

Rbrul Manual

b

STUDENT ASSISTANTS LINGUISTICS
ENGLISH DEPARTMENT UNIVERSITY OF BERN

Rbrul Manual

Welcome To The Wonderful World Of Rbrul! Or So...

Rbrul is a script that is written in the programming language R. It is specifically designed to make statistical analysis of linguistic data fast and simple. Relatively fast and simple.

There are a few guides to Rbrul. For instance:

The official manual by Daniel Ezra Johnson (2010), the creator of Rbrul
http://www.danielezrajohnson.com/Rbrul_manual.html

A manual by Meredith Tamminga (2011), linked on the Rbrul website
http://www.danielezrajohnson.com/Getting_started_with_Rbrul.pdf

A very comprehensive guide by Agata Daleszynska, also linked on the Rbrul website
http://agatad.co.uk/images/Rbrul/rbrul%20handout_daleszynska.pdf

These and other guides you may find elsewhere are often great in terms of statistical explanations, therefore, if you struggle in the process of analysing your results, you may find answers in these guides. However, most of them do not provide detailed installation instructions or troubleshooting sections.

The Rbrul Manual of the Student Assistants Linguistics, English Department, University of Bern

This manual provides detailed instructions for the installation of both R and Rbrul (and even Rstudio), as well as an overview of a number of basic functions. Remember that there is also a troubleshooting section and a short glossary with explanations of the most important terms.

This manual has been put together - and is continuously enhanced - by the Student Assistants of the English Department, University of Bern. All persons involved are clearly more linguists than statisticians. Any error reports, feedback and suggestions for additions and enhancements are greatly appreciated. Please get in touch via the following email address:

sa-linguistics.ens@lists.unibe.ch

- thank you

Manual Contents

Installation

- 0.1 Installation of R
- 0.2 Installation of Rbrul Packages

Basic Functions

- 1. Launching of R and Rbrul
 - 1.1 Launching of R
 - 1.2 The Rbrul source code
 - 1.3 The Rbrul launch command

2. Preparing of data

- 2.1 Data preparation in Excel
- 2.2 Saving format

3. Loading of data into Rbrul

- 3.1 The 'load data' menu
- 3.2 Specification of file type

4. Crosstabs

- 4.1 The 'crosstabs' menu
- 4.2 One-layer tab
- 4.3 Two-layer tab
- 4.4 Three-layer tab
- 4.5 Output format

5. Plotting

- 5.1 The 'plotting' menu
- 5.2 Creation of visual representations

6. Modeling

- 6.1 The 'modeling' menu
- 6.2 Preparing a test on effects and significances
- 6.3 Running a test on effects and significances

7. Adjusting data

- 7.1 The 'adjust data' menu
- 7.2 Excluding of variable response(s)
- 7.3 Recoding of variable response(s)

8. Saving data

- 8.1 The 'load/save' menu
- 8.2 Saving of current data

9. Grouping variables and variable responses

- 9.1 Grouping in Excel
- 9.2 Grouping in Rbrul

10. Exporting Data to Excel

Troubleshooting

- o Launch Rbrul in RStudio
- o Alternative launches
- o Scatterplots are not displayed
- o Crashes while modeling & scatterplots are not displayed contd.
- o Normal shut down
- o Force shut down
- o Getting lost / Running wrong commands
- o Step-up and step-down mismatch
- o Error messages in step-up/step-down models
- o Results are not displayed in the preferred order
- o Change number categories in Excel

Glossary

- o Variable
- o Variable response/variant
- o Dependent variable/DV
- o Independent variable/IV
- o Random vs fixed effect
- o Step-up/step-down model
- o Factor weight
- o R²
- o Intercept
- o Significance

Installation

- 0.1 Installation of R
 - 0.1.1 Visit the following web site:
<https://cran.r-project.org/>
 - 0.1.2 Install the newest version of R with default settings

- 0.2 Installation of Rbrul Packages
 - 0.2.1 Open R
 - 0.2.2 Go to the "Packages" menu
 - 0.2.3 Click on "Install package(s)" (Windows) / "Package installer" (Mac)
 - 0.2.4 Click on "Get list" (Mac only)
 - 0.2.5 Select cran mirror (Switzerland)
 - 0.2.6 Tick the box for "Install Dependencies" (Mac only)
 - 0.2.7 Enter the following code to install essential packages:

```
> install.packages(c("boot", "car", "Hmisc", "lattice", "lme4", "Matrix", "MuMIn"), dependencies = TRUE)
```

