

Cómo realizar un análisis estadístico multivariado en Rbrul con datos numéricos/continuos

Introducción

Este manual sirve como guía para estudiantes que quiere aprender a realizar análisis estadísticos en Rbrul. Trabajaremos específicamente con datos numéricos o continuos¹ ya que existen varios manuales que se enfoca en datos categóricos (véase <http://www.danielezrajohnson.com/rbrul.html>). Rbrul tiene la capacidad de realizar análisis de regresión, el cual es un proceso estadístico que evalúa las relaciones entre variables. El análisis de regresión permite observar cómo el valor de una variable dependiente cambia cuando cualquiera de las variables independientes es variada, mientras que se mantienen fijas las otras variables independientes. Hay regresiones logísticas, las cuales se realizan con datos binarios, y por lo tanto, los datos se expresan en proporciones o porcentajes. También hay regresiones lineales, las cuales se realizan con datos numéricos, y por lo tanto, los datos se expresan como promedios. Nuestro enfoque será la última.

Terminología

Antes de continuar, es necesario aclarar algunos términos. Una **variable** es una propiedad que tiene valores distintos. Es decir que varía. Por ejemplo, la /s/ española en posición de coda es una propiedad lingüística que puede ser pronunciada de formas diferentes, una sibilante, una aspirada o una forma elidida. Otra propiedad variable que se encuentra en cualquier lengua es acentuación silábica. Generalmente hablando, hay sílabas átonas y sílabas tónicas.

En las estadísticas, existen **variables dependientes** y **variables independientes**. Una variable dependiente es aquella que responde a los valores de otra variable. Es decir, los valores de la variable dependiente *dependen* de los valores de otra variable. También es conocida como *variable respuesta*. Los investigadores de la sociolingüística cuantitativa, por lo general, hablan de una *variable lingüística*, la cual viene siendo una variable dependiente; por ejemplo, la /s/ española en posición de coda. Recuerde que es la variable bajo estudio. Los valores de la variable dependiente (o respuesta o lingüística) son conocidos como **variantes**.

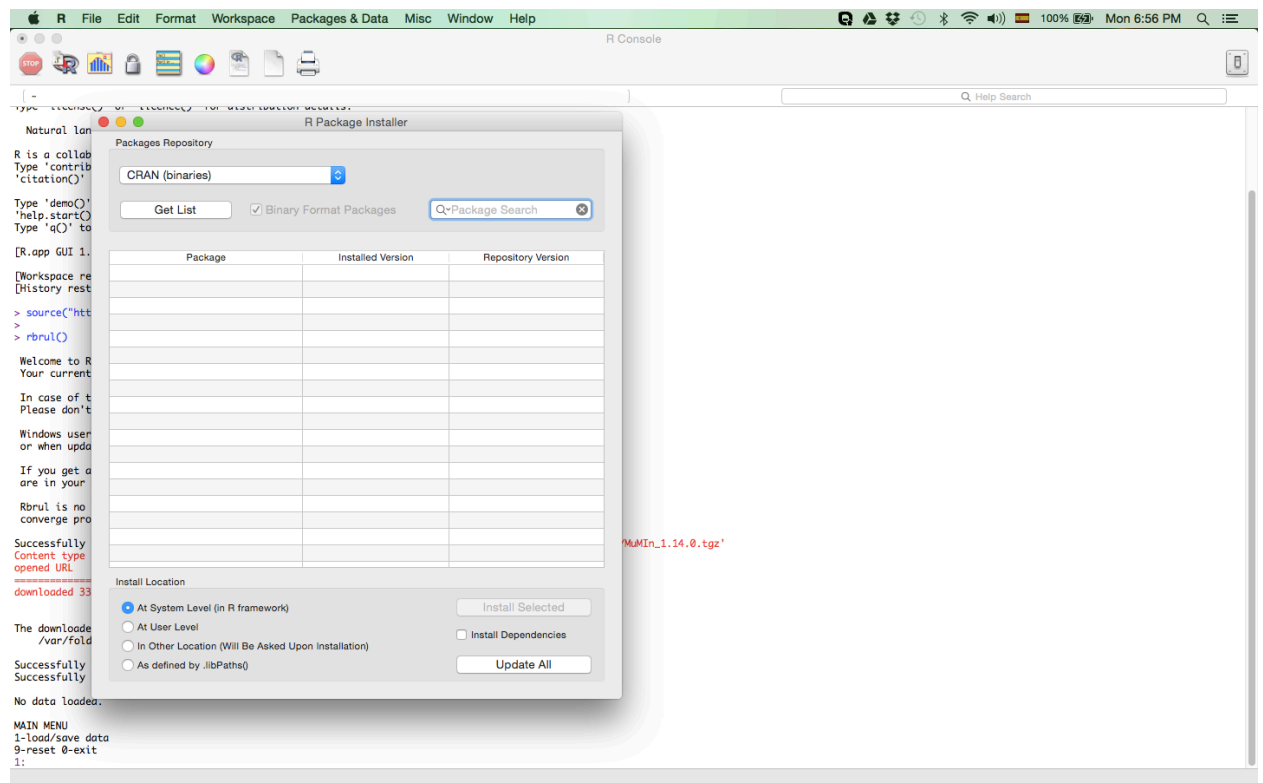
Las variables independientes son aquellas cuyos valores influyen en la variable dependiente y, por lo general, no dependen de los valores de otras variables. También se conocen como *variables explicativas* porque *explican* la variación de la variable dependiente. Por

¹ Los datos utilizados en este manual viene de un corpus de español panameño que contiene hablantes monolingües y bilingües. Se trata de un análisis de la producción de la /s/ en posición de coda en términos de una propiedad acústica conocida como centro de gravedad (o *center of gravity* en inglés). El centro de gravedad (CDG) es un rasgo acústico de sonidos fricativos. Mide la energía acumulada en la fricación. Mientras mayor el centro de gravedad, más energía concentrada tiene el sonido. En el caso de la /s/, se usa para detectar la reducción o debilitamiento de dicho sonido. Los sonidos fricativos linguales que tienen un punto de articulación anterior, como por ejemplo la sibilante, tienden a tener mucha energía en la fricación, y por lo tanto, un centro de gravedad alto. Los sonidos fricativos linguales que tienen un punto de articulación posterior, como la aspirada glotal (una forma reducida de la /s/), tienden a tener poca energía de la fricación, y por lo tanto, un centro de gravedad bajo (véase File-Muriel & Brown, 2011 para más información sobre esta propiedad y otras relacionadas a la /s/).

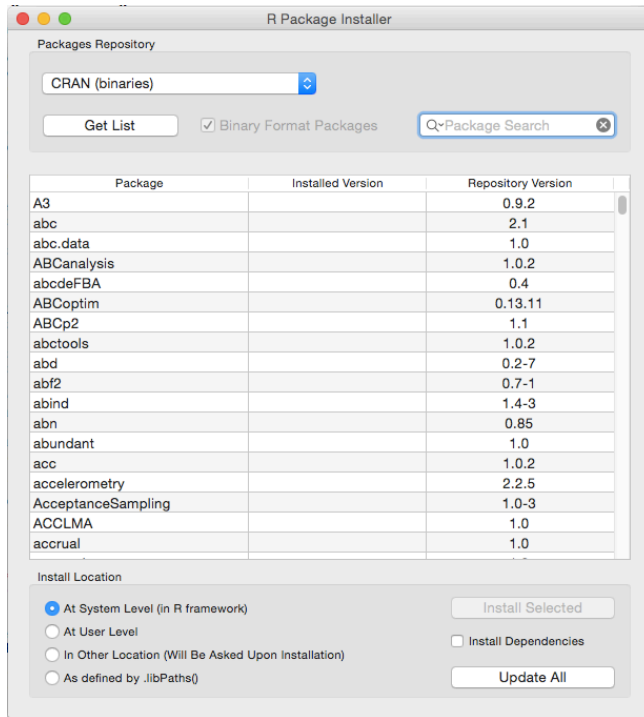
ejemplo, una sílaba átona tiende a favorecer el uso de una [s] aspirada mientras que una sílaba tónica tiende a favorecer el uso de una sibilante. La acentuación silábica es una variable independiente que explica la variabilidad de la propiedad lingüística /s/. En la tradición de la sociolingüística cuantitativa, las variables independientes son conocidos como *grupos factoriales* porque cada uno de sus valores es un factor, o sea que es un grupo de factores. Veremos que en Rbrul se usa el término *predictor* para referirse a factor ya que el predictor *predice* el comportamiento de la variable dependiente. En lo que sigue, se utilizarán indistintamente los términos variables independientes y grupos factoriales.

