{X}$ y \bar{X} es el valor medio de X_i para toda la muestra.

Si los sujetos reciben diferentes probabilidades de asignación al tratamiento en función de sus covariables, entonces nuestro método de estimación debe tener en cuenta esto (nuevamente, consulte 10 cosas que necesita saber sobre la aleatorización para obtener más detalles).

4 Por qué hacer ajuste de covariables

No es necesario controlar las covariables cuando estimamos el efecto promedio del tratamiento en un ensayo controlado aleatorio que asigna a cada sujeto la misma probabilidad de recibir el tratamiento. La diferencia en el promedio no ajustado de la variable de resultado entre tratamiento y control es un estimador insesgado del ATE. Sin embargo, el ajuste de covariables tiende a mejorar la precisión si las covariables son buenos predictores de la variable resultado⁴.

En muestras grandes, la asignación aleatoria tiende a producir grupos de tratamiento y control con características iniciales similares. Aún así, por “simple azar”, los sujetos en grupo puede tener un mayor nivel de educación, o un grupo puede tener tasas de votación ligeramente más altas en elecciones anteriores, o un grupo puede ser un poco mayor en promedio. Por esta razón, el ATE estimado está sujeto a una “variabilidad de muestreo”, lo que significa que obtendrá estimaciones del ATE que se produjeron con un método insesgado pero que no dieron en el blanco⁵. Una alta variabilidad de muestreo contribuye al ruido (imprecisión), no al sesgo.

El control de estas covariables tiende a mejorar la precisión si las covariables predicen los resultados potenciales. Puede revisar el siguiente ejemplo, que se basa en el experimento de Giné y Mansuri sobre el comportamiento del voto femenino en Pakistán⁶. En este experimento, los autores asignaron al azar una campaña de información a mujeres en Pakistán para estudiar los efectos en su comportamiento de participación política, la independencia de su elección de candidato y su conocimiento político. Llevaron a cabo una encuesta de referencia inicial que les brindó información sobre varias covariables.

El siguiente código recrea este experimento creando datos falsos para cuatro de las covariables que los autores recopilan: si la mujer posee una tarjeta de identificación, si la mujer tiene educación formal, la edad de la mujer y si la mujer tiene acceso a la televisión. También crea dos resultados potenciales (los resultados que ocurrirían si fueran asignadas al tratamiento y si no) para tener una medida de hasta qué punto la elección del candidato de las mujeres era independiente de las opiniones de los hombres de su familia. Los resultados potenciales están correlacionados con las cuatro covariables, y el efecto de tratamiento “real” incorporado en

⁴Una breve revisión del sesgo y la precisión: Imagine replicar el experimento muchas veces (sin cambiar la muestra y las condiciones experimentales, pero volviendo a realizar la asignación aleatoria cada vez). Un estimador insesgado puede sobrestimar o subestimar el ATE en cualquier réplica dada, pero su valor esperado (el promedio de todas las réplicas posibles) será igual al ATE verdadero. Por lo general, preferimos estimadores insesgados o aproximadamente insesgados, pero también nos importa la precisión (que se define formalmente como la inversa de la varianza). Imagine que está lanzando un dardo a un tablero de dardos. Si en promedio tiende a golpear el centro del tablero de dardos, pero sus tiros a menudo están lejos de la marca, tiene un estimador imparcial pero impreciso. Si golpea cerca del centro cada vez, su estimador es más preciso. Un investigador puede preferir aceptar un pequeño sesgo a cambio de una gran mejora en la precisión. Un posible criterio para evaluar los estimadores es el error cuadrático medio, que es igual a la varianza más el cuadrado del sesgo. Véase, por ejemplo, Sharon Lohr (2010), *Sampling: Design and Analysis*, 2ª ed., Págs. 31-32.

⁵“La variabilidad de muestreo” se refiere a la dispersión de las estimaciones que se produce simplemente debido a las diferentes asignaciones aleatorias que podrían haberse realizado. Cuando por simple azar la asignación aleatoria produce un grupo de tratamiento con más “A” y un grupo de control con más “B”, es más difícil separar las características de contexto (A y B) de la asignación de tratamiento como predictor de los resultados observados.