Basic Functions

1. Launching of R and Rbrul
 - 1.1 Launch R
 - 1.2 Enter the Rbrul source code:

```
> source("http://www.danielezrajohanson.com/Rbrul.R")
```
 - 1.3 Enter the Rbrul launch command:

```
> rbrul()
```

Note: When Rbrul is launched for the first time, it will install necessary components. This may take a while. Rbrul has successfully launched when the Main Menu appears:

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

If the Main Menu does not show up, go to "Troubleshooting\Alternative launches"
If Rbrul is launched in Rstudio go to "Troubleshooting\Launch Rbrul in RStudio"
2. Preparing of data
 - 2.1 Use Excel to prepare data before loading it into Rbrul.
Make sure that the data file does not include unpermissible entries or typos
 - 2.2 Save the Excel file
(Recommendation: Save as Text (Tab delimited) file.)
3. Loading of data into Rbrul
 - 3.1

```
> 1
```

 Command: load data
 - 3.2

```
> t
```

 Command: file is tab delimited
 - 3.2.1 Select the file from the folder
Rbrul has successfully loaded data when the following appears:

```
Current data file is: [file location/file]
```



```
Current data structure:
```



```
Total tokens: [####]
```



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

4. Crosstabs

Note: Examples contain random numbers, counts and percentages do not match.

4.1 `> 4` Command: `crosstabs`

4.2 `columns` Define the variable for the columns
`columns`

<i>variable: st.</i>	
stand	non-st
501	323

4.3 `rows` (optional) Define the variable for the rows
`columns x rows`

		<i>variable: st.</i>	
		stand	non-st
<i>variable:</i>	female	123	45
<i>sex</i>	male	67	89

4.4 `pages` (optional) Define whether there should be several cross tabs representing a third variable
`columns X rows X pages`
`variable: age`

<i>young</i>		<i>variable: st.</i>	
		stand	non-st
<i>variable:</i>	female	27	13
<i>sex</i>	male	14	2

<i>old</i>		<i>variable: st.</i>	
		stand	non-st
<i>variable:</i>	female	38	4
<i>sex</i>	male	15	16

4.5 `cells` Select output format (Enter for counts, > 1 for proportions/means)
 counts: see above
 proportions/means:

binary variables:

		<i>variable: st.</i>	
		stand	non-st
<i>variable:</i>	female	63.3	15.2
<i>sex</i>	male	57.4	21.9

cells = proportions for e.g. young speakers vs other age groups
 (i.e. variable: age)

continuous variables:

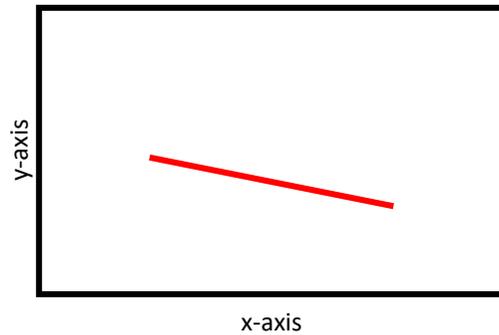
		<i>variable: st.</i>	
		stand	non-st
<i>variable:</i>	female	422.3	440.9
<i>sex</i>	male	401.7	419.1

cells = means for e.g. first formant frequencies (F1)
 (i.e. variable: age)

5. Plotting

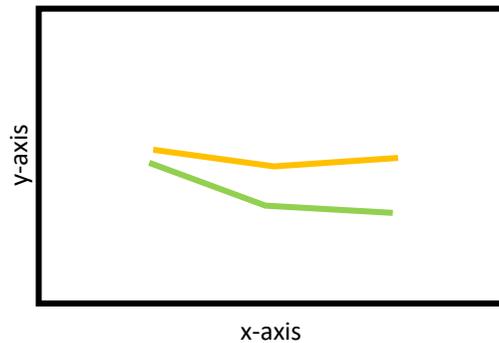
Note: The aspect ratio of the scatterplots are dependent on the window size and of the screen size.
In RStudio, use the 'Export' function to manually specify the ratio.

```
5.1 > 6 Command: plotting
5.2 > 1 Command: custom scatterplot
5.2.2 y-axis Define what the y-axis represents, e.g. DV standard/non-standard
5.2.3 x-axis Define what the x-axis represents, e.g. IV sex
```