1. Abra R
2. INSTALAR PAQUETES

Por lo general, este paso sólo es necesario la primera vez: En el menú principal, seleccione “Package & Data” y dele clic a “Package Installer”. Se abrirá una nueva ventana. Esta ventana permite hacer una búsqueda de todos los paquetes disponibles en el mundo o ambiente de R.

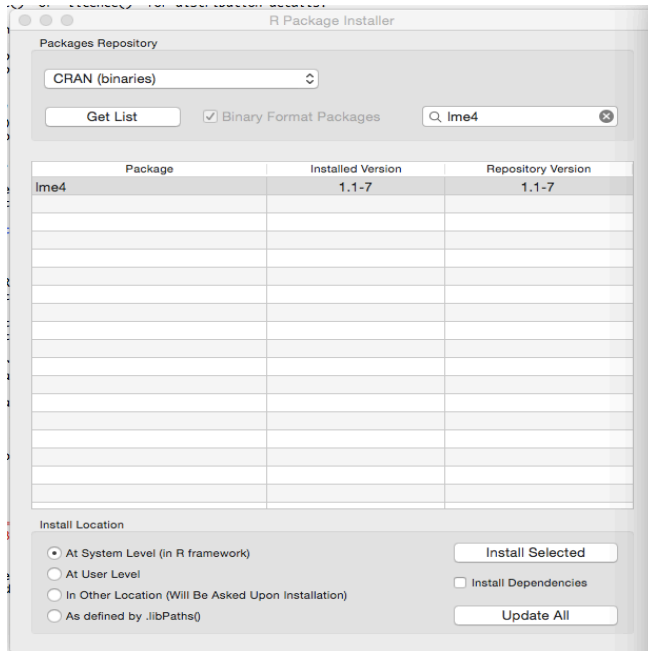


En el primer menú de esa ventana, debe estar seleccionado “CRAN (binaries)”. Luego, dele clic a “Get List”. Se llenará la lista de paquetes.



A la derecha de “Get List”, verá una barra de búsqueda donde escribirá **lme4**. Este paquete permite hacer análisis estadísticos multivariados (y otras cositas divertidas) en Rbrul. Al escribir el nombre del paquete, verá que se reducirá la lista y aparecerá el paquete necesario. Si ya está instalado, la segunda columna indicará la última versión instalada. La tercera columna indicará la versión actual disponible. Si no está instalado (o si tiene una versión antigua), instálelo seleccionando el paquete en la lista y dándole clic a “Install Selected”, el cual se encuentra por debajo de la lista de paquetes.

RECOMENDACIÓN: ponga un gancho ✓ en el cuadrado de “Install Dependencies”. Esta opción permite instalar paquetes adicionales de los cuales depende el paquete lme4. Algunos son absolutamente necesarios y otros no. Si no le pone gancho, tendrá que instalar estos paquetes manualmente. Después de la instalación, puede cerrar la ventana.



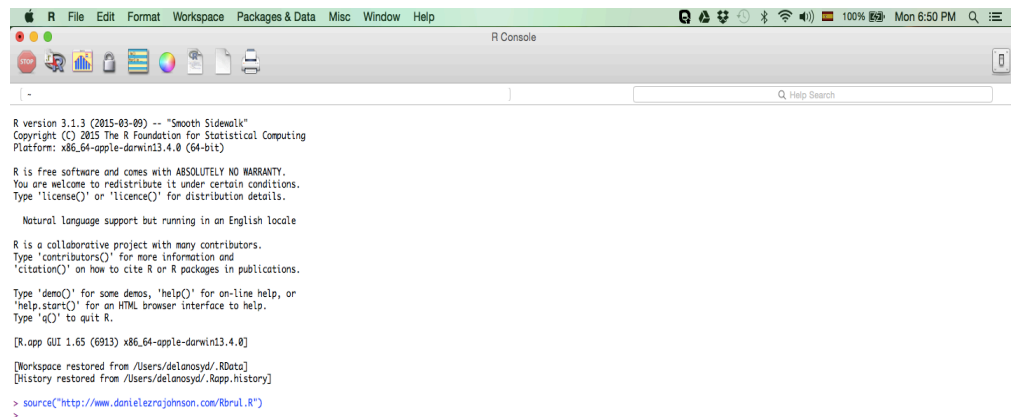
3. CARGAR RBRUL

El programa Rbrul es un archivo de texto con el nombre “Rbrul.R”. Para usarlo, tiene que cargarlo en R. Hay dos maneras de hacerlo:

PROCEDIMIENTO A

Puede usar el siguiente comando:

`source("http://www.danielezrajohanson.com/Rbrul.R")`. Cópelo y péguelo en la ventana principal de R. Después de un momento, aparecerá una flecha “>” debajo del comando. **OJO:** Este procedimiento requiere una conexión al internet ya que estará accediendo al programa desde la página web de Dan Johnson.



PROCEDIMIENTO B

Para acceder al programa sin internet, es necesario guardar el programa en su computadora. Diríjase a este enlace:

<http://www.danielezrajohnson.com/rbrul.html>. Allí se encuentra una lista de enlaces. El primero es el enlace a la última versión del programa Rbrul.

RBRUL

by Daniel Ezra Johnson

"A tool such as Rbrul offers a compromise of the old and new that I believe will be widely used in the near future." – [R. Harald Baayen](#)

"Using mixed models and adding individual speaker as a random effect results in interesting, logical results for my data. The results are conservative, but I like that. If I don't use speaker as random, I get loads of extra factors as significant, but lots of these make no sense and simply can't be explained. This again gives me confidence in my conservative approach." – a satisfied customer

"I've been using it a lot and finding it so much easier than trying to do the same in R." – another satisfied customer

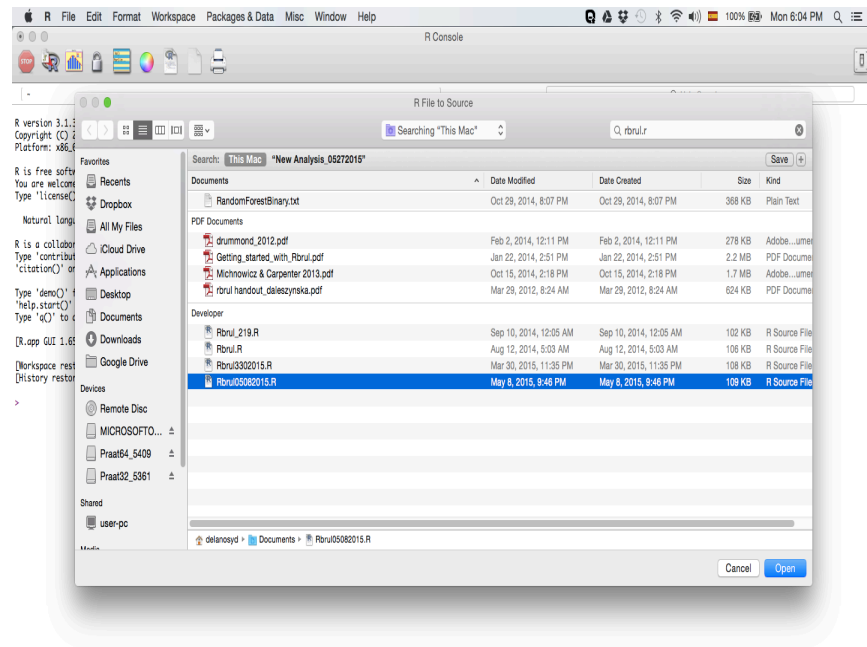
"Thanks Dan. You cast light upon my path into statistical wisdom." – another real testimonial

