

Medida de desequilibrio multivariante:  $L_1 = 0,735$

Porcentaje de apoyo común local: LCS = 12,4%

Univariante Medidas de desequilibrio:

	estadística	tipo	L1 min 25%	
edad	0,179203803 (diferencia)	4,705882e-03	0	1
educación	0,192236086 (diferencia)	9,811844e-02	1	0
negro	0.001346801 (diferencia)	1.346801e-03	0	0
casado	0.010703110 (diferencia)	1.070311e-02	0	0
no graduado	- 0.083477916 (diferencia)	8.347792e-02	0	- 1
re74	- 101.486184085 (diferencia)	5.551115e-17	0	0
re75	39.415450601 (diferencia)	5.551115e-17	0	0
Hispano	- 0.018665082 (diferencia)	1.866508e-02	0	0
u74	- 0.020099030 (diferencia)	2.009903e-02	0	0
u75	- 0.045086156 (diff)	4.508616e-02	50% 75% máx.	0 0
edad	0,00000 -1,0000 - 6,0000			
educación	1,00000 1,0000 2,0000			
negro	0,00000 0,0000 0,0000			
casado	0,00000 0,0000 0,0000			
no graduado	0,00000 0,0000 0,0000			
re74	69.73096 584.9160 -2139.0195			
re75	294.18457 660.6865 490.3945			
Hispano	0,00000 0,0000 0,0000			
u74	0,00000 0,0000 0,0000			
u75	0,00000 0,0000 0,0000			

La primera línea de nuestros informes de salida  $L_1$ , que es una medida del desequilibrio multivariado creado por Iacus et al. (2011). Una explicación más completa está disponible en ese artículo, pero en general, esta estadística varía de 0 a 1, y los valores más bajos indican una mejor equilibrio. Cuando  $L_1 = 0$  las dos distribuciones se superponen perfectamente, y cuando  $L_1 = 1$  las dos distribuciones no se superponen en absoluto. Volviendo a la mesa, cada fila muestra varias estadísticas de equilibrio para una covariable individual. Para todas estas estadísticas, los valores más cercanos a cero son mejores. La columna etiquetada estadística muestra la diferencia en significa entre las variables y la columna etiquetada L1 calcula  $L_1$ . Cuál es el misma medida que  $L_1$  pero solo calculado para la covariable individual. Las columnas restantes muestran diferencias de cuantiles entre los dos grupos (p. Ej., La diferencia en los mínimos respectivos de los dos grupos, la diferencia entre los percentiles 25 respectivos de los grupos, etc.).

A continuación, usaremos el `cem` comando para realizar `Cen` `furecido` `mixact` `METRO` capturando nuestros datos. Dentro `decem` comando, enumeramos nuestra variable de tratamiento con el tratamiento argumento, nuestro conjunto de datos con el `datos` argumento, y cualquier variable que no queramos que coincida con el `soltar` argumento. El argumento siempre debe incluir nuestra variable de resultado, si está en el mismo conjunto de datos, así como cualquier dato

índices o variables irrelevantes. Podríamos usar un vector para enumerar todas las variables que queremos que se ignoren, como hicimos con el `desequilibrio` comando antes, pero en este caso, solo el resultado `re78` debe omitirse. Escribimos:

```
cem.match.1 <- cem (tratamiento = "tratado", datos = LL, gota = "re78") cem.match.1
```

Nuestro resultado inmediato de esto es simplemente el siguiente:

	G0	G1
Todas	425	297
Emparejado	222	163
Inigualable	203	134

Esto nos dice que nuestros datos originales tenían 425 observaciones de control y 297 observaciones tratadas. El CEM incluyó 222 del control y 163 de las observaciones tratadas en la muestra emparejada, y el resto se eliminó. Para ser claros: todas las observaciones todavía están contenidas en el original `LL` conjunto de datos, pero ahora el objeto `cem.match.1` detalla qué observaciones coinciden o no.

Dado que CEM procede agrupando valores similares de covariables en estratos, una característica importante de esto es cómo establecemos los intervalos ordenados de cada predictor en el engrosamiento. `Lacem` El comando tiene valores predeterminados razonables si el usuario no establece estos intervalos, pero es importante registrar cuáles son los intervalos. Para ver cuáles fueron nuestros intervalos para los valores de nuestros predictores, podríamos escribir: `cem.match.1 $ descansos`.

Esto nos daría el siguiente resultado:

```
$ edad
[1] 17,0 20,8 24,6 28,4 32,2 36,0 39,8 43,6 47,4 51,2 55,0

$ educación
[1] 3,0 4,3      5,6   6,9   8.2   9,5 10,8 12,1 13,4 14,7 16,0

$ negro
[1] 0,0 0,1 0.      2 0,3 0,4 0,5 0,6 0,7 0,8 0,9 1,0

$ casado
[1] 0,0 0,1 0.      2 0,3 0,4 0,5 0,6 0,7 0,8 0,9 1,0

$ nodegree
[1] 0,0 0,1 0.      2 0,3 0,4 0,5 0,6 0,7 0,8 0,9 1,0

$ re74
[1]      0.000    3957.068 7914.136 11871.204 15828.272 19785.340
[7] 23742.408 27699.476 31656.544 35613.612 39570.680

$ re75
[1]      0.000 3743.166 7486.332 11229.498 14972.664 18715.830
[7] 22458.996 26202.162 29945.328 33688.494 37431.660

$ hispano
[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

\$ u74

[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

\$ u75

[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Para ilustrar lo que esto significa, considere edad. La categoría más baja de edad tosca agrupa a todas las personas de 17 a 20,8 años, la segunda categoría agrupa a todas las personas de 20,8 a 24,6 años, y así sucesivamente. Variables comocasado, nodegree, hispano, negro, u74, y u75 son en realidad binarios, por lo que la mayoría de las categorías que se crean son innecesarias, aunque los contenedores vacíos no dañarán nuestro análisis. Por supuesto, los usuarios no están obligados a utilizar los valores predeterminados de software, y Iacus et al. instar a los investigadores a utilizar un conocimiento sustancial de la medición de cada variable para establecer los rangos de los contenedores de engrosamiento (2012, pag. 9). Sección 8.3.2 ofrece detalles sobre cómo hacer esto.

Ahora podemos evaluar el desequilibrio en la nueva muestra emparejada escribiendo:

```
desequilibrio (LL $ tratado [cem.match.1 $ igualado],
              LL [cem.match.1 $ coincidente,], drop = todrop)
```

Nuestro resultado de esto es el siguiente:

Medida de desequilibrio multivariante: L1 = 0,592 Porcentaje de apoyo local común: LCS = 25,2%

Medidas de desequilibrio univariante:

	estadística	tipo	L1 min 25%		50%
edad	- 0,42486044 (diferencia)	0,00000000	0	- 1	- 2,0000
educación	- 0,10855027 (diferencia)	0,10902006	0	0	- 1,0000
negro	- 0,01771403 (diferencia)	0,01771403	0	0	0,0000
casado	- 0,01630465 (diferencia)	0,01630465	0	0	0,0000
no graduado	0,09022827 (diferencia)	0,09022827	0	0	0,0000
re74	- 119,33548135 (diferencia)	0,00000000	0	0	0,0000
re75	- 50,01527694 (diferencia)	0,00000000	0	0	- 49,3559
Hispano	0,01561377 (diferencia)	0,01561377	0	0	0,0000
u74	0,01619411 (diferencia)	0,01619411	0	0	0,0000
u75	0,02310286 (diferencia)	0,02310286	0	0	0,0000
	75%	max			
edad	0,00	1,000			
educación	0,00	0,000			
negro	0,00	0,000			
casado	0,00	0,000			
no graduado	0,00	0,000			
re74	- 492,95	416,416			
re75	- 136,45	- 852,252			
Hispano	0,00	0,000			
u74	0,00	0,000			
u75	0,00	0,000			

Comparar esto a los datos originales. Ahora tenemos  $L_1 D_0 = 592$ , que es menor que nuestra puntuación de 0,735 para los datos brutos, lo que indica que el equilibrio multivariado es mejor en el emparejado muestra. En cuanto a las covariables individuales, puede ver algo de una bolsa mixta, pero en general el equilibrio parece mejor. Por ejemplo, conedad la

La diferencia de medias es en realidad un poco mayor en valor absoluto con la muestra emparejada (0,42) que con los datos brutos (0,18). Sin embargo,  $L_{\text{eda}}$  ahora es minúsculo en la muestra emparejada y menor que el valor 0.0047 para los datos sin procesar. Esto probablemente se deba al hecho de que los datos brutos tienen una mayor discrepancia en el extremo superior que la muestra emparejada. El usuario ahora debe decidir si los casos tratados y de control están lo suficientemente equilibrados o si probar otros engrosamientos para mejorar el equilibrio.

Si el usuario está satisfecho con el nivel de equilibrio, puede proceder a estimar el **A** media **T**efecto de regeneración en el **T**reaccionó (ATT) usando el comando `att`. Esta cantidad representa el efecto causal sobre el tipo de individuo que recibió el tratamiento. En este comando especificamos qué muestra coincidente está usando `elobj` argumento, la variable de resultado (**re78**) y `tratamiento` (**tratado**) utilizando la fórmula argumento, y nuestros datos usando el `datos` argumento. Esto nos da:

```
est.at.1 <- att (obj = cem.match.1, fórmula = re78 ~ tratado, datos = LL) est.att.1
```

Nuestro resultado de esto es:

	G0	G1
Todas	425	297
Emparejado	222	163
Inigualable	203	134

Modelo de regresión lineal sobre datos emparejados CEM:

Estimación puntual de SATT: 550,962564 (p.value = 0,368242) 95% conf.  
intervalo: [-647.777701, 1749.702830]

El resultado resume las características de nuestra muestra emparejada y luego informa nuestra estimación del efecto promedio del tratamiento de la muestra en los tratados (SATT): Estimamos en nuestra muestra que las personas que recibieron el tratamiento ganaron \$ 551 más en promedio que aquellas que no recibieron el tratamiento. Sin embargo, este efecto no es estadísticamente discernible de cero (*pag* D 0: 368). Ésta es una conclusión marcadamente diferente de la que extrajimos en el cap. 5, cuando observamos una diferencia de \$ 886 que fue estadísticamente significativa. Esto ilustra la importancia del control estadístico.

### 8.3.2 Explorando diferentes soluciones CEM

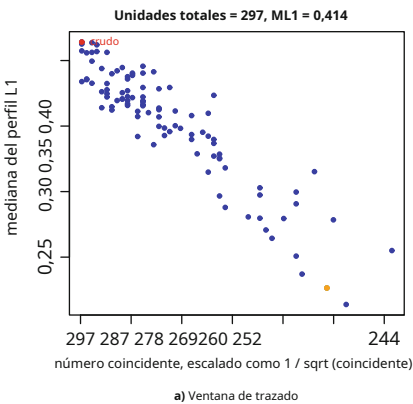
Como punto final, si un investigador no está satisfecho con el nivel de equilibrio o el tamaño de la muestra en la muestra emparejada, entonces una herramienta para encontrar un mejor equilibrio es la `cemspace` comando. Este comando produce aleatoriamente varios grosores diferentes para las variables de control (250 diferentes grosores por defecto). Luego, el comando traza el nivel de equilibrio contra el número de observaciones tratadas incluidas en la muestra emparejada. El siguiente código llama a este comando:

```
cem.explore <- cemspace (tratamiento = "tratado", datos = LL, gota = "re78")
```

La sintaxis de cemspace es parecido a cem, aunque dos opciones más son importantes: mínimo y máximo. Estos establecen cuál es el número mínimo y máximo permitido de intervalos aproximados para las variables. El comando anterior usa los valores predeterminados de 1 y 5, lo que significa que no se pueden incluir más de cinco intervalos para una variable. Por lo tanto, todas las muestras coincidentes de este comando serán más burdas que las que usamos en la sección.8.3.1, y por lo tanto menos equilibrado. Sin embargo, el usuario podría aumentarmáximo a 12 o incluso a un número más alto para crear intervalos más finos y potencialmente mejorar el equilibrio sobre nuestro resultado anterior.

Nuestra salida de cemspace se muestra en la Fig. 8.2. *Debido al elemento aleatorio en la elección de grosores, sus resultados no coincidirán exactamente con esta cifra.* Figura 8.2a muestra la figura interactiva que se abre. El eje horizontal de esta figura muestra el número de unidades de tratamiento emparejadas en orden descendente, mientras que el eje vertical muestra el nivel de desequilibrio. En general, una muestra emparejada en la esquina inferior izquierda del gráfico sería ideal, ya que indicaría el mejor equilibrio (reducción del sesgo) y la muestra más grande (aumento de la eficiencia). Normalmente, sin embargo, tenemos que tomar una decisión sobre esta compensación, generalmente poniendo un poco más de peso en minimizar el desequilibrio. Al hacer clic en diferentes puntos del gráfico, la segunda ventana quecemspace crea, que se muestra en la Fig. 8.2b mostrará los intervalos utilizados en ese engrosamiento particular. El usuario puede copiar los vectores de los puntos de corte del intervalo y pegarlos en su propio código. *Nota:* R no continuará con nuevos comandos hasta que estas dos ventanas estén cerradas.

En la Fig. 8.2a, se ha elegido uno de los posibles grosores, y es iluminado y amarillo. Si queremos implementar este engrosamiento, podemos copiar los vectores que se muestran en la segunda ventana ilustrada en la Fig.8.2B. Pegando estos en el nuestroR script produce el siguiente código:



a) Ventana de trazado

Matched units=253, ML1=0.226
Variable: age
c(17, 26.5, 36, 45.5, 55)
Variable: education
c(3, 6.25, 9.5, 12.75, 16)
Variable: black
c(0, 0.5, 1)
Variable: married
c(0, 0.5, 1)
Variable: nodegree
c(0, 0.5, 1)
Variable: re74
c(0, 19785.34, 39570.68)
Variable: re75
c(0, 9357.92, 18715.83, 28073.75, 37431.66)
Variable: hispanic
c(0, 0.5, 1)
Variable: u74
c(0, 0.5, 1)
Variable: u75
c(0, 1)
Additional CEM args:
Run CEM with these coarsening

B) Ventana X11

**Figura 8.2** Gráfica de las estadísticas de equilibrio para 250 muestras emparejadas de groserías aleatorias contra el número de observaciones tratadas incluidas en la muestra emparejada respectiva. **(a)** Ventana de trazado; **(B)** Ventana X11

```
age.cut <-c (17, 26.5, 36, 45.5, 55) education.cut <-c (3, 6.25,
9.5, 12.75, 16) black.cut <-c (0, 0.5, 1)

casado.cut <-c (0, 0.5, 1) nodogree.cut <-c (0, 0.5,
1) re74.cut <-c (0, 19785.34, 39570.68)

re75.cut <-c (0, 9357.92, 18715.83, 28073.75, 37431.66) hispanic.cut <-c (0,
0.5, 1)
u74.cut <-c (0, 0.5, 1)
u75.cut <-c (0, 1)
new.cuts <-list (age = age.cut, education = education.cut,
black = black.cut, casado = casado.cut, nodegree = nodegree.cut, re74 =
re74.cut, re75 = re75.cut, hispanic = hispanic.cut, u74 = u74.cut, u75 =
u75.cut)
```

Terminamos este código creando un *lista* de todos estos vectores. Si bien nuestros vectores aquí se han creado utilizando un engrosamiento creado por `cemspace`, este es el procedimiento que usaría un programador para crear sus propios puntos de corte para los intervalos. Sustituyendo los vectores anteriores con vectores de punto de corte creados por el usuario, un investigador puede utilizar su propio conocimiento de la medición de las variables para engrosar.

Una vez que hemos definido nuestros propios puntos de corte, ya sea usando `cemspace` o conocimientos sustantivos, ahora podemos aplicar CEM con el siguiente código:

```
cem.match.2 <- cem (tratamiento = "tratado", datos = LL, gota = "re78",
puntos de corte = new.cuts)
```

Nuestra adición clave aquí es el uso de la puntos de corte opción, donde ingresamos nuestra lista de intervalos. Al igual que en la Secta.8.3.1, ahora podemos evaluar las cualidades de la muestra emparejada, los niveles de desequilibrio y calcular el ATT si lo deseamos:

```
cem.match.2
desequilibrio (LL $ tratado [cem.match.2 $ igualado],
LL [cem.match.2 $ coincidente,], drop = todrop)
est.at.2 <- att (obj = cem.match.2, fórmula = re78 ~ tratado, datos = LL) est.att.2
```

En este caso, en parte debido a los bins más gruesos que estamos usando, el saldo es peor que el que encontramos en la sección anterior. Por lo tanto, estaríamos mejor en este caso si nos quedamos con nuestro primer resultado. Se anima al lector a tratar de encontrar un engrosamiento que produzca un mejor equilibrio que los valores predeterminados del software.

## 8.4 Análisis de lista legislativa con dominar

Los metodólogos en ciencias políticas y otras disciplinas han desarrollado una amplia gama de modelos de medición, varios de los cuales están disponibles para que los usuarios los implementen en R. Sin duda, uno de los modelos de medición más destacados en la disciplina es NOMINATE, abreviatura de **nominal** **t**tres **p**asos **m**iestimulación (McCarty et al. 1997; Poole y Rosenthal 1997; Poole y col. 2011). El modelo NOMINATE analiza los datos nominales de las votaciones legislativas, ubicando a los legisladores y las alternativas políticas que votan en el espacio ideológico. El modelo es particularmente prominente porque Poole, Rosenthal y sus colegas hacen que las puntuaciones DW-NOMINATE estén disponibles

para la Cámara de Representantes y el Senado de los Estados Unidos para cada período del Congreso. Innumerables autores han utilizado estos datos, interpretando típicamente la puntuación de la primera dimensión como una escala de ideología liberal-conservadora.

En resumen, la lógica básica del modelo se basa en el modelo de proximidad espacial de la política, que esencialmente establece que tanto las preferencias ideológicas de los individuos como las alternativas políticas disponibles pueden representarse en un espacio geométrico de una o más dimensiones. Un individuo generalmente votará por la opción de política que más se acerque en el espacio a su propio punto ideal ideológico (Black<sup>1958</sup>; Downs<sup>1957</sup>; Hotelling<sup>1929</sup>). El modelo NOMINATE se basa en estos supuestos y coloca a los legisladores y las opciones de políticas en el espacio ideológico según la forma en que los votos de los legisladores se dividen en el transcurso de muchas votaciones nominales y cuando los legisladores se comportan de manera impredecible (produciendo errores en el modelo). Por ejemplo, en el Congreso de los Estados Unidos, los miembros liberales suelen votar de manera diferente a los miembros conservadores, y los ideólogos extremos son los más propensos a estar en una pequeña minoría siempre que haya un amplio consenso sobre un tema. Antes de aplicar el método NOMINATE a sus propios datos, e incluso antes de descargar datos DW-NOMINATE premedidos para incluirlos en un modelo que calcule, asegúrese de leer más sobre el método y sus suposiciones porque comprender a fondo cómo funciona un método es esencial antes usándolo. En particular, el Cap. 2 y Apéndice A de Poole y Rosenthal (<sup>1997</sup>) y el Apéndice A de McCarty et al. (<sup>1997</sup>) ofrecen descripciones detalladas, pero intuitivas, de cómo funciona el método.

En R, la dominar El paquete implementa W-NOMINATE, que es una versión del algoritmo NOMINATE que solo debe aplicarse a una sola legislatura. Los puntajes de W-NOMINATE son válidos internamente, por lo que es justo comparar los puntajes de los legisladores dentro de un solo conjunto de datos. Sin embargo, los puntajes no se pueden comparar externamente con los puntajes cuando W-NOMINATE se aplica a un período diferente de la legislatura oa un cuerpo de actores completamente diferente. Por tanto, es un buen método para intentar hacer comparaciones transversales entre legisladores de un mismo organismo.

Si bien la aplicación más común para W-NOMINATE ha sido el Congreso de los Estados Unidos, el método podría aplicarse a cualquier cuerpo legislativo. Con ese fin, el ejemplo práctico de esta sección se centra en las votaciones nominales emitidas en las Naciones Unidas. Esto Naciones Unidas el conjunto de datos está disponible en el dominar paquete, y reúne 237 votos nominales emitidos en las primeras tres sesiones de la ONU (1946-1949) por 59 países miembros. Las variables están etiquetadas **V1** a **V239**. **V1** es el nombre de la nación miembro, y **V2** es una variable categórica codificada como "WP" para un miembro del Pacto de Varsovia, u "Otro" para todas las demás naciones. Las variables restantes identifican secuencialmente las votaciones nominales.

Para empezar, limpiamos, instalamos el dominar paquete la primera vez que lo usamos, cargamos la biblioteca y cargamos el Naciones Unidas datos:<sup>9</sup>