This graph could represent how many tokens females and males (IV 'sex' plotted on the x-axis) realise in the standard way (DV 'st' plotted on y-axis)

5.2.4 Color separation (optional) Define which variable is to be plotted by means of different colors, e.g. IV age

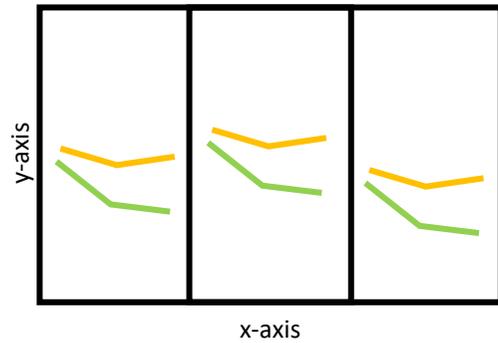


This graph could represent how many tokens young and old (IV 'age', color) females and males (IV 'sex', x-axis) realise in the standard way (DV 'st', y-axis)

5.2.5 Averages (optional) Indicate whether the graph should contain a black average line

5.2.6 Horizontal panels and/or vertical panels

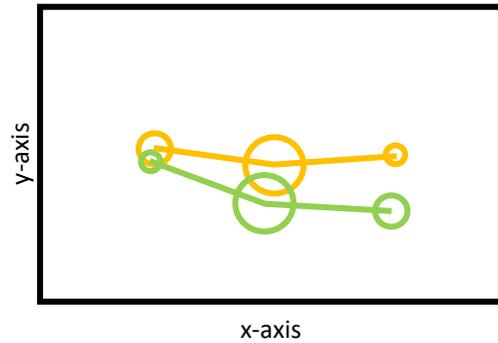
(optional) Indicate whether there should be different panels for different responses of further IVs



This graph could represent how many tokens young and old (IV 'age', color) females and males (IV 'sex', x-axis), who speak a different dialect (IV 'dialect', panels), realise in the standard way (DV 'st', y-axis)

5.2.7 Type of points

Recommendation: Enter for mean points



This creates circles on the graph that represent the number of tokens for a given category - the bigger the circle, the more data it represents

5.2.8 Scaling of points

Recommendation: Enter for 1

5.2.9 Type of lines

Recommendation: Enter for mean lines

5.2.10 Reference line

Recommendation: Enter for none

6. Modeling

- 6.1 > 5 Command: modeling
6.2 > 1 Command: choose variables
- 6.2.1 DV Define the DV
- 6.2.2 Type of response
continuous In cases where the DV has an infinite number of numerical values (e.g. frequencies: all values between 0 and 20'000 Hz possible)
binary In cases where the DV only has fixed values (e.g. yes/no, standard/non-standard, 1/2/3, a/b/c, etc.)
- (continued with binary DV)
- 6.2.3 Application value(s) Which response of the DV is of interest?
Should Rbrul count instances of e.g. "standard" and do calculations for this response, or of "non-standard"?
Note: These selections may be taken over into further calculations. Reset Rbrul in that case.
- 6.2.4 IV(s) Define the IV(s)
(Recommendation: in order to have a good overview of the results, use as few IVs as necessary - especially beginners of Rbrul.)
- 6.2.5 Continuous IV(s) Define those IV(s) that have an infinite number of numerical values
- 6.2.6 Interaction(s) (optional) Define whether certain IVs are to be paired up (e.g. How does age AND sex influence the realisation of X?)
- 6.2.7 Random effects Define those IV(s) that are characterised by significant heterogeneity in terms of the number of occurrences of their responses
E.g. 1: 540 tokens for speaker A, 660 tokens for speaker B - relatively similar number of occurrences, i.e. fixed effect;
E.g. 2: 1100 tokens for speaker A, 100 tokens for speaker B - relatively dissimilar number of occurrences, i.e. random effect.

The variables have been defined successfully when the following appears:

Current variables are:

.