⁶Giné, Xavier y Ghazala Mansuri (2012). “Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan.”

la medida de independencia aquí es 1. Para determinar si nuestro estimador tiene sesgo o no, simulamos 10,000 réplicas de nuestro experimento. En cada repetición, asignamos aleatoriamente un tratamiento y luego hacemos una regresión del resultado observado Y en el indicador de tratamiento Z , con y sin controlar las covariables. Por lo tanto, estamos simulando dos métodos (no ajustado y ajustado por covariables) para estimar el ATE. Para estimar el sesgo de cada método, tomamos la diferencia entre el promedio de las 10,000 estimaciones simuladas y el efecto “real” del tratamiento.

```
rm(list=ls())

set.seed(20140714)
N = 2000
N.treated = 1000
Replications = 10000

true.treatment.effect = 1

# Crea covariables pre-tratamiento
owns.id.card = rbinom(n = N, size = 1, prob = .18)
has.formal.schooling = rbinom(n = N, size = 1, prob = .6)
age = round(rnorm(n = N, mean = 37, sd = 16))
age[age<18] = 18
age[age>65] = 65
TV.access = rbinom(n = N, size = 1, prob = .7)
epsilon = rnorm(n = N, mean = 0, sd = 2)

# Crea resultados potenciales correlacionados con las variables pre-tratamiento
Y0 = round(owns.id.card + 2*has.formal.schooling + 3*TV.access + log(age) + epsilon)
Y1 = Y0 + true.treatment.effect

# Asigna el tratamiento repetidamente
Z.mat = replicate(Replications, ifelse(1:N %in% sample(1:N, N.treated), 1, 0))

# Genera la variable de resultado observada
Y.mat = Y1 * Z.mat + Y0 * (1 - Z.mat)

diff.in.means = function(Y, Z) {
  coef(lm(Y ~ Z))[2]
}

ols.adjust = function(Y, Z) {
  coef(lm(Y ~ Z + owns.id.card + has.formal.schooling + age + TV.access))[2]
}

unadjusted.estimates = rep(NA, Replications)
adjusted.estimates = rep(NA, Replications)

for (i in 1:Replications) {
  unadjusted.estimates[i] = diff.in.means(Y.mat[,i], Z.mat[,i])
  adjusted.estimates[i] = ols.adjust(Y.mat[,i], Z.mat[,i])
}

# Variabilidad estimada (desviación estándar) de cada estimador
sd.of.unadj = sd(unadjusted.estimates)
sd.of.unadj
```

```
sd.of.adj = sd(adjusted.estimates)
sd.of.adj

# Sesgo estimado de cada estimador
mean(unadjusted.estimates) - true.treatment.effect
mean(adjusted.estimates) - true.treatment.effect

# Margen de error ( 95% nivel de confianza) para cada sesgo estimado
1.96 * sd.of.unadj / sqrt(Replications)
1.96 * sd.of.adj / sqrt(Replications)
```

Ambos métodos, con y sin covariables, producen el efecto real del tratamiento de 1 en promedio. Cuando ejecutamos la regresión sin covariables, nuestro ATE estimado promedió 1.0008 entre las 10.000 réplicas, y con covariables, promedió 1.0003. Como la estimación ajustada por regresión es esencialmente insesgada a pesar de que nuestro modelo de regresión está mal especificado; controlamos la edad linealmente cuando el proceso de generación de datos verdaderos involucra el logaritmo de la edad⁷.

Los beneficios reales provienen de la precisión de nuestras estimaciones. El error estándar (la desviación estándar de la distribución muestral) de nuestro ATE estimado cuando ignoramos las covariables es 0.121. Cuando incluimos covariables en el modelo, nuestra estimación se vuelve un poco más estricta: el error estándar es 0.093. Debido a que nuestras covariables predicen la variable de resultado, incluirlas en la regresión explica algo de ruido en nuestros datos, lo que ajusta nuestra estimación del ATE.

5 ¿Qué tanto ayuda el ajuste de covariables?

¿Cuándo es más probable que el ajuste de covariables mejore la precisión?

El ajuste de covariables será más útil cuando sus covariables sean fuertemente predictivas (o “pronósticas”) de sus resultados. El ajuste de covariables esencialmente le permite hacer uso de la información sobre las relaciones entre las características pre-tratamiento y su variable de resultado para que pueda identificar mejor la relación entre el tratamiento y la variable de resultado. Pero si las características de contexto de base están solo débilmente correlacionadas con la variable de resultado, el ajuste de covariables no servirá de mucho. Las covariables por las que querrá ajustar son las que están fuertemente correlacionadas con los resultados.