```
rm(list = ls())
install.packages("wnominate") biblioteca
(wnominate)
datos (ONU)
```

<sup>9</sup>Los datos de la ONU también están disponibles en el archivo UN.csv, disponible en el Dataverse (consulte la página vii) o el contenido del capítulo (consulte la página 125)

Una vez que se cargan los datos, se pueden ver con los comandos estándar como `reparar`, pero para una vista rápida de cómo se ven los datos, simplemente podríamos escribir: `cabeza (ONU [, 1: 15])`. Esto mostrará la estructura de los datos a través de las primeras 13 votaciones nominales.

Antes de que podamos aplicar `W-NOMINATE`, tenemos que reformatear los datos a un objeto de clase Llamada de rol. Para hacer esto, primero necesitamos redefinir nuestra Naciones Unidas conjunto de datos como una matriz, y dividir los nombres de los países, si el país estaba en el Pacto de Varsovia, y el conjunto de llamadas de lista en tres partes separadas:

```
UN <-as.matrix (UN)
UN.2 <-UN [, - c (1,2)]
UNnames <-UN [, 1]
legData <-matrix (UN [, 2], length (UN [, 2]), 1)
```

La primera línea giró el Naciones Unidas marco de datos en una matriz. (Para obtener más información sobre los comandos de matriz enR, ver Cap. 10.) La segunda línea creó una nueva matriz, que hemos denominado `UN.2`, que ha eliminado las dos primeras columnas (nombre del país y miembro del Pacto de Varsovia) para dejar solo las llamadas de lista. La tercera línea exportó los nombres de las naciones al vector `UNnames`. (En muchos otros entornos, este sería el nombre del legislador o una variable de identificación.) Por último, nuestra variable de si una nación estaba en el Pacto de Varsovia se ha guardado como una matriz de una columna llamada `legData`. (En muchos otros entornos, este sería el partido político de un legislador.) Una vez que tengamos estos componentes juntos, podemos usar el Llamada de rol comando para de fi nir un Llamada de rol-objeto de clase que nombramos `rc` como sigue:

```
rc <-rollcall (data = UN.2, yea = c (1,2,3), nay = c (4,5,6),
  missing = c (7,8,9), notInLegis = 0, legis.names = UNnames, legis.data = legData,
  desc = "UN Votes", source = "voteview.com")
```

Especificamos nuestra matriz de votaciones nominales con el `data` argumento. Basado en cómo se codifican los datos en la matriz de votación nominal, usamos `yea`, `nay`, y `missing` argumentos para traducir códigos numéricos a su significado sustantivo. Adicionalmente, `notInLegis` nos permite especificar un código que específicamente significa que el legislador no era miembro en el momento de la votación (por ejemplo, un legislador falleció o dimitió). No tenemos tal caso en estos datos, pero el valor predeterminado `notInLegis = 9`, y `9` significa algo más para nosotros, por lo que debemos especificar un código no utilizado de `0`. Con `legis.names` especificamos los nombres de los votantes, y con `legis.data` especificamos variables adicionales sobre nuestros votantes. Finalmente, `desc` y `source` nos permite registrar información adicional sobre nuestros datos.

Con nuestros datos formateados correctamente, ahora podemos aplicar el modelo `W-NOMINATE`. el comando simplemente se llama `wnominate`:

```
resultado <-wnominate (rcObject = rc, polarity = c (1,1))
```

Con el `rcObject` argumento, simplemente nombramos nuestros datos formateados correctamente. La polaridad El argumento, por el contrario, requiere una aportación sustancial del investigador: el usuario debe especificar un vector que enumere qué observación debe caer claramente en el lado positivo de cada dimensión estimada. Dada la política del Frío



Guerra, usamos la observación n.º 1, los EE. UU., Como ancla en ambas dimensiones que estimamos. Por defecto, dominar coloca a los votantes en ideología bidimensional espacio (aunque esto podría cambiarse especificando el atenua opción).

Para ver los resultados de nuestra estimación, `lata` comienzo `por` mecanografía: `resumen (resultado)`. Esto imprime la siguiente salida.

RESUMEN DEL OBJETO NOMINADO W

-----

Número de legisladores:	59 (0 legisladores	eliminado)
Número de votos: 219	(18 votos eliminados)	
Número de dimensiones:	2	
Sí pronosticado:	4693	de 5039 (93,1%) predicciones correcto
No previstos:	4125	de 4488 (91,9%) predicciones correcto
Clasificación correcta: APRE: 0.574	89,5%	92,56%
0.698		
BPF: 0,783	0,841	

Las primeras 10 estimaciones de los legisladores son:

	coord1D	coord2D
Estados Unidos	0,939	0.344
Canadá	0,932	0.362
Cuba	0.520	- 0.385
Haití	0.362	- 0.131
Rep.Dominicana	0,796	- 0,223
México	0,459	0,027
Guatemala	0.382	0.364
Honduras	0.588	- 0,266
El Salvador	0,888	- 0,460
Nicaragua	0,876	- 0,301

La salida comienza recapitulando varias características descriptivas de nuestros datos y luego se ajusta a los índices. Enumera, por ejemplo, la predicción correcta de sí y no, y luego, en "Clasificación correcta", enumera el porcentaje predicho correctamente por la primera dimensión sola y luego ambas dimensiones juntas. Con un 89,5%, la primera dimensión puede explicar mucho por sí sola. El resultado finaliza enumerando las estimaciones de nuestras primeras 10 observaciones. Como puede verse, EE. UU. Tiene un valor positivo en ambas dimensiones, según nuestra especificación en el modelo.

Podemos obtener resultados adicionales escribiendo: `trama (resultado)`. El resultado de este comando se presenta en la Fig. 8.3. El panel superior izquierdo visualiza los puntajes W-NOMINATE para las 59 naciones votantes. El eje horizontal es la primera dimensión de la puntuación, el eje vertical es la segunda dimensión y cada punto representa la posición de una nación. Por lo tanto, en el espacio ideológico bidimensional definido internamente por la ONU

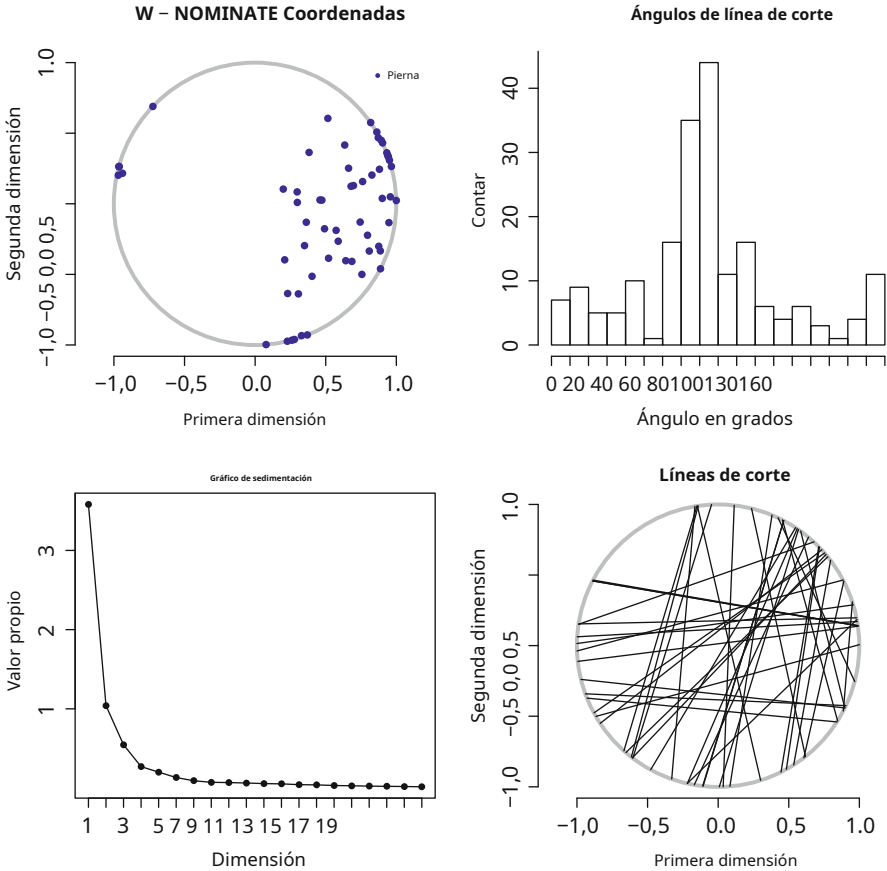


Figura 8.3 Producción gráfico a partir de la estimación de las puntuaciones W-NOMINATE de las tres primeras sesiones del Naciones Unidas

procedimientos, aquí es donde cae cada nación. El panel de la parte inferior izquierda muestra un gráfico de pantalla, que enumera el valor propio asociado con cada dimensión. Los valores propios más grandes indican que una dimensión tiene más poder explicativo. Como en todas las gráficas de pedregal, cada dimensión adicional tiene un valor explicativo menor que la anterior.<sup>10</sup> El panel superior muestra la distribución de los ángulos de las líneas de corte. Las líneas de corte dividen el sí de los votos negativos sobre un tema determinado. El hecho de que tantas líneas de corte estén cerca de los 90° marca indica que la primera dimensión es importante para muchos de los

<sup>10</sup>Al elegir cuántas dimensiones incluir en un modelo de medición, muchos académicos usan la "regla del codo", lo que significa que no incluyen ninguna dimensión más allá de un codo visual en el diagrama de la pantalla. En este caso, un erudito ciertamente no incluiría más de tres dimensiones y podría contentarse con dos. Otro límite común es incluir cualquier dimensión para la cual el valor propio exceda 1, lo que nos haría detenernos en dos dimensiones.

votos. Finalmente, el panel inferior derecho muestra la malla de Coombs de este modelo, una visualización de cómo todas las líneas de corte de los 237 votos se unen en un solo espacio.

Si el usuario está satisfecho con los resultados de este modelo de medición, entonces es sencillo escribir las puntuaciones en un formato de datos utilizable. Dentro de nuestrodominar salida nombrada resultado podemos llamar al atributo llamado legisladores, lo que guarda nuestros puntos ideales para todos los países, cualquier variable no nominal que especifiquemos (por ejemplo, Pacto de Varsovia o no) y una variedad de otras medidas. Guardamos esto como un nuevo marco de datos llamado puntuaciones y luego escríbalo en un archivo CSV:

```
puntuaciones <-resultado $ legisladores
write.csv (puntuaciones, "UNscores.csv")
```

Solo recuerda usar el setwd comando para especificar la carpeta en la que desea guardar el archivo de salida.

Una vez que tenemos nuestros puntos ideales W-NOMINAR en un marco de datos separado, podemos hacer cualquier cosa que haríamos normalmente con datos en R, como dibujar nuestros propios gráficos. Supongamos que quisiéramos reproducir nuestro propio gráfico de los puntos ideales, pero queríamos marcar qué naciones eran miembros del Pacto de Varsovia frente a las que no lo eran. Podríamos hacer esto fácilmente usando nuestropuntuaciones datos. La forma más sencilla de hacer esto podría ser utilizar el subconjunto comando para crear marcos de datos separados de nuestros dos grupos:

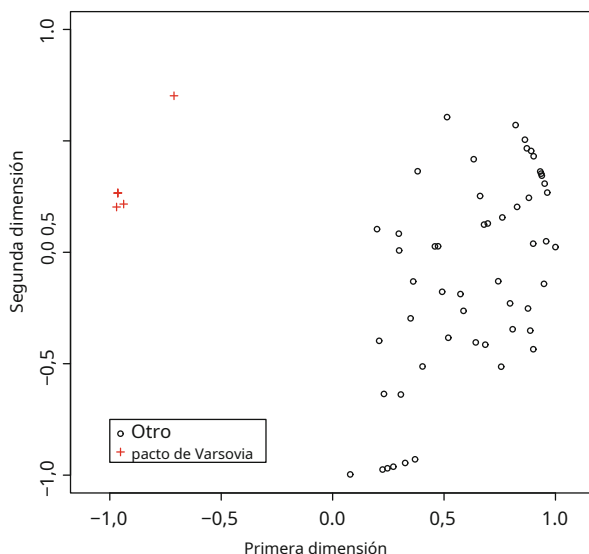
```
wp.scores <-subconjunto (puntuaciones, V1 == "WP")
other.scores <-subconjunto (puntuaciones, V1 == "Otro")
```

Una vez que tengamos estos subconjuntos en la mano, podemos crear el gráfico relevante en tres líneas de código.