"I really like how you designed Rbrul. It is so user-friendly! And fast, too! Perhaps the best thing about using Rbrul is Dan's speedy responses to questions I pose to him. Thank you, Dan!" – more not-made-up feedback

- [Rbrul version 2.28](#) (Jun 3 2015)
 - Open Link in New Tab
 - Open Link in New Window
 - Open Link in New Private Window
 - [Rbrul release](#)
 - [Getting off to a good start](#) (If you are using Rbrul, you should read this paper.)
 - [Rbrul manual](#) (Mar 11 2010 - very out of date!)
 - Bookmark This Link
 - Save Link As...
 - Save Link to Pocket
 - Copy Link Location
 - Search Google for "Rbrul version 2..."
 - Inspect Element
 - [Workshop exercise 1](#) (.R text file, open in R)
 - [Solution to Workshop exercise 1](#) (.R text file, open in R)
 - [Workshop on using Rbrul](#) (by [Lynn Clark](#))
- Other workshop materials are available [here](#).

Haga clic derecho en ese enlace y seleccione “Save Link As...” para guardar el programa como archivo de texto en su computadora. El nombre del archivo debe ser “Rbrul.R”. **OJO:** Dan Johnson actualiza regularmente el programa, por lo que tendrá que volver a guardar el archivo periódicamente para asegurarse de que tenga la versión más actualizada. Se recomienda editar el nombre del archivo según el día en que fue guardado en su computadora, por ejemplo, “Rbrul_06222015.R”. Esto le recuerda de cuándo fue la última vez que guardó el archivo. También, si por alguna razón no funciona la versión actualizada en su computadora, puede volver a usar la versión anterior.

Para cargar el programa en R, en el menú de “File”, dele clic a “Source file”, busque el archivo de texto “Rbrul.R” (o cómo lo haya denominado) y selecciónelo.



4. RBRUL

Ahora puede correr Rbrul escribiendo el comando **rbrul()** en R y dándole a “Enter”. Después de un momento, aparecerá una descripción del programa. También se cargarán los paquetes necesarios para realizar los análisis estadísticos multivariados.

El más importante es el paquete **lme4**. Si los paquetes no están actualizados, se actualizarán automáticamente y luego se cargarán. Por último, aparecerá el menú principal de Rbrul.

5. CARGAR LOS DATOS

Escriba “1” para cargar o leer los datos que se analizarán y dele a “Enter”. El archivo debe estar en formato .csv o “comma-separated values”.

No data loaded.

```
MAIN MENU
1-load/save data
9-reset 0-exit
1: 1
```

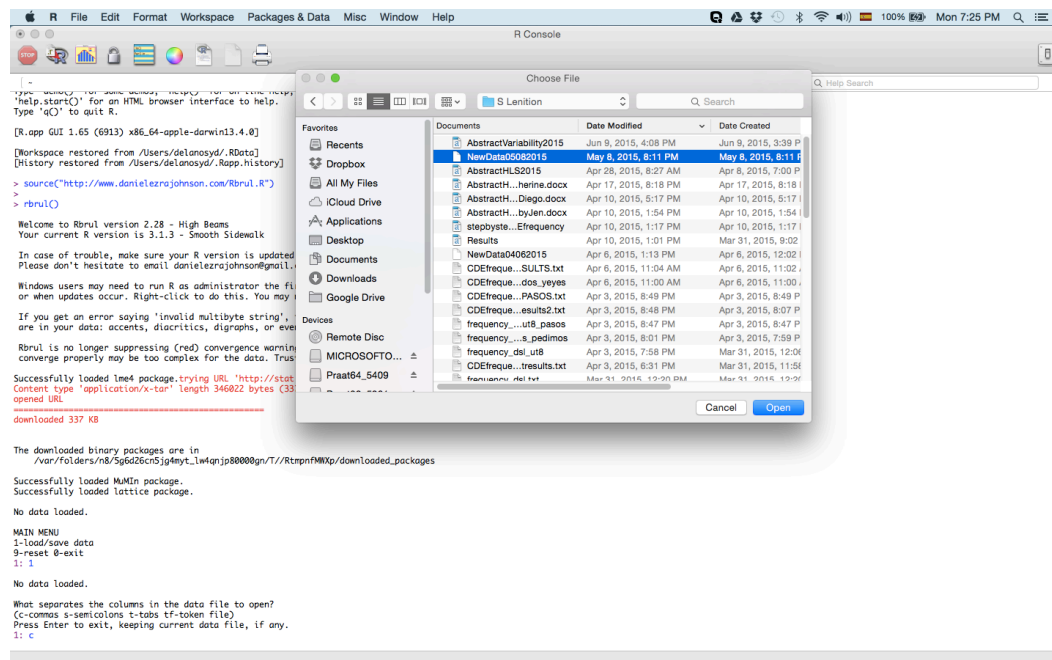
Luego, escriba “c” y dele a “Enter”.

```
MAIN MENU
1-load/save data
9-reset 0-exit
1: 1
```

No data loaded.

```
What separates the columns in the data file to open?
(c-commas s-semicolons t-tabs tf-token file)
Press Enter to exit, keeping current data file, if any.
1: c
```

Se abrirá una ventana para seleccionar el archivo de datos.



Una vez seleccionado el archivo, aparecerá en la ventana de R la estructura de los datos. Es decir, verá una lista de todas las columnas y los códigos utilizados en cada columna. Este es un buen momento de revisar sus datos para asegurarse de que todo esté codificado adecuadamente. Al final de la lista, aparece el total de los tokens u ocurrencias. Asegúrese de que sea correcto este número. Si encuentra algún error (código mal escrito, códigos duplicados en una sola columna, etc.) vuelva al archivo .csv y arregle el problema, y luego, vuelva a realizar este paso.