El siguiente gráfico representa la relación entre la capacidad de pronóstico de la covariable y la ganancia que obtenemos al ajustarla. En el eje x está el tamaño de la muestra, y en el eje y está la raíz del error cuadrado medio (Root-Mean-Square Deviation, RMSE), la raíz cuadrada de la diferencia DE media entre el estimador y el ATE verdadero al cuadrado. Queremos que nuestro RMSE sea pequeño y el ajuste de covariables debería ayudarnos a reducirlo.

```
rm(list=ls())
library(MASS) # para mvrnorm()
set.seed(1234567)
num.reps = 10000

# Error real del efecto del tratamiento es 0 para cada unidad

adj.est = function(n, cov.matrix, treated) {
  Y.and.X = mvrnorm(n, mu = c(0, 0), Sigma = cov.matrix)
  Y = Y.and.X[, 1]
  X = Y.and.X[, 2]
  coef(lm(Y ~ treated + X))[2]
}
```

⁷El sesgo estimado es 0.0003 con un margen de error (en el nivel de confianza del 95%) de 0,0018.

```

unadj.est = function(n, treated) {
  Y = rnorm(n)
  coef(lm(Y ~ treated))[2]
}

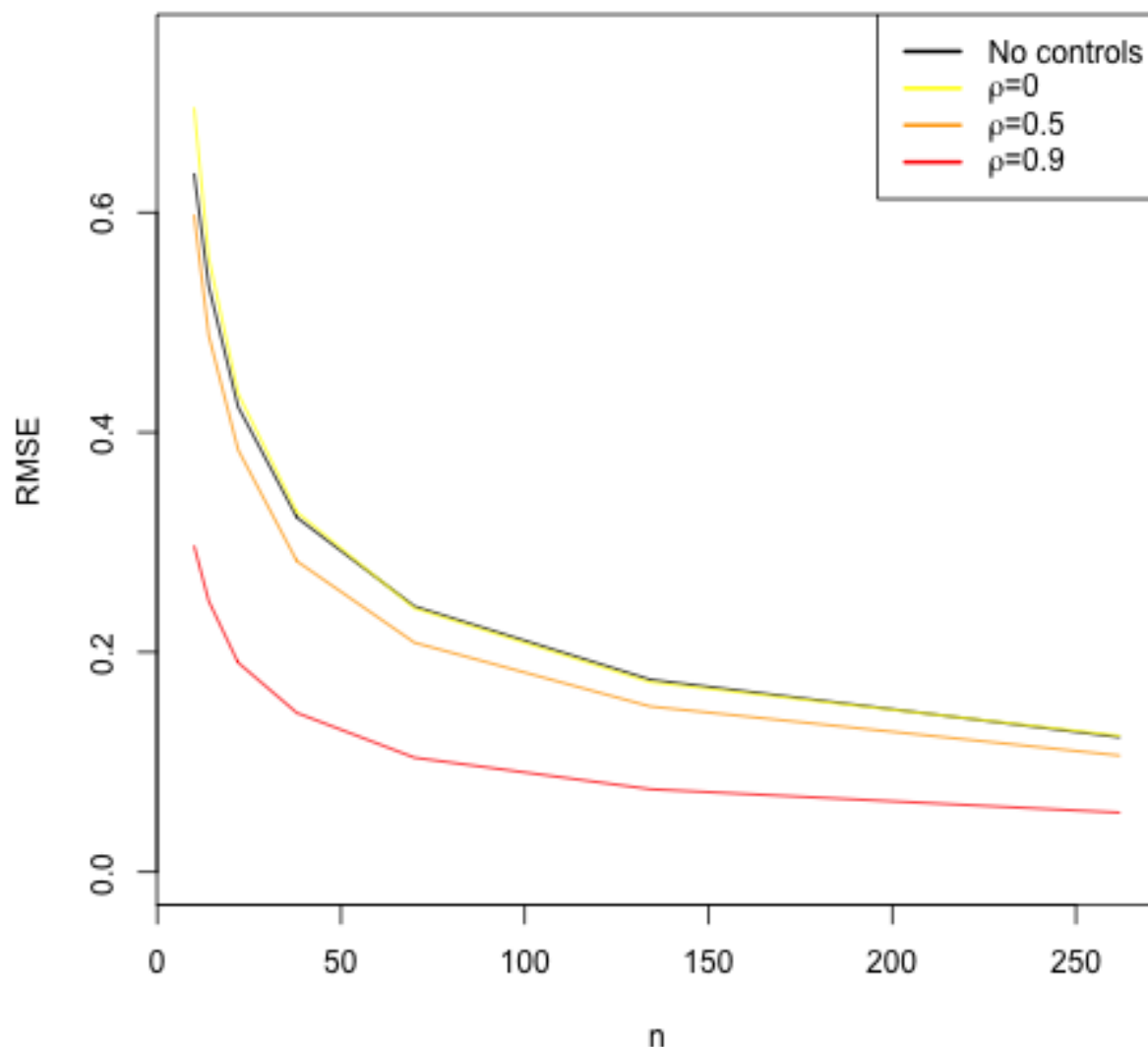
rmse = function(half.n, rho = 0, control = TRUE) {
  treated = rep(c(0, 1), half.n)
  n = 2 * half.n

  if (control) {
    cov.matrix = matrix(c(1, rho, rho, 1), nrow = 2, ncol = 2)
    return( sqrt(mean(replicate(num.reps, adj.est(n, cov.matrix, treated)) ^ 2)) )
  }
  else {
    return( sqrt(mean(replicate(num.reps, unadj.est(n, treated)) ^ 2)) )
  }
}

half.n = c(5, 7, 11, 19, 35, 67, 131)
n = 2 * half.n
E = sapply(half.n, rmse, control = FALSE)
E0 = sapply(half.n, rmse, rho = 0)
E1 = sapply(half.n, rmse, rho = 0.5)
E2 = sapply(half.n, rmse, rho = 0.9)

plot(n, E, type = "l", ylab = "RMSE", xlim = c(min(n), max(n)), ylim = c(0, .75))
lines(n, E0, col = "yellow")
lines(n, E1, col = "orange")
lines(n, E2, col = "red")
legend(x = 'topright',
      c("No controls",
        expression(paste(rho, "=0")), expression(paste(rho, "=0.5")),
        expression(paste(rho, "=0.9"))),
      col=c("black", "yellow", "orange", "red"), lty = 1, lwd=2)