MODELING MENU

1-choose variables 2-one-level 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

6.3 > 5 Command: step-up/step-down model
Rbrul is running the calculations as specified. This may take a while.
Rbrul has successfully run its analysis when the following appears:

```
STEPPING UP...
.
.
.
.
.
STEP-UP AND STEP-DOWN MATCH!*

Current variables are:
response.binary: glottalisation (glottalized vs. not glottalized)
fixed.factor: sex age

MODELING MENU
1-choose variables 2-one-level 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
```

* see Troubleshooting sheet in case MISMATCHES or ERRORS occur.

6.3.1 Binary DV: Analysis of results

6.3.1.1 Locate the very top of the step-up/step-down model, marked by the last command entry: > 5
This is the beginning of Rbrul's calculations.

6.3.1.1.1 Locate the following:
BEST STEP-UP MODEL IS WITH ... (0.00143)
[p-values building from null model]

```
$(IV)
factor      logodds      tokens      [DV responses]      centered factor weight
[IV resp.]      0.012      345      0.910      0.789
[IV resp.]      -0.012      678      0.123      0.211
...
```

```
$misc
...
```

STEPPING DOWN...

STEP 0 - Run 0 - full model with

```
$(IV)
factor      logodds      tokens      [DV responses]      centered factor weight
[IV resp.]      0.012      345      0.910      0.789
[IV resp.]      -0.012      678      0.123      0.211
...
```

This is the summary of the stepping-up calculations, followed by the stepping-down calculations.
For the interpretation of the step-up/step-down calculations, this area is where further analysis should be started
(Recommendation: Copy, paste and save such blocks, in a word or excel file, and add comments about what is being calculated in each block.)

6.3.1.3 Interpretation (binary)

Check the Glossary for a more detailed description of the expressions used in this section.

tokens 345/678
Number of tokens for an IV response of an IV, e.g. IV sex: female, male

DV responses 0.910/0.090
Percentage of tokens of an IV response occurring with a specified DV response, e.g. female and standard, male and standard
In the example above, 91% of 345 tokens are realised as X (X = possible DV response, e.g. standard)

factor weight* 0.789/0.211
Does the IV response, e.g. female, have a favouring or disfavouring effect on the specified DV response?
How strong is this effect?
Key value: 0.5 (< 0.5: disfavouring; > 0.5: favouring)

p* 0.00143
The significance p indicates whether this favouring or disfavouring effect is accidental or the manifestation of a clear pattern
(see Glossary entry)
Key values:
0.05 (< 0.05: significant; > 0.05: not significant)
0.01 (< 0.01: significant; > 0.01: not significant)
0.001 (< 0.001: significant; > 0.001: not significant)
(e = 10^x, e.g. 3.46e-15 = 3.46 * 10⁻¹⁵ = 0.000000000000000346)

* Factor weights of significant IVs should be taken from the step-up summary or from the individual step-up calculations
Factor weights of non-significant IVs should be taken from the first iteration of the step-down calculations

6.3.2 Continuous DV: Analysis of results

6.3.2.1 Locate the very top of the step-up/step-down model, marked by the last command entry: > 5

This is the beginning of Rbrul's calculations.

6.3.2.2 Locate individual IV calculations:

Trying with [IV]...

```
$(IV)
factor      coef      tokens      mean
[IV resp.]  202.827      99      1729.505
[IV resp.]  163.793      68      1690.471
...
```

```
$(misc.1)
n      df      intercept      overall mean
1172   8      1526.678      1504.713
```

```
$(misc.2)
deviance      AIC      AICc      R2
168000194     17259.16   17259.31   0.068
```

Run X (above) with [IV] is better/worse than Run X-1 without [IV], p = 3.46e-15
...

(Recommendation: Copy, paste and save such blocks, in a word or excel file, and add comments about what is being calculated in each block.)

6.3.2.3 Interpretation (continuous)

Check the Glossary for a more detailed description of the expressions used in this section.

tokens 99/68
Number of tokens for an IV response of an IV, e.g. IV sex: female, male

overall mean 1504.713
The mean of all DV values (numerical) in the entire data set, e.g. DV formant frequency: 1403.873, 1694.551, 1852.309, 1553.078 ...

mean 1729.505/1690.471
The mean of the DV values for a given IV response, e.g. IV speaker: speaker A mean 1729.505, speaker B mean 1690.471, ...