The screenshot shows the Rbrul application window. The title bar includes standard macOS window controls and the text 'R Console'. The menu bar contains 'R', 'File', 'Edit', 'Format', 'Workspace', 'Packages & Data', 'Misc', 'Window', and 'Help'. The main area displays the following text:

```

No data loaded.

What separates the columns in the data file to open?
(C-c commas s-semicolons t-tabs tf-token file)
Press Enter to exit, keeping current data file, if any.
1: c

Current data file is: /Users/delanosyd/Documents/S Lention/NewData05082015.csv

Current data structure:
ORDER (integer with 1892 values): 1 2 3 4 5 ...
Filename (factor with 32 values): Luis Angela.2 Helena.2SPN Sonia Marianela ...
PART (integer with 28 values): 11 2 0 17 14 ...
context (factor with 1493 values): de@pues despueH ella e! miHmo salon conocern@H de de viHta ...
new.word (factor with 544 values): despues mismo nos visto nosotros ...
word (factor with 548 values): despues mismo nos visto nosotros ...
variant (factor with 3 values): 0 h s
binaryvar (factor with 2 values): reduced sibilant
value (integer with 3 values): 0 1 2
Duration.ms.. (numeric with 1128 values): 0 81.903 45.529 83.354 45.459 ...
COG_MID (numeric with 1131 values): 0 2610.267243 1978.536722 2004.89107 2778.714104 ...
UNVOI (numeric with 212 values): 100 0 0.096 40 33.333 ...
posinword (factor with 2 values): m f
preseghheight.bin (factor with 2 values): minus plus
new.folsegh.bin (factor with 4 values): consonants minus zero plus
folseghheight.bin (factor with 3 values): minus plus zero
sylstress (factor with 2 values): a t
wordlength.bin (factor with 3 values): short long
freq.plus1 (integer with 320 values): 4895 3643 8706 1254 5241 ...
log.freq.plus1 (numeric with 320 values): 3.689752096 3.561459171 3.939818663 3.098297536 3.71941416 ...
coqstat (factor with 2 values): n y
eec (factor with 2 values): n y
sex (factor with 2 values): male female
hablante (factor with 2 values): mono bi
wordlength (integer with 9 values): 2 1 3 5 4 ...
typepresegh (factor with 2 values): vowel consonant
presegh (factor with 6 values): e i o u a ...
typefolsegh (factor with 3 values): cons vowel pause
folsegh (factor with 19 values): p e m d t ...
preseghheight (factor with 3 values): minus plus coronal
folseghheight (factor with 5 values): minuscor minus pluscor zero plus
frequency (integer with 320 values): 4894 3642 8705 1253 5240 ...
ras (integer with 2 values): 1 0
START (numeric with 837 values): 14.1864154 16.24798922 21.08678266 29.00596592 30.92423238 ...
END (numeric with 837 values): 14.24645281 16.31188092 21.16855473 29.04021842 31.23155956 ...
DUR (numeric with 834 values): 0.060037405 0.063891704 0.081772077 0.034252501 0.307327185 ...
Total tokens: 1892

MAIN MENU
1-load/save data 2-adjust data
4-crosstabs 5-modeling 6-plotting
8-restore data 9-reset 0-exit
1:

```

- a. **OJO:** En el menú principal de Rbrul (que aparece otra vez debajo de la lista), está la opción 2 que permite ajustar la codificación. Aunque es preferible arreglar la codificación en el mismo archivo .csv, también es posible utilizar esta función para hacer lo mismo. Escriba “2” y dele a “Enter”. Aparecerá un segundo menú “ADJUSTING MENU” en el cual escogerá la opción “5-recode”. Luego, puede escoger la columna a recodificar escribiendo el número correspondiente. Aparecen los códigos con sus números correspondientes. Simplemente, seleccione los que quiera ajustar o unir como un grupo dándole a “Enter” después de cada número. Una vez seleccionados los números deseados, dele a “Enter” otra vez y escriba el nuevo código y dele a “Enter”. Puede seguir recodificando la columna escribiendo otros números o puede darle a “Enter” para completar. Cuando le da “Enter”, tendrá la opción de crear una nueva columna para la nueva codificación (escriba el nombre de la nueva columna y dele a “Enter”), o puede darle a “Enter” para seguir con la columna original. Escriba “9-main menu” para volver al menú principal. Aparecerá de nuevo la lista de columnas, la cual incluirá la nueva codificación (aparece al final si es una nueva columna).

```

MAIN MENU
1-load/save data 2-adjust data
4-crosstabs 5-modeling 6-plotting
8-restore data 9-reset 0-exit
1: 2

ADJUSTING MENU
1-change class 2-rename 3-exclude 4-retain 5-recode
6-relevel 7-center/transform 8-count 9-main menu 0-exit
10-make full interaction group 11-make partial interaction group
1: 5
Factor group to recode? (press Enter to exit) (1-ORDER 2-Fileame 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNWOI 13-posimword 14-presesheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 23-sex 24-hablante 25-wordlength 26-typeprese 27-preses 28-typefolseg 29-folseg 30-presesheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR)
1: 31
Factor(s) of folsegheight to recode together? (1-minus 2-minusor 3-plus 4-plusor 5-zero Enter-done)
1: 1
2: 2
3:
Recode minus minusor as what?
1: minus
Factor(s) of folsegheight to recode together? (1-minus 2-minus 3-plus 4-plusor 5-zero Enter-done)
1:
Recode to new column? (Yes-type new column name No-press Enter)
1: example.folsegheight

ADJUSTING MENU
1-change class 2-rename 3-exclude 4-retain 5-recode
6-relevel 7-center/transform 8-count 9-main menu 0-exit
10-make full interaction group 11-make partial interaction group
1: 9

```

6. TABULACIONES CRUZADAS

Antes de realizar un análisis estadístico multivariado, es recomendable hacer una serie de tabulaciones cruzadas para familiarizarse con sus datos. Estas tabulaciones, también conocidas como tablas de contingencia, permiten detectar errores en la codificación y multicolinealidad (i.e., correlaciones fuertes entre variables independientes). Además, puede visualizar las distribuciones generales de la variable bajo estudio. Escriba la opción 4. Decida qué grupos factoriales o variables independientes quisiera analizar. La primera serán las columnas y la segunda, las filas o líneas. Dele a “Enter” después de cada selección. Luego, habrá una opción de “pages”. Ignoraremos esta opción por el momento. Déjelo en blanco y dele a “Enter”. Ahora, tendrá que decidir cómo quisiera ver los datos, como una proporción o un promedio o como una suma (“count”). Si quiere verlos como una suma, la cual es necesaria en los casos en que quiere hacer una prueba de chi cuadrado, dele a “Enter”. Si quiere ver los datos como una proporción o un promedio, escriba “1”. Por ahora, seleccionaremos la opción 1 puesto que las cifras tienen más significado. Dele a “Enter”. Luego, tiene que escribir el número que corresponda a la columna de la variable dependiente. Dele a “Enter”. Ahora, decida si la variable es binaria o continua (numérica). Este manual se trata de cómo realizar un análisis estadístico con datos numéricos así que escribirá la opción 1. Dele a “Enter”. Aparecerá la tabulación cruzada. Si no la ve, corra para arriba un poco utilizando la barra a la derecha.

```

MAIN MENU
1-load/save data 2-adjust data
4-crosstabs 5-modeling 6-plotting
8-restore data 9-reset 0-exit
1: 4
Cross-tab: factor for columns? (1-ORDER 2-Fileame 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNWOI 13-posimword 14-presesheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 23-sex 24-hablante 25-wordlength 26-typeprese 27-preses 28-typefolseg 29-folseg 30-presesheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight)
1: 23
Cross-tab: factor for rows? (1-ORDER 2-Fileame 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNWOI 13-posimword 14-presesheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 24-hablante 25-wordlength 26-typeprese 27-preses 28-typefolseg 29-folseg 30-presesheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight Enter-none)
1: 24
Cross-tab: factor for 'pages'? (1-ORDER 2-Fileame 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNWOI 13-posimword 14-presesheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 25-wordlength 26-typeprese 27-preses 28-typefolseg 29-folseg 30-presesheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight Enter-none)
1:
Cross-tab: variable for cells? (1-response proportion/mean Enter-counts)
1: 1
Choose response (dependent variable) by number (1-ORDER 2-Fileame 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNWOI 13-posimword 14-presesheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 23-sex 24-hablante 25-wordlength 26-typeprese 27-preses 28-typefolseg 29-folseg 30-presesheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight)
1: 11
Type of response? (1-continuous Enter-binary)
1: 1
mean of COG_MID
sex
hablante  female  male  total
bl  2298.844  1862.207  2080.914
mono  1569.917  1139.121  1402.198
total  1967.578  1610.888  1885.410

```

Primero que todo, en *la columna* de total, vemos que los bilingües tienen un promedio de CDG mayor que los monolingües, 2080.9 Hz vs. 1402.2 Hz. Esta observación nos indica que los bilingües reducen menos que los monolingües. Si miramos *la línea* de total, vemos que los hombres tienen un promedio de 1610.8 Hz, mientras que las mujeres tienen un promedio de 1967.6 Hz. Esto quiere decir que los hombres reducen más que las mujeres. Ahora, se puede apreciar también que los hombres monolingües son los que más reducen la /s/ con un promedio de 1139.1 Hz, luego siguen las mujeres monolingües con 1569.9 Hz, después siguen los hombres bilingües con 1862.2 Hz, y por último, las mujeres bilingües con 2298.8 Hz.