```



La línea negra muestra el RMSE cuando no ajustamos la covariable. La línea roja muestra el RMSE cuando ajustamos la covariable de alto pronóstico (la correlación entre la covariable y el resultado es 0.9). Como puede ver la línea roja siempre está por debajo de la línea negra, lo que quiere decir que el RMSE es más bajo cuando se ajusta una covariable de pronóstico. La línea naranja representa el RMSE cuando ajustamos una covariable de pronóstico moderado (la correlación entre la covariable y el resultado es 0.5). Todavía estamos obteniendo ganancias en precisión en relación con la línea negra, pero no tanto como lo hicimos con la línea roja. Finalmente, la línea amarilla muestra lo que sucede si controla una covariable que no predice en absoluto el resultado. La línea amarilla es casi idéntica a la línea negra. No recibió ninguna mejora en la precisión al controlar una covariable no pronóstica; de hecho, pagó una pequeña multa porque desperdició cierto grado de libertad, lo que es especialmente costoso cuando el tamaño de la muestra es pequeño. Este ejercicio demuestra que obtendrá la mayor ganancia en precisión al controlar las covariables que predicen fuertemente la variable de resultado.

¿Cómo puede saber qué covariables tienen capacidad de predecir la variable de resultado antes de iniciar su experimento? Experimentos exploratorios o incluso estudios observacionales pueden servir como guía para

entender qué características de referencia inicial predicen mejor los resultados.

6 Control de las covariables con poder de pronóstico independientemente de si presentan imbalances

Las covariables generalmente deben elegirse en función de su capacidad esperada para ayudar a predecir los resultados, independientemente de si muestran “imbalances” (es decir, independientemente de si existen diferencias notables entre el grupo de tratamiento y el grupo de control en los valores promedio u otros aspectos de distribuciones de covariables). Hay dos razones por las que hacemos esta recomendación:

1. La inferencia estadística frecuentista (errores estándar, intervalos de confianza, valores p , etc.) supone que el análisis sigue una estrategia predefinida. La elección de covariables sobre la base de los imbalances observados dificulta que podamos hacer inferencias que reflejen la estrategia real. Por ejemplo, supongamos que decidimos no controlar el género porque los grupos de tratamiento y control tienen una composición de género similar, pero *debería* haberlo controlado el género si hubiera habido un imbalance notable. Los métodos convencionales para estimar los errores estándar suponen incorrectamente que no controlamos el género sin importar cuánto imbalance se observe.
2. Como explicamos anteriormente, el ajuste de una covariable con alta capacidad de pronóstico tiende a mejorar la precisión. Para recibir el debido crédito por esta mejora en la precisión, debe ajustar las covariables incluso si no hay imbalance. Por ejemplo, suponga que el género está altamente correlacionado con la variable de resultado, pero sucede que el grupo de tratamiento y el grupo de control tienen exactamente la misma composición de género. En este caso, la estimación no ajustada del ATE será exactamente la misma que la estimación ajustada de una regresión de la variable de resultado sobre el tratamiento y el género, pero sus errores estándar serán diferentes. El EE de la estimación no ajustada tiende a ser mayor porque asume que incluso si los grupos de tratamiento y control tuvieran composiciones de género muy diferentes, seguiríamos usando el tratamiento no ajustado, es decir, control de la diferencia del valor medio de la variable de resultado (que probablemente estaría lejos del verdadero ATE en ese caso). Si especifica con anticipación que se ajustará el género independientemente de cuánto o qué poco imbalance haya, tenderá a obtener SE más pequeños, intervalos de confianza más estrictos y pruebas de significancia más poderosas.