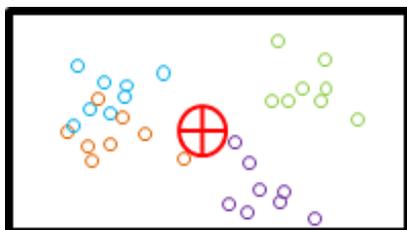
intercept 1526.678
The mean of the means of the DV values within a given IV response, e.g. the mean of 1729.505, 1690.471, ...

coef 202.827/163.793
The deviation of a given IV response mean from the intercept
e.g. mean formant frequency for speaker A = intercept + coef(speaker A) = 1526.678 + 202.827 = 1729.505

R² 0.068
Multiple R-squared
Do the IV(s), e.g. speaker, etc., have a strong or weak effect on the DV?
How much of the variance between DV values can be explained due to IV effect?
In the example above, only 6.8% of variance between DV values (e.g. formant frequency) is due to the effect of all IVs in the calculation (e.g. speaker, etc.).
Individual IV effects can be taken from the individual step-up calculations

p 3.46e-15
The significance p indicates whether the effect of the IV on the DV is accidental or the manifestation of a clear pattern (see Glossary entry)
Key values:
0.05 (< 0.05: significant; > 0.05: not significant)
0.01 (< 0.01: significant; > 0.01: not significant)
0.001 (< 0.001: significant; > 0.001: not significant)
(e = 10^x, e.g. 3.46e-15 = 3.46 * 10⁻¹⁵ = 0.00000000000000346)

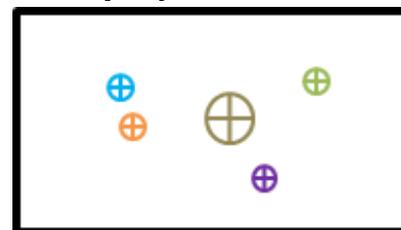
overall mean (red):



means (crosshairs):



intercept (golden):



7. Adjusting data

- 7.1 > 2 Command: adjust data
- 7.2 > 3 Command: exclude
 - 7.2.1 Factor group Select the variable containing the variable response(s) to be excluded
 - 7.2.2 Factor(s) Select the variable response(s) to be excluded

- 7.3 > 5 Command: recode
 - 7.3.1 Factor group Select the variable containing the variable response(s) to be recoded
 - 7.3.2 Factor(s) Select the variable response(s) to be recoded
 - 7.3.3 Recode as Name the new variable or variable response
 - 7.3.4 Factor(s) Select other variable (response(s) to be recoded, if applicable
 - 7.3.5 Recode to new column
 - Yes In case a new variable is to be created; name the new variable (the original variable response(s) remain unchanged)
 - No In case a new variable response is to be created (the original variable response(s) will be replaced)

8. Saving data

- 8.1 > 1 Command: load/save
- 8.2 Save current data
 - Yes In case a new file is to be created; name the new file (the original file remains unchanged)
 - No In case new data is to be loaded (the changes applied to the data loaded in Rbrul will be lost)
- 8.3 Load new data
 - Specify file type In case new data is to be loaded
 - Enter In case the current data is to be kept

9. Grouping variables and variable responses

- 9.1 Grouping in Excel
 - Advantage: making changes in the Excel master file allows for a better overview of what is being done with the data before any Rbrul calculations
- 9.1.1 Add a new column on the right of the table
- 9.1.2 Enter the concatenate formula (in german: verkettten) in the first cell that needs a new entry:
 - Either: =concatenate([columnX, row#],[columnY, row#], ...) e.g. =concatenate(A1, A2, A3, A4)
 - Or: =[columnX, row#]&[columnY, row#]& ... e.g. =(A1&A2&A3)
 - This will simply string the entries of the columns (2 or more) together.
- 9.1.3 Apply the same formula to the entire new column
 - Click on the cell where the formula is entered, click at the bottom right corner and pull it all the way down to the end of the table
- 9.1.4 Rbrul analysis
 - Save the new Excel file - Text (tab delimited) - and load it into Rbrul.