mean of COG_MID			
	sex		
hablante	female	male	total
bi	2298.844	1862.207	2080.914
mono	1569.917	1139.121	1402.198
total	1967.578	1610.808	1805.410

Se recomienda hacer tabulaciones cruzadas con todas las variables independientes. También se puede hacer tabulaciones cruzadas con las variables independientes y la variable dependiente. En las opciones de columnas y filas, escoja una variable independiente para las columnas y la variable dependiente para las filas. En este caso, necesitará una columna con la variable dependiente codificada categóricamente, por ejemplo, reducida vs. no reducida. Este tipo de tabulación permite observar si la codificación categórica corresponde a la variable continua. Por ejemplo, en la tabla abajo, se desprende que las variantes reducidas tienen un promedio de CDG menor en sílabas átonas (969.9 Hz) que las variantes no reducidas o sibilantes (5735.5 Hz). Se observa lo mismo en el caso de las sílabas tónicas. Además, las reducidas tienen un promedio menor en sílabas átonas (969.9 Hz) que en sílabas tónicas (1276.2 Hz). Por razones articulatorias, se esperan las dos observaciones.

mean of COG_MID			
	sylstress		
binaryvar	a	t	total
reduced	969.941	1276.152	1027.054
sibilant	5735.487	6181.546	5855.411
total	1671.868	2340.285	1805.410

Considero que este último ejemplo de tabulación cruzada se realiza mejor en Excel utilizando tablas Pivot.

Ahora, hagamos una tabulación cruzada con variables lingüísticas independientes y visualizando los datos como sumas. Esta vez las columnas serán de “altura del segmento siguiente” y las filas serán de “posición de /s/ en la palabra”, dos predictores que han sido importantes en estudios previos. En vez de proporción, se escoge la opción de “count” dándole a “Enter”.

```

MAIN MENU
1-load/save data 2-adjust data
4-crosstabs 5-modeling 6-plotting
8-restore data 9-reset 0-exit
1: 4
Cross-tab: factor for columns? (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNVOI 13-posinword 14-presegheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-wordlength 26-typeprese 27-preseg 28-typefolseg 29-folseg 30-presegheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight)
1: 31
Cross-tab: factor for rows? (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNVOI 13-posinword 14-presegheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-wordlength 26-typeprese 27-preseg 28-typefolseg 29-folseg 30-presegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight Enter-none)
1: 13
Cross-tab: factor for 'pages'? (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNVOI 14-presegheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-wordlength 26-typeprese 27-preseg 28-typefolseg 29-folseg 30-presegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight Enter-none)
1:
Cross-tab: variable for cells? (1-proportion/mean of COG_MID 2-change response variable Enter-counts)
1:
counts
      folsegheight
posinword minus minuscor plus pluscor zero total
f      266    446    75    340    310    1437
m       0     200     0    255     0     455
total   266    646    75    595    310    1892

```

Lo que vemos en la tabulación producida es que hay tres casos donde la suma de variantes es 0, lo cual quiere decir que no existe, por ejemplo, una ocurrencia en el interior de la palabra seguida de cero (o pausa). Esto tiene sentido porque una /s/ dentro de palabra siempre es seguida por algo. También se observa que las ocurrencias en el interior de la palabra nunca se encuentran seguidas de vocales altas o no altas². Es decir, nunca le siguen vocales. Esto también tiene sentido porque el análisis se trata de la /s/ en posición de coda, la cual siempre es seguida de una consonante. Si le siguiera una vocal no se consideraría una /s/ en coda, y por tanto, estaría fuera del contexto variable.

Estas observaciones son hechos del sistema lingüístico en general y nos dicen que se está midiendo lo mismo cuando se considera el efecto de posición en la palabra y la altura del segmento siguiente. Una variable parece depender de la otra. Esto puede causar un problema en el análisis estadístico ya que las variables independientes deben ser, pues, independientes. Esto es muy común y hay diferentes formas de lidiar con este asunto. Una es recodificar los factores (véase Sección 5a). En este caso, podríamos clasificar los segmentos siguientes como +cor/alto vs. –cor/alto, o sea, según la posición de la lengua. Sin embargo, la categoría de “cero” presenta todavía un problema. La otra solución es simplemente dejar fuera una de las variables independientes mientras que se analice la otra. Volveremos a este asunto en otro momento.

Counts						
	folsegheight					
posinword	minus	minuscor	plus	pluscor	zero	total
f	266	446	75	340	310	1437
m	0	200	0	255	0	455
total	266	646	75	595	310	1892

Después de hacer todas las posibles tabulaciones, puede proceder al análisis estadístico multivariado.

² “minus” se refiere a vocales no altas como “a”, “e” y “o”. “plus” se refiere a vocales altas como “i” y “u”. “minuscor” se refiere a consonantes no corales como “p” o “k”. “pluscor” se refiere a consonantes corales como “t”, “n” o “d”.

7. ELABORACIÓN DE MODELOS ESTADÍSTICOS

Para diseñar el modelo estadístico, se escoge la opción 5-modeling en el menú principal. Aparecerá otro menú, “MODELING MENU”. Luego, escriba “1” y dele a “Enter” para seleccionar las variables. Primero escoja la variable dependiente, en este caso, será el centro de gravedad, una variable continua. Dele a “Enter” después de escribir el número correspondiente. Después tendrá que especificar si la variable es continua o binaria. Escriba “1” para continua. Ahora tiene que escoger todas las variables independientes que quisiera incluir en el modelo. Escriba los números dándole a “Enter” después de cada uno. Al final, dele a “Enter” para seguir al próximo paso donde tendrá que especificar cuáles son las variables independientes continuas, por ejemplo, duración, frecuencia, velocidad de habla, etc. Por ahora, sólo vamos a trabajar con variables independientes categóricas, así que puede darle a “Enter” sin escribir una opción. Ahora Rbrul le pregunta si quiere incluir algunas pruebas de interacción entre las variables independientes. Ignoraremos esta parte por el momento. Dele a “Enter”. Luego, le pregunta si algunas de las variables independientes son aleatorias. También ignoraremos esto por el momento. Dele a “Enter”. Ahora volvemos a “MODELING MENU”.

```
MAIN MENU
1-load/save data 2-adjust data
4-crossabs 5-modeling 6-plotting
8-restore data 9-reset 0-exit
1: 5

Current variables are:
response.binary: binaryvar (reduced vs. sibilant)
fixed.factor: folseghheight sylstress sex hablante

MODELING MENU
1-choose variables 2-one-level (recommended)
3-step-up 4-step-down 5-step-up/step-down
6-trim 7-plotting 8-settings 9-main menu 0-exit
10-chi-square test
1: 1
Choose response (dependent variable) by number, or Enter to keep binaryvar (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms.. 11-COG_MID 12-UNVOI 13-
posinword 14-presegheight.bin 15-new.folseg.bin 16-folseghheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-wordlength 26-typepresseg 27-presseg
28-typefolseg 29-folseg 30-presegheight 31-folseghheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folseghheight)
1: 11
Type of response? (1-continuous Enter-binary)
1: 1
Choose predictors (independent variables) by number, or Enter to keep folseghheight & sylstress & sex & hablante (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-
Duration.ms.. 12-UNVOI 13-posinword 14-presegheight.bin 15-new.folseg.bin 16-folseghheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-
wordlength 26-typepresseg 27-presseg 28-typefolseg 29-folseg 30-presegheight 31-folseghheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folseghheight)
1: 31
2: 13
3: 17
4: 23
5: 24
6:
Are any predictors continuous? (31-folseghheight 13-posinword 17-sylstress 23-sex 24-hablante Enter-none)
1:
Consider a pairwise interaction of fixed effects (for interactions between fixed and random effects, use random slopes, below)? (Choose two at a time.
(31-folseghheight 13-posinword 17-sylstress 23-sex 24-hablante Enter-done)
1:
Any random intercepts? (31-folseghheight 13-posinword 17-sylstress 23-sex 24-hablante Enter-none)
1:

Current variables are:
response.continuous: COG_MID
Fixed.factor: folseghheight posinword sylstress sex hablante
```

Hay varias opciones de cómo podemos correr el análisis estadístico.