Suponiendo que la asignación aleatoria se implementó correctamente, ¿debería la prueba de imbalances desempeñar algún papel en la elección de las covariables por las que se debe ajustar? Aquí hay una muestra de diferentes perspectivas:

- Mutz, Pemantle y Pham (2016) argumentan que a menos que exista un deserción diferencial, la práctica de seleccionar covariables basada en imbalances observados “no sólo es innecesaria” sino que “ni siquiera es útil ... y de hecho puede ser perjudicial”, porque invalida los intervalos de confianza, empeora la precisión (relativo al ajuste pre-especificado para las covariables de pronóstico), y abre la puerta a la “pesca” de resultados⁸.

Basándose en teoría y en el uso de simulaciones, Permutt (1990) estudia escenarios específicos y encuentra que cuando se usa una prueba de imbalance para decidir si ajustar una covariable, la prueba de significancia del efecto del tratamiento es conservadora (es decir, tiene una verdadera probabilidad de error Tipo I por debajo de su nivel nominal). Según Putnam, “se puede lograr un mayor poder ajustando una covariable que esté altamente correlacionada con la variable de resultado independientemente de su distribución entre los grupos”. Sin embargo, no descarta por completo considerar los imbalances observados: “La elección de covariables sobre la base de la diferencia entre el promedio de los grupos de tratamiento y control no es irracional. Después de todo, algunos errores de tipo I pueden ser más graves que otros. Una diferencia significativa en el resultado que puede explicarse como el efecto de una covariable puede ser un error más penoso que informar sobre uno que desaparece en la replicación pero sin una explicación obvia. Consideraciones similares pueden aplicarse a los errores de tipo II. Un resultado positivo que depende del ajuste de una covariable puede verse

⁸Diana C. Mutz, Robin Pemantle y Philip Pham (2016), “Model Choice in Experimental Design: Messy Analyses of Clean Data.”

como menos convincente que una prueba positiva de dos muestras de todos modos, por lo que el error de no llegar a una conclusión tan positiva puede ser menos grave. Estas justificaciones, sin embargo, son ajenas a la teoría formal de probar hipótesis⁹.