```
plot (x = other.scores $ coord1D, y = other.scores $ coord2D,
      xlab = "Primera dimensión", ylab = "Segunda dimensión", xlim = c
      (-1,1), ylim = c (-1,1), asp = 1)
puntos (x = wp.scores $ coord1D, y = wp.scores $ coord2D,
        pch = 3, col = 'rojo')
leyenda (x = -1, y = -. 75, leyenda = c ("Otro", "Pacto de Varsovia"),
        col = c ("negro", "rojo"), pch = c (1,3))
```

En la llamada a gráfico, graficamos las 53 naciones que no eran miembros del Pacto de Varsovia, colocando la primera dimensión en el eje horizontal y la segunda en el eje vertical. Etiquetamos nuestros ejes apropiadamente usando xlab y ylab. También establecemos los límites de nuestro gráfico de 1 a 1 en ambas dimensiones, ya que las puntuaciones están limitadas a caer en estos rangos. Es importante destacar que garantizamos que la escala de las dos dimensiones es la misma, como deberíamos generalmente para este tipo de modelo de medición, estableciendo el **áspidect** relación a 1 (asp = 1). En la segunda línea de código, usamos el puntos comando para agregar las seis observaciones que estaban en el Pacto de Varsovia, coloreando estas observaciones de rojo y usando un carácter de trama diferente. Por último, agregamos una leyenda.

Figura 8.4 presenta la salida de nuestro código. Este gráfico transmite inmediatamente que la primera dimensión está capturando la división de la Guerra Fría entre Estados Unidos y sus aliados versus la Unión Soviética y sus aliados. Especificamos que EE. UU. Tomaría coordenadas positivas en ambas dimensiones, por lo que podemos ver que los aliados soviéticos



**Figura 8.4** Gráfico de la primera y segunda dimensiones de las puntuaciones W-NOMINATE de las tres primeras sesiones de las Naciones Unidas. A *Cruz Roja* indica un miembro del Pacto de Varsovia, y un *circulo negro* indica todos los demás miembros de la ONU (figura en color en línea)

(representados con cruces rojas) están en los extremos de los valores negativos en la primera dimensión.

En resumen, este capítulo ha ilustrado cuatro ejemplos de cómo utilizar R paquetes para implementar métodos avanzados. El hecho de que estos paquetes estén disponibles gratuitamente hace que el trabajo de vanguardia en metodología política y de una variedad de disciplinas esté fácilmente disponible para cualquier persona. R usuario. Ningún libro podría mostrar todos los paquetes aportados por los investigadores que están disponibles, sobre todo porque los nuevos paquetes están disponibles de forma regular. La próxima vez que se encuentre enfrentando un problema metodológico impositivo, es posible que desee verificar los servidores de CRAN para ver si alguien ya ha escrito un programa que proporcione lo que necesita.

## 8.5 Problemas de práctica

Este conjunto de problemas de práctica considera cada una de las bibliotecas de ejemplo por turno, y luego sugiere que intente usar un paquete nuevo que no se ha discutido en este capítulo. Cada pregunta requiere un conjunto de datos único.

1. Regresión logística multinivel: revise Singh's (2015) datos sobre la participación electoral en función de las reglas de votación obligatorias y varios otros predictores. Si no tiene el archivo `stdSingh.dta`, descárguelo del Dataverse (consulte la página vii) o el contenido del capítulo (consulte la página 125). (Estos datos se introdujeron por primera vez en

Secta. 7.4.) Ajuste este modelo de regresión logística utilizando el `resplandecer` comando, e incluir una intercepción aleatoria por país-año (**cntryyear**). Recuerde que el resultado es la participación (**votado**). La severidad de las reglas de voto obligatorio (**gravedad**) interactúa con los primeros cinco predictores: edad (**edad**), conocimiento político (**polinfrel**), ingreso (**ingreso**), eficacia (**eficacia**), y partidismo (**partyID**). Se deben incluir cinco predictores más solo para efectos aditivos: magnitud del distrito (**dist\_magnitude**), número de fiestas (**enep**), margen de victoria (**vicmarg\_dist**), sistema parlamentario (**parlamentario**), y PIB per cápita (**desarrollo**). Nuevamente, todas las variables predictoras se han estandarizado. ¿Qué aprende de este modelo de regresión logística multinivel estimado con `resplandecer` que no aprende de un modelo de regresión logística agrupado estimado con `glm`?

2. Modelo Bayesiano de Poisson con MCMC: Determine cómo estimar un modelo de Poisson con MCMC usando `MCMCpack`. Recargar Peake y Eshbaugh-Soha's (2008) datos sobre la cobertura de noticias sobre política energética, discutidos por última vez en la secc. 7.3. Si no tiene el archivo `PESEnergy.csv`, puede descargarlo del Dataverse (consulte la página vii) o del contenido del capítulo (consulte la página 125). Estime un modelo Bayesiano de Poisson en el que el resultado es la cobertura energética (**Energía**), y los insumos son seis indicadores para los discursos presidenciales (**rmn1173**, **grf0175**, **grf575**, **jec477**, **jec1177**, y **jec479**), un indicador del embargo petrolero árabe (**embargo**), un indicador de la crisis de los rehenes en Irán (**rehenes**), el precio del petróleo (**oilc**), aprobación presidencial (**Aprobación**), y la tasa de desempleo (**Desempleo**). Utilice una prueba de Geweke para determinar si existe alguna evidencia de no convergencia. ¿Cómo debería cambiar su código en R si la no convergencia es un problema? Resuma sus resultados en una tabla y muestre una gráfica de densidad del coeficiente parcial del discurso de Richard Nixon de noviembre de 1973 (**rmn1173**).
3. Coincidencia exacta aproximada: en el cap. 5, los problemas de práctica introducidos por Álvarez et al. (2013) datos de un experimento de campo en Salta, Argentina, en el que algunos votantes emitieron su voto a través del voto electrónico y otros votaron en el entorno tradicional. Carga el extranjero biblioteca y abra los datos en formato Stata. Si no tiene el archivo `alpl2013.dta`, puede descargarlo del Dataverse (consulte la página vii) o del contenido del capítulo (consulte la página 125). En este ejemplo, la variable de tratamiento es si el votante utilizó el voto electrónico o el voto tradicional (**EV**). Las covariables son el grupo de edad (**grupo de edad**), la educación (**educ**), trabajador de cuello blanco (**de cuello blanco**), no es un trabajador a tiempo completo (**not\_full\_time**), masculino (**masculino**), una variable de conteo para el número de seis posibles dispositivos tecnológicos utilizados (**tecnología**), y una escala ordinal para el conocimiento político (**pol\_info**). Utilizar el `cem` biblioteca para responder a lo siguiente:

una. ¿Cuán equilibradas están las observaciones de tratamiento y control en los datos brutos?

B. Realice una coincidencia exacta aproximada con `elcem` mando. ¿Cuánto ha mejorado el equilibrio como resultado?

C. Considere tres posibles variables de respuesta: si el votante evaluó positivamente la experiencia de votación (**eval\_voting**), si el votante evaluó la velocidad de la votación como rápida (**velocidad**), y si el votante está seguro de que se está contando su voto (**seguro\_contado**). ¿Cuál es el efecto del tratamiento promedio en los tratados (ATT) en su conjunto de datos emparejados para cada una de estas tres respuestas?

- D. ¿En qué se diferencian sus estimaciones de los efectos promedio del tratamiento en los tratados de las pruebas simples de diferencia de medias?
4. W-NOMINATE: De vuelta en la Secta. 2.1, presentamos los datos del pase de lista de Lewis y Poole para el 113° Senado de los Estados Unidos. Consulte el código allí para leer estos datos, que están en formato de ancho fijo. El nombre del archivo `essen113kh.ord`, y está disponible en Dataverse (consulte la página vii) y el contenido del capítulo (consulte la página 125).
- una. Formatee los datos como una matriz y cree lo siguiente: una matriz separada solo de las 657 pasadas de lista, un vector de los números de identificación de ICPSR y una matriz de las variables de las pasadas de lista. Utilice todos estos para crear un Llamada de rol-objeto de clase. Las votaciones nominales se codifican de la siguiente manera: 1 D Sí; 6 D No; 7 y 9 D desaparecido, y 0 D no es un miembro.
- B. Estime un modelo W-NOMINATE bidimensional para este objeto de lista. Del resumen de sus resultados, informe lo siguiente: ¿Cuántos legisladores fueron eliminados? ¿Cuántos votos se eliminaron? ¿Fue la clasificación general correcta?
- C. Examine la gráfica de salida de su modelo estimado, incluidas las coordenadas W-NOMINATE y la gráfica de la pantalla. Con base en el diagrama de la pantalla, ¿cuántas dimensiones cree que son suficientes para caracterizar el comportamiento de votación en el 113 ° Senado? ¿Por qué?
5. Bono: intente aprender a usar un paquete que nunca antes haya usado. Instala el Amelia paquete, que realiza una imputación múltiple de los datos faltantes. Eche un vistazo a Honaker et al. (2011) artículo en el *Revista de software estadístico* tener una idea de la lógica de la imputación múltiple y aprender a hacer esto en R. Ajustar un modelo lineal en conjuntos de datos imputados utilizando el libre comercio datos del Amelia Biblioteca. ¿Qué encuentras?