- La opción 2-one-level: un análisis en el cual se incluyen todos los factores en una corrida para determinar cuáles tienen un efecto significativo en la distribución de la variable dependiente.
- La opción 3-step-up: se añade un factor a la vez comenzando con el que tiene el mayor efecto en la variable dependiente. Se repite el proceso hasta que no haya más factores con un efecto significativo.
- La opción 4-step-down: se incluyen todos los factores en el modelo y luego Rbrul va eliminando los factores que no son significativos.
- La opción 5-step-up/step-down es simplemente los dos análisis anteriores en sólo paso.

Vemos en paréntesis que la opción 2-one-level es la recomendada así que solo trabajaremos con esta opción. Escriba “2” y dele a “Enter”. Aparecerán los resultados del análisis.

```
MODELING MENU
1-choose variables 2-one-level (recommended)
3-step-up 4-step-down 5-step-up/step-down
6-trim 7-plotting 8-settings 9-main menu 0-exit
10-chi-square test
1: 2
```

La primera línea de los resultados es una lista de los factores que fueron incluidos en el análisis. Después de esa línea, aparecen los grupos factoriales de nuevo con sus respectivos factores o niveles. Luego sigue información sobre el análisis en general. Vemos que hay varias cifras. Estas cifras son importantes para entender cómo está distribuida la variable según los factores.

```
ONE-LEVEL ANALYSIS OF RESPONSE COG_MID WITH PREDICTOR(S): folsegheight (2.15e-46) + hablante (2.41e-13) + posinword (1.46e-10) + sex (4.87e-09) + sylstress (7.8e-09)

$folsegheight
factor      coef tokens    mean
zero 1332.591    310 3125.326
plus  145.698     75 2090.127
minus -203.440    266 1706.716
pluscor -511.668   595 1680.277
minuscor -763.181  646 1294.851

$posinword
factor      coef tokens    mean
m    387.593    455 2036.464
f   -387.593   1437 1732.251

$sylstress
factor      coef tokens    mean
t    340.604    378 2340.285
a   -340.604   1514 1671.868

$sex
factor      coef tokens    mean
female 276.038   1032 1967.578
male  -276.038    860 1610.808

$hablante
factor      coef tokens    mean
bi    349.677   1124 2080.914
mono -349.677    768 1402.198

$misc
deviance      AIC df intercept grand mean R2.fixed
7702110978 34184.31 9 2346.171 1805.41 0.151

Current variables are:
response.continuous: COG_MID
fixed.factor: folsegheight posinword sylstress sex hablante
```

- e. **Valores p:** la primera línea de los resultados contiene números en paréntesis conocidos como valores p. Este valor indica la significancia estadística de los grupos factoriales. Un valor p por debajo de 0.05 indica que el grupo factorial condiciona significativamente la variable dependiente. En el ejemplo, se observa que todos los grupos factoriales son significativos ya que todos los valores p están por debajo de 0.05 (e-46 significa que el decimal se mueve 46 veces a la izquierda, o sea, muy por debajo de 0.05).
- f. **Coef (coeficientes):** reflejan la fuerza de la relación entre el factor y la variable dependiente. Los coeficientes negativos indican una correlación negativa y los coeficientes positivos indican una correlación positiva. Mientras más grande el valor, más fuerte es la relación. Se puede expresar esa relación también en términos de favorecimiento hacia el promedio de la variable dependiente. Los coeficientes negativos indican que el factor favorece un promedio menor. Los positivos indican que el factor favorece un promedio mayor. Se observan los promedios en orden descendente para cada factor en la última columna de cada grupo. **OJO:** Muchos investigadores que trabajan dentro del marco variacionista

se preocupan cuando los promedios (o porcentajes en los casos de variables binarias) no coinciden con los coeficientes. Según la tradición variacionista, cuando esto ocurre, es un indicio de que hay interacción entre dos grupos factoriales. Hay varios pasos que se pueden tomar, primero, para determinar cuáles son los grupos factoriales que interactúan, y segundo, para resolver la interacción. Volveremos a este asunto en otro momento.

A modo de ejemplo, vamos a interpretar los resultados del efecto de “altura del segmento siguiente”. El valor p de esta variable independiente es $2.15e-46$, lo cual quiere decir que tiene un efecto significativo en la variable dependiente puesto que está por debajo de 0.05. Este hallazgo es importante para el estudio de la /s/ en coda porque se ha observado en la literatura que el contexto fonético en el que aparece el sonido afecta cómo se produce. Al considerar los coeficientes, se desprende que cuando la /s/ aparece en posición final del enunciado, o sea, hay una pausa o cero, el promedio de CDG es más alto porque el coeficiente es positivo. Es decir, este contexto fonético favorece la retención de la /s/ o un promedio más alto. Se ve también que cuando la vocal es +alta también se favorece la retención de /s/ o un promedio alto. Además, una vocal –alta favorece el debilitamiento de la /s/ o un promedio más bajo ya que el coeficiente es negativo. En el caso del punto de articulación de la consonante siguiente, se favorece un promedio más bajo tanto con +coronal como con –coronal. Note que el orden es +coronal primero y luego –coronal.

Estos hallazgos se alinean con lo que se ha visto en estudios previos y explican lo que ocurre articulatoriamente en la producción de la /s/. Con vocales altas y consonantes coronales, el ápice de la lengua se encuentra más hacia la región alveolar. Esto facilita la producción de la /s/ ya que este sonido tiene un punto de articulación alveolar. Cuando la lengua está más abajo en la boca, como en el caso de las vocales no altas y las consonantes no coronales, se dificulta la elevación de la lengua hacia los alvéolos para producir la /s/ y, por lo tanto, se debilita dicho fono.

\$folsheight			
factor	coef	tokens	mean
zero	1332.591	310	3125.326
plus	145.698	75	2090.127
minus	-203.44	266	1706.716
pluscor	-511.668	595	1680.277
minusc	-763.181	646	1294.851

Note que, después de la última variable independiente, hay una sección titulada \$misc. Esta es información general que provee Rbrul sobre el modelo producido. Es particularmente importante cuando diseñamos varios modelos con los mismos datos. Se utiliza esta información para comparar el ajuste de los modelos.

Deviance (desviación) – medida de cuán bien se ajusta el modelo a los datos o cuánto se desvían los datos de las predicciones del modelo. En otras palabras, mide lo bien que se

explican los datos con el modelo. Mientras más grande la desviación, peor es el ajuste. La desviación tiende a aumentar a medida de que se vayan añadiendo predictores o variables independientes al modelo.

Degrees of freedom (df – grado de libertad) – df es el número de parámetros en el modelo, una medida de la complejidad del modelo. Mientras más factores en el modelo, mayor es el grado de libertad.

Intercept (intercepto) – Es el valor estimado de la variable dependiente si $x=0$.

Grand mean (promedio general) – es el promedio general de todos los datos.

R² – se refiere a la proporción de la variación explicada por el modelo.