- Altman (2005) nos dice que: “parece mucho más preferible elegir qué variables ajustar sin tener en cuenta el conjunto de datos real disponible”. Altman recomienda controlar las covariables de alto pronóstico, así como las que se usaron para crear bloques. Sin embargo, también analiza un dilema: “En la práctica, el imbalance puede surgir cuando no se ha anticipado la posible necesidad de ajuste. ¿Qué deberían hacer los investigadores? Podrían ignorar el imbalance; esto sería correcto como se señaló anteriormente. La dificultad es entonces la credibilidad. Los lectores de su artículo (incluidos los revisores y editores) pueden cuestionar si el hallazgo observado ha sido influenciado por la distribución desigual de una o más covariables de referencia inicial. Es posible y aconsejable, llevar a cabo un análisis ajustado, pero ahora siendo explícito de que se trata de un análisis exploratorio más que definitivo, y que el análisis no ajustado debe tomarse como el principal. Obviamente, si los análisis simples y ajustados arrojan sustancialmente el mismo resultado, entonces no hay dificultad de interpretación. Este suele ser el caso. Sin embargo, si los resultados de los dos análisis difieren, entonces tendríamos un problema real. El hecho de que haya esta diferencia debe generar dudas sobre la veracidad del resultado general (no ajustado). La situación es similar a las dificultades de interpretación que surgen con las comparaciones de subgrupos no planificadas. Una sugerencia en tales circunstancias es tratar de imitar lo que se habría hecho si el problema se hubiera anticipado, es decir, ajustar no por las variables que se observa que están desequilibradas, sino por todas las variables que se habrían identificado de antemano con poder de pronóstico. Se podría utilizar una fuente independiente para identificar tales variables. Alternativamente, los datos del ensayo podrían usarse para determinar qué variables tienen la capacidad de predicción. Esta estrategia también podría especificarse previamente en el protocolo del estudio. Debido a que este análisis se realizaría condicionalmente sobre el imbalance observado, no elimina el sesgo y, por lo tanto, no puede considerarse completamente satisfactorio”¹⁰.
- Tukey (1991) señala que los imbalances que observamos nos pueden servir para justificar el ajuste de covariables como una forma para verificar qué tan robusto son nuestros resultados: aunque “la mayoría de los estadísticos” aceptarían un análisis de un ensayo clínico aleatorio que no se ajusta a las covariables, “algunos médicos, y algunos estadísticos serían más escépticos y (tal vez como análisis complementario) pedirían un análisis que tenga en cuenta los imbalances observados en estas covariables registradas. Tener mayor seguridad sobre los resultados de tal análisis es ideal, ya que así nos protegemos contra las consecuencias de una aleatorización inadecuada o la ocurrencia (aleatoria) de una aleatorización inusual, las cuales aumentan considerablemente con el ajuste. *Mayor seguridad, en lugar de mayor precisión . . . a menudo será la razón básica para el ajuste de covarianza en un ensayo aleatorizado. El propósito principal de permitir [ajustar] por covariables en un ensayo aleatorizado es defensivo: dejar en claro que el análisis ha cumplido con su obligación científica*”¹¹.
- Algunos estadísticos afirman que nuestras inferencias deben estar condicionadas a una medida de imbalance de covariables, en otras palabras, al evaluar el sesgo, la varianza y el error cuadrático medio de una estimación puntual o la probabilidad de cobertura de un intervalo de confianza. Entonces, en lugar de considerar todas las aleatorizaciones posibles, puede ser más relevante considerar solo aquellas aleatorizaciones que producirían un imbalance de covariables similar al que observamos. Desde esta perspectiva, los imbalances observados pueden ser relevantes para la elección del estimador¹².
- Lin, Green y Coppock (2016) aseguran que “por lo general, las covariables deben elegirse sobre la base

⁹Thomas Permutt (1990), “Testing for Imbalance of Covariates in Controlled Experiments,” *Statistics in Medicine* 9: 1455–1462.

¹⁰Douglas G. Altman (2005), [“Covariate Imbalance, Adjustment for,”] (<http://doi.org/10.1002/0470011815.b2a01015>) en *Encyclopedia of Biostatistics*.

¹¹John W. Tukey (1991), “Use of Many Covariates in Clinical Trials,” *International Statistical Review* 59: 123–137. Cursiva en el título original.

¹²See, e.g.: D. R. Cox and N. Reid (2000), *The Theory of the Design of Experiments*, pp. 29–32; D. Holt y T. M. F. Smith (1979), “Post Stratification,” *Journal of the Royal Statistical Society, Series A (General)* 142: 33–46; Richard M. Royall (1976), “Current Advances in Sampling Theory: Implications for Human Observational Studies,” *American Journal of Epidemiology* 104: 463–474. Para una introducción a los desacuerdos filosóficos sobre la inferencia estadística, puede ver Bradley Efron (1978), “Controversies in the Foundations of Statistics,” *American Mathematical Monthly* 85: 231–246.

de su capacidad esperada para ayudar a predecir los resultados, independientemente de si parecen estar bien balanceadas entre los brazos de tratamiento o no. Pero puede haber ocasiones en las que la lista de covariables especificada en el PAP [plan de análisis previo] omitió una covariable potencialmente importante (debido a un descuido o la necesidad de mantener la lista corta cuando N es pequeño) con un imbalance no trivial. La protección contra el sesgo ex post (condicionado al imbalance observado) es entonces una preocupación legítima.”Sin embargo, recomiendan que si se permite que los desequilibrios observados influyan en la elección de las covariables”, las comprobaciones de equilibrio y las decisiones sobre el ajuste deben finalizar antes de que veamos datos de resultado no cegados”, “la *dirección* del desequilibrio observado (por ejemplo, si el grupo de tratamiento o el grupo de control parece más aventajado en la fase inicial del estudio) no debe poder influir en las decisiones sobre el ajuste” y el estimador originalmente pre-especificado deben “ser siempre reportados y etiquetados como tales, incluso si también se reportan estimaciones alternativas.

7 Cuándo no debemos hacer ajuste de covariables

No es una buena idea ajustar las covariables cuando creemos que nuestro tratamiento podría haber influido en esas covariables. Ésta es una de las razones por las que se recopilan muchas covariables de las encuestas de referencia inicial; a veces, las covariables que se recopilan en las encuestas después de la intervención podrían reflejar los efectos del tratamiento en lugar de las características subyacentes del sujeto. El ajuste de las covariables que se ven afectadas por el tratamiento (covariables “posteriores al tratamiento”) puede causar sesgo.

Supongamos, por ejemplo, que Giné y Mansuri hubieran recopilado datos sobre el número de mítines políticos al que asistieron las mujeres después de recibir el tratamiento. Al estimar el efecto del tratamiento sobre la independencia de la elección política, puede tener la tentación de incluir esta variable como una covariable en su regresión. Pero incluir esta variable, incluso si predice fuertemente el resultado, puede distorsionar el efecto estimado del tratamiento.

En el código presentado a continuación, creamos esta variable falsa, que está correlacionada (como la variable de resultado) con las covariables de referencia y también con el tratamiento. Aquí, por construcción, el efecto del tratamiento sobre el número de mítines políticos a los que asistió es 2. Cuando incluimos la variable de mítines como una covariable, el efecto del tratamiento promedio estimado sobre la independencia de la elección del candidato promedió 0.54 en las 10,000 repeticiones. Recuerde que el verdadero efecto del tratamiento en este resultado es 1. Se trata de un sesgo severo, ¡todo porque controlamos una covariable posterior al tratamiento! ¹³ Este sesgo se debe al hecho de que la covariable está correlacionada con el tratamiento.

```
# Crea una covariable posterior al tratamiento
# que se correlacione con las covariables previas al tratamiento
rallies0 = round(.5*owns.id.card + has.formal.schooling + 1.5*TV.access + log(age))
rallies1 = rallies0 + 2
rallies.mat = rallies1 * Z.mat + rallies0 * (1-Z.mat)

# Estima el ATE con nuevo modelo que incluye la covariable postratamiento

adjust.for.post = function(Y, Z, X) {
  coef(lm(Y ~ Z + X + owns.id.card + has.formal.schooling + age + TV.access))[2]
}

post.adjusted.estimates = rep(NA, Replications)

for (i in 1:Replications) {
  post.adjusted.estimates[i] = adjust.for.post(Y.mat[,i], Z.mat[,i], rallies.mat[,i])
}
```

¹³El sesgo estimado es \$-\$0.459 con un margen de error (en el nivel de confianza del 95%) de 0,002.

```

}

# Sesgo estimado del nuevo estimador
mean(post.adjusted.estimates) - true.treatment.effect

# Margen de error (a un nivel de confianza del 95%) para el sesgo estimado
1.96 * sd(post.adjusted.estimates) / sqrt(Replications)

```

El hecho de que no deba ajustar las covariables que son afectadas por el tratamiento no significa que no pueda recopilar datos de las covariables después del tratamiento, pero debe tener cuidado. Algunas medidas podrían recopilarse después del tratamiento, pero es poco probable que se vean afectadas por el tratamiento (por ejemplo, edad y sexo). Sin embargo, tenga cuidado con las medidas que pueden estar sujetas a efectos impulsados por la evaluación: por ejemplo, las mujeres tratadas pueden ser más conscientes de la expectativa de participación política y pueden informar retrospectivamente que fueron más activas políticamente de lo que realmente eran varios años antes.

8 Preocupaciones sobre el sesgo en muestras pequeñas

En muestras pequeñas, el ajuste de regresión puede producir una estimación sesgada del efecto promedio del tratamiento¹⁴. El uso de simulaciones nos ha permitido ver que este sesgo tiende a ser insignificante cuando el número de unidades asignadas al azar es mayor que veinte¹⁵. Si está trabajando con una muestra pequeña, es posible que desee utilizar un método de ajuste de covariables insesgado como post-estratificación (dividir la muestra en subgrupos en función de los valores de una o más covariables iniciales, calcular la diferencia de control-tratamiento en los resultados medios para cada subgrupo y tomar un promedio ponderado de estas estimaciones del efecto del tratamiento específico de subgrupo, con ponderaciones proporcionales al tamaño de la muestra)¹⁶.

9 Cómo hacer que las decisiones sobre el ajuste de covariables sean transparentes

En aras de la transparencia, si ajusta las covariables, pre-especifique sus modelos e informe estimaciones no ajustadas y ajustadas por covariables.