8. COMPARACIÓN DE DIFERENTES MODELOS

Recuerde que al final de la Sección 6 hicimos una tabulación cruzada entre “altura del segmento siguiente” y “posición en la palabra” y nos percatamos de que cuando aparece la /s/ en el interior de la palabra nunca le siguen vocales ni pausas. No le siguen vocales porque siempre es una /s/ en coda, un contexto en que solo siguen consonantes; si siguiera una vocal, no sería /s/ en coda. No le siguen pausas porque está dentro de la palabra. Concluimos de esto que hay una correlación entre estas dos variables independientes, la cual se conoce como multicolinealidad. La multicolinealidad puede afectar los valores estimados o coeficientes que produce el modelo estadístico. Es decir, puede sesgar los resultados del análisis aun cuando las dos variables salen significativas. Para evitar coeficientes imprecisos, se recomienda excluir una de las variables mientras que se analiza el efecto de la otra. Esto quiere decir que tendremos dos modelos, uno con “altura del segmento siguiente” y otro con “posición en la palabra”. Luego, hacemos una comparación de los dos modelos para ver cuál tiene el mejor ajuste a los datos.

Vuelva a hacer los pasos para diseñar el modelo (note que si ya se especificó la variable dependiente previamente sólo hace falta darle a “Enter”). Vamos a elaborar un modelo con todas las variables independientes previamente incluidas menos “posición en la palabra”. Es decir, el modelo tendrá “altura del segmento siguiente”, “acentuación silábica”, “sexo” y “tipo de hablante”.


```

MODELING MENU
1-choose variables 2-one-level (recommended)
3-step-up 4-step-down 5-step-up/step-down
6-trin 7-plotting 8-settings 9-main menu 0-exit
10-chi-square test
1: 1
Choose response (dependent variable) by number, or Enter to keep COG_MID (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms... 11-COG_MID 12-UNVOI 13-
postword 14-presheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 23-sex 24-hablante 25-wordlength 26-typepresg 27-presg
28-typefolseg 29-folseg 30-presheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight)
1:
Choose predictors (independent variables) by number, or Enter to keep folsegheight & sylstress & sex & hablante (1-ORDER 2-Filename 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-
Duration.ms... 12-UNVOI 13-postword 14-presheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-ec 23-sex 24-hablante 25-
wordlength 26-typepresg 27-presg 28-typefolseg 29-folseg 30-presheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight)
1: 31
2: 17
3: 23
4: 24
5:
Are any predictors continuous? (31-folsegheight 17-sylstress 23-sex 24-hablante Enter-none)
1:
Consider a pairwise interaction of fixed effects (for interactions between fixed and random effects, use random slopes, below)? Choose two at a time.
(31-folsegheight 17-sylstress 23-sex 24-hablante Enter-done)
1:
Any random intercepts? (31-folsegheight 17-sylstress 23-sex 24-hablante Enter-none)
1:

Current variables are:
response.continuous: COG_MID
fixed.factor: folsegheight sylstress sex hablante

MODELING MENU
1-choose variables 2-one-level (recommended)
3-step-up 4-step-down 5-step-up/step-down
6-trin 7-plotting 8-settings 9-main menu 0-exit
10-chi-square test
1: 2

```

Los resultados aparecen en la imagen abajo.

```

ONE-LEVEL ANALYSIS OF RESPONSE COG_MID WITH PREDICTOR(S): folsegheight (8.1e-38) + hablante (3.64e-14) + sylstress (2.32e-10) + sex (4.93e-08)

$folsegheight
  factor    coef tokens    mean
    zero 1220.815    310 3125.326
    plus  21.797     75 2090.127
  pluscor -291.971    595 1680.277
   minus -318.452    266 1706.716
 minuscor -632.188    646 1294.851

$sylstress
  factor    coef tokens    mean
    t 376.267    378 2340.285
    a -376.267   1514 1671.868

$sex
  factor    coef tokens    mean
 female 259.618   1032 1967.578
  male -259.618    860 1610.808

$hablante
  factor    coef tokens    mean
    bi 364.971   1124 2080.914
   mono -364.971    768 1402.198

$misc
  deviance    AIC df intercept grand mean R2.fixed
7870132433 34223.14  8 2090.607   1805.41   0.133

Current variables are:
response.continuous: COG_MID
fixed.factor: folsegheight sylstress sex hablante

```

Note que todos los grupos factoriales son significativos. También note la desviación: 7870132433. Ahora vamos a elaborar otro modelo, pero esta vez, en lugar de “altura del segmento siguiente” incluiremos “posición de la palabra” junto con las otras variables independientes. Los resultados aparecen en la imagen abajo.

ONE-LEVEL ANALYSIS OF RESPONSE COG_MID WITH PREDICTOR(S): hablante (1.7e-13) + sylstress (1.46e-08) + sex (4.51e-06) + posinword (0.0339)

```
$posinword
factor      coef tokens    mean
m    122.343    455 2036.464
f   -122.343   1437 1732.251

$sylstress
factor      coef tokens    mean
t    349.023    378 2340.285
a   -349.023   1514 1671.868

$sex
factor      coef tokens    mean
female  227.629   1032 1967.578
male   -227.629    860 1610.808

$hablante
factor      coef tokens    mean
bi    371.241   1124 2080.914
mono  -371.241    768 1402.198

$misc
deviance    AIC df intercept grand mean R2.fixed
8600847302 34385.12 5 1987.925 1805.41 0.052
```

Current variables are:
 response.continuous: COG_MID
 fixed.factor: posinword sylstress sex hablante

Vemos en los resultados del segundo modelo que todos los grupos factoriales son significativos. Note la desviación: 8600847302. Observamos que del primer modelo al segundo, aumentó la desviación, lo cual quiere decir que el segundo modelo tiene un peor ajuste. Es decir, los datos son explicados mejor con “altura del segmento siguiente”, “acentuación silábica”, “sexo” y “hablante” como variables independientes. Ahora queremos saber si esta diferencia entre los dos modelos es significativa. Si el primer modelo es significativamente mejor, nos quedamos con ése y descartamos el segundo. Para determinar esto, tendremos que realizar una prueba de chi cuadrado, la cual se encuentra como opción 10 en “MODELING MENU”. Escriba “10” y dele a “Enter”. Rbrul le pide la primera desviación. Introdúzcala y dele a “Enter”. Ahora introduzca la segunda desviación y dele a “Enter”. Luego Rbrul le pregunta si los valores que fueron introducidos fueron valores de “log likelihood”. La respuesta es no, así que dele a “Enter”. Ahora le pide que introduzca la diferencia entre los grados de libertad de los dos modelos. El df del primer modelo es 8 y el df del segundo es 5, así que la diferencia es 3. Escriba 3 y dele a “Enter”. Los resultados de la prueba de chi cuadrado aparecen en la imagen abajo.

```
MODELING MENU
1-choose variables 2-one-level (recommended)
3-step-up 4-step-down 5-step-up/step-down
6-trim 7-plotting 8-settings 9-main menu 0-exit
10-chi-square test
1: 10
Enter first deviance or log likelihood.
1: 7870132433
Enter second deviance or log likelihood.
1: 8600847302
If these were log likelihood values, press 1. Press Enter if they were deviances.
1:
Enter difference in degrees of freedom.
1: 3
Chi-square = 730714869, df = 3, p = 0
```

El número que más importa aquí es el valor p. Debe ser 0.05 o menos. El valor p de la prueba es 0. Es decir, la diferencia entre los dos modelos es significativa, y el primer

modelo es significativamente mejor que el segundo. Así que descartamos el segundo modelo.

Recuerde que en la sección 7f se mencionó que muchos variacionistas se preocupan cuando los promedios no coinciden con los coeficientes. En nuestro análisis, resulta que tenemos este caso. Observe en los resultados de “altura del segmento siguiente” que “pluscor” y “minus” no están en orden descendente; “minus” debe preceder a “pluscor”. Puede que haya otra variable independiente que esté interactuando con esta. Podemos determinar cuál es esa variante si hacemos una especie de “step-up” (véase 7b) manual en el que elaboramos varios modelos comenzando con “altura del segmento siguiente” (los promedios deben coincidir con los coeficientes cuando se analiza solo) y añadiendo grupos factoriales hasta que vemos el cambio en el orden. Sabremos que es el último grupo factorial añadido que interactúa con “altura del segmento siguiente”. Cuando hago este procedimiento, resulta que es “acentuación silábica” la que interactúa con “altura del segmento siguiente”.

ONE-LEVEL ANALYSIS OF RESPONSE COG_MID WITH PREDICTOR(S): folsegheight (1.82e-34)

```
$folsegheight
  factor    coef tokens    mean
    zero 1145.866   310 3125.326
    plus  110.667    75 2090.127
    minus -272.743   266 1706.716
 pluscor -299.182   595 1680.277
 minuscor -684.608   646 1294.851

$misc
 deviance      AIC df intercept grand mean R2.fixed
8347361567 34328.53  5 1979.459   1805.41    0.08
```

Current variables are:
response.continuous: COG_MID
fixed.factor: folsegheight

ONE-LEVEL ANALYSIS OF RESPONSE COG_MID WITH PREDICTOR(S): folsegheight (1.23e-35) + sylstress (6.47e-09)

```
$folsegheight
  factor    coef tokens    mean
    zero 1201.645   310 3125.326
    plus  29.468    75 2090.127
 pluscor -285.646   595 1680.277
    minus -311.139   266 1706.716
 minuscor -634.328   646 1294.851

$stylstress
  factor    coef tokens    mean
    t  351.128   378 2340.285
    a -351.128  1514 1671.868


$misc
 deviance      AIC df intercept grand mean R2.fixed
8200871776 34297.03  6 2168.338   1805.41    0.096
```

Current variables are:
response.continuous: COG_MID
fixed.factor: folsegheight sylstress

Se puede hacer una prueba de interacción para ver si es significativa o no. Cuando se elabora el modelo, Rbrul le permite escoger dos factores a la vez para hacer una prueba de interacción.

```
MODELING MENU
1-choose variables 2-one-level (recommended)
3-step-up 4-step-down 5-step-up/step-down
6-trin 7-plotting 8-settings 9-main menu 0-exit
10-chi-square test
1: 1
Choose response (dependent variable) by number, or Enter to keep COG_MID (1-ORDER 2-File name 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-Duration.ms... 11-COG_MID 12-UNVOI 13-
posimord 14-presgheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-wordlength 26-typepreseg 27-preseg
28-typefolseg 29-folseg 30-presgheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight 38-nueva.altseg)
1:
Choose predictors (independent variables) by number, or Enter to keep folsegheight.bin & sylstress & sex & hablante (1-ORDER 2-File name 3-PART 4-context 5-new.word 6-word 7-variant 8-binaryvar 9-value 10-
Duration.ms... 12-UNVOI 13-posimord 14-presgheight.bin 15-new.folseg.bin 16-folsegheight.bin 17-sylstress 18-wordlength.bin 19-freq.plus1 20-log.freq.plus1 21-cogstat 22-eec 23-sex 24-hablante 25-
wordlength 26-typepreseg 27-preseg 28-typefolseg 29-folseg 30-presgheight 31-folsegheight 32-frequency 33-ros 34-START 35-END 36-DUR 37-example.folsegheight 38-nueva.altseg)
1: 31
2: 17
3: 23
4: 24
5:
Are any predictors continuous? (31-folsegheight 17-sylstress 23-sex 24-hablante Enter-none)
1:
Consider a pairwise interaction of fixed effects (for interactions between fixed and random effects, use random slopes, below)? Choose two at a time.
(31-folsegheight 17-sylstress 23-sex 24-hablante Enter-done)
1: 31
2: 17
Consider a pairwise interaction of fixed effects (for interactions between fixed and random effects, use random slopes, below)? Choose two at a time.
(31-folsegheight 17-sylstress 23-sex 24-hablante Enter-done)
1:
Any random intercepts? (23-sex 24-hablante Enter-none)
1:

Current variables are:
response.continuous: COG_MID
fixed.factor: folsegheight sylstress sex hablante
fixed.interaction: folsegheight:sylstress
```



Los resultados del nuevo modelo se encuentra en la imagen abajo. Observamos que la interacción entre “acentuación silábica” y “altura del segmento siguiente” es significativa. También notamos que cuando se incluye la prueba de interacción, se arregla el orden de los promedios en “altura del segmento siguiente”. A veces la interacción tiene implicaciones importantes; depende del interés del investigador. No entraremos en detalle de lo que significa la interacción en este análisis.

ONE-LEVEL ANALYSIS OF RESPONSE COG_MID WITH PREDICTOR(S): hablante (1.03e-14) + sex (7.6e-08) + folsegheight:sylstress (9.33e-05) + folsegheight [main effect, not tested] + sylstress [main effect, not tested]

\$folsegheight

factor	coef	tokens	mean
zero	1382.147	310	3125.326
plus	214.584	75	2090.127
minus	-320.772	266	1706.716
pluscor	-497.107	595	1680.277
minuscor	-778.853	646	1294.851

\$sylstress

factor	coef	tokens	mean
t	564.073	378	2340.285
a	-564.073	1514	1671.868

\$sex

factor	coef	tokens	mean
female	254.701	1032	1967.578
male	-254.701	860	1610.808

\$hablante

factor	coef	tokens	mean
bi	371.115	1124	2080.914
mono	-371.115	768	1402.198

\$`folsegheight:sylstress`

factor:factor	coef	tokens	mean
plus:t	651.177	26	3622.931
pluscor:a	432.775	469	1633.311
minuscor:a	309.211	543	1218.696
zero:t	137.115	47	4302.892
minus:a	46.306	190	1443.447
minus:t	-46.306	76	2364.888
zero:a	-137.115	263	2914.886
minuscor:t	-309.211	103	1696.330
pluscor:t	-432.775	126	1855.095
plus:a	-651.177	49	1276.802

\$misc

deviance	AIC	df	intercept	grand mean	R2.fixed
7772300037	34207.48	12	2153.911	1805.41	0.143

Current variables are:

response.continuous: COG_MID

fixed.factor: folsegheight sylstress sex hablante

fixed.interaction: folsegheight:sylstress

9. PRESENTACIÓN DE LOS RESULTADOS

Lo siguiente es un ejemplo de cómo se presentan los resultados en una tabla.

Center of Gravity of Coda /s/				
Total				1892
Deviance				7772300037
Intercept				2153.911
	Factor	Coef	N	Mean
Speaker Type				
	Bilingual	371.115	1124	2080.914
	Monolingual	-371.115	768	1402.198
<i>p-value</i>	<i>1.03e-14</i>			
Speaker Sex				
	Female	254.701	1032	1967.578
	Male	-254.701	860	1610.808
<i>p-value</i>	<i>7.6e-08</i>			
Interaction –Following Segment Height:Syllable Stress				
	+High:Tonic	651.177	26	3622.931
	+Coronal:Atonic	432.775	469	1633.311
	-Coronal:Atonic	309.211	543	1218.696
	Zero:Tonic	137.115	47	4302.892
	-High:Atonic	46.306	190	1443.447
	-High:Tonic	-46.306	76	2364.888
	Zero:Atonic	-137.115	263	2914.886
	-Coronal:Tonic	-309.211	103	1696.33
	+Coronal:Tonic	-432.775	126	1855.095
	+High:Atonic	-651.177	49	1276.802
<i>p-value</i>	<i>9.33e-05</i>			
Following Segment Height				
	Zero/Pause	1382.147	310	3125.326
	+High	214.584	75	2090.127
	-High	-320.772	266	1706.716
	+Coronal	-497.107	595	1680.277
	-Coronal	-778.853	646	1294.851
Syllable Stress				
	Tonic	564.073	378	2340.285
	Atonic	-564.073	1514	1671.868