Las simulaciones anteriores han demostrado que los resultados pueden cambiar levemente o no tan levemente según las covariables que elija incluir en su modelo. Hemos resaltado algunas reglas generales aquí: incluya solo las covariables previas al tratamiento que sean predictivas de los resultados. Sin embargo, decidir qué covariables incluir es a menudo una tarea subjetiva más que objetiva, por lo que otra regla general es ser totalmente transparente sobre las decisiones de covariables. Incluya siempre el modelo más simple, la regresión simple del resultado del tratamiento sin controlar las covariables, en su artículo o apéndice para complementar los hallazgos de su modelo, incluidas las covariables.

Otra forma de minimizar la preocupación de sus lectores para que no crean que usted eligió una combinación particular de covariables que dio resultados favorables a sus hipótesis es pre-especificar sus modelos en un plan de pre-análisis¹⁷. Esto le da la oportunidad de explicar antes de ver los resultados qué covariables medidas antes del tratamiento espera que sean predictivas del resultado. Incluso puede escribir estas regresiones en R usando datos falsos, como hacemos aquí, de modo que cuando lleguen los resultados del campo, todo lo que

¹⁴David A. Freedman (2008), “On Regression Adjustments in Experiments with Several Treatments,” *Annals of Applied Statistics* 2: 176–196. También puede ver las publicaciones del blo de Winston Lin (parte I y parte II) sobre su respuesta a Freedman.

¹⁵Green, Donald P. and Aronow, Peter M., Analyzing Experimental Data Using Regression: When Is Bias a Practical Concern? (March 7, 2011). Working paper: <http://ssrn.com/abstract=1466886>

¹⁶Miratrix, Sekhon, and Yu (2013), citados anteriormente.

¹⁷Para más información sobre los planes de pre-análisis, ver, por ejemplo, Benjamin A. Olken (2015), [“Promises and Perils of Pre-Analysis Plans”] (<http://doi.org/10.1257/jep.29.3.61>) *Journal of Economic Perspectives* 29 (3) : 61–80.

necesita hacer es ejecutar su código con los datos reales. Estos esfuerzos son una forma útil “de atar sus manos” como investigador y mejorar su credibilidad.

10 Las covariables pueden ayudarlo a investigar la integridad de la asignación aleatoria

A veces puede que no nos quede claro si las unidades fueron realmente asignadas aleatoria (o si se utilizó el procedimiento que el investigador previó). Por ejemplo, cuando los académicos analizan asignaciones aleatorias que ocurren naturalmente (por ejemplo, las realizadas por una agencia gubernamental), es útil evaluar estadísticamente si el grado de imbalance entre los grupos de tratamiento y control está dentro del margen de error esperado. Una prueba estadística consiste en hacer una regresión de la asignación de tratamiento en todas las covariables y calcular el estadístico F. La importancia de esta estadística se puede evaluar simulando un gran número de asignaciones aleatorias y para cada una calculando la estadística F; la distribución resultante se puede utilizar para calcular el valor p del estadístico F observado. Por ejemplo, si se realizan 10,000 simulaciones y solo 30 simulaciones generan una estadística F mayor que la que se obtuvo realmente a partir de los datos, el valor p es 0.003, lo que sugiere que el nivel de imbalance observado es muy inusual. En tales casos, es posible que desee investigar el procedimiento de aleatorización más de cerca.

Lecturas complementarias

Athey, Susan y Guido W. Imbens (2017). “The Econometrics of Randomized Experiments.” In *Handbook of Economic Field Experiments*, vol. 1 (E. Duflo and A. Banerjee, eds.). arXiv DOI

Gerber, Alan S. y Donald P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation*, chapter 4.

Hennessy, Jonathan, Tirthankar Dasgupta, Luke Miratrix, Cassandra Pattanayak y Pradipta Sarkar (2016). “A Conditional Randomization Test to Account for Covariate Imbalance in Randomized Experiments.” *Journal of Causal Inference* 4: 61–80.

Judkins, David R. y Kristin E. Porter (2016). “Robustness of Ordinary Least Squares in Randomized Clinical Trials.” *Statistics in Medicine* 35: 1763–1773.

Lin, Winston (2012). “Regression Adjustment in Randomized Experiments: Is the Cure Really Worse than the Disease?” Development Impact blog post, part I and part II.

Raudenbush, Stephen W. (1997). “Statistical Analysis and Optimal Design for Cluster Randomized Trials.” *Psychological Methods* 2: 173–185.

Wager, Stefan, Wenfei Du, Jonathan Taylor y Robert Tibshirani (2016). “High-Dimensional Regression Adjustments in Randomized Experiments.” *Proceedings of the National Academy of Sciences* 113: 12673–12678. arXiv DOI