9.2 Grouping in Rbrul
 Advantage: changes can be applied without having to go back to the Excel master file

9.2.1 Grouping in the Modeling Menu
 Purpose: testing of combined occurrence without changing the data set in Rbrul
 Disadvantage: this method does not yield p-values for individual combinations

9.2.1.1 Go to the Modeling Menu

9.2.1.2 > 5 Command: modeling

9.2.1.3 > 1 Command: choose variables

9.2.1.4 DV Define the DV

9.2.1.5 Type of response Is the DV binary or continuous?

9.2.1.6 Application value(s) What is the realisation of the DV that is of interest?

9.2.1.7 IV(s) Define the IV(s)
 Choose the two variables that are of interest for their combined occurrence, e.g. age and sex

9.2.1.8 Continuous IV(s) Define those IV(s) that have an infinite number of numerical values

9.2.1.9 Pairwise interaction Choose the two variables that are of interest for their combined occurrence, e.g. age and sex

9.2.1.10 Random effects Define those IV(s) that are characterised by significant heterogeneity in terms of the number of occurrences of their responses
 E.g. 1: 540 tokens for speaker A, 660 tokens for speaker B - relatively similar number of occurrences, i.e. fixed effect
 E.g. 2: 1100 tokens for speaker A, 100 tokens for speaker B - relatively dissimilar number of occurrences, i.e. random effect

9.2.1.11 > 5 Command: step-up/step-down model
 Locate the calculations for the combined effect of the two chosen variables.
 It is recommended to focus on the first iteration of the step-down calculations.

9.2.2 Creating full interaction groups
 Purpose: analysis of all possible combinations of variable responses of two variables

9.2.2.1 Go to the Main Menu

9.2.2.2 > 2 Command: adjust data

9.2.2.3 > 10 Command: make full interaction group

9.2.2.4 Choose two factors Define which two variables are to be grouped together to form a new variable, e.g. age and sex
 Rbrul has successfully created a new variable when the Main Menu shows the following:

```

Current data file is: [file location/file]

Current data structure:
.
.
.
[new variable]
Total tokens: [####]

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

```

More than two variables can be grouped together, e.g. age, sex and dialect, by firstly creating a full interaction group of age and sex, and then by creating a full interaction group with the newly created variable (age:sex) and dialect (age:sex:dialect)

9.2.2.5 Rbrul analysis

It is now possible to analyse this new variable, which is a combinations of other variables.

9.2.3 Creating partial interaction groups

Purpose: analysis of specific combinations of two variables, especially useful for variables with many variable responses

9.2.3.1 Go to the Main Menu

9.2.3.2 > 2 Command: adjust data

9.2.3.3 > 11 Command: make partial interaction group

9.2.3.4 Choose two factors Define which from which two variables responses are to be grouped together to form a new variable, e.g. age and sex

9.2.3.5 Choose factor values Define which variable responses of the chosen two variables are to be grouped together, e.g. 20, 21, 22 and female
Rbrul has successfully created a new variable when the Main Menu shows the following:

```
Current data file is: [file location/file]
```

```
Current data structure:
```

```
.  
. .  
[new variable]  
Total tokens: [####]
```

```
MAIN MENU
```

```
1-load/save data
```

```
9-reset 0-exit
```

```
1:
```

9.2.4.6 Rbrul analysis

It is now possible to analyse this new variable, which is a combination of other variable responses.

10. Exporting Data to Excel

10.1 Select and copy data

10.2 Open Notepad (pre-installed on your computer)

10.3 Paste the data into Notepad

10.4 Select and copy the data in Notepad again

10.5 Open Excel

10.6 In the 'Home' tab, locate the 'Clipboard' section, and click on the arrow below 'Paste'

10.7 Click on 'Use Text Import Wizard'

10.8 When specifying the type of data to be pasted into Excel, chose 'fixed width' and finish the data import

It is also possible to chose 'delimited', specify the delimiter (space, tab, commas, whatever is applicable) and then finish the data import.

Excel then imports the original Rbrul data by nicely distributing them into separate cells

It is now possible to costumise these data: change fonts, cell sizes, create tables and graphs, etc.

If imported data is not computed correctly by Excel for further calculations or creations of graphs, go to "Troubleshooting\Change number categories in Excel"

Troubleshooting

Rbrul has a tendency to crash - some times for no apparent reason. In most cases, a reboot is necessary - which is why it is convenient to save the source code and have it handy at all times when working with Rbrul (see Basic Functions, 1.2.1).

Launch Rbrul in RStudio

```
RSt 1  Launch RStudio
RSt 2  Select "Open File" in the Files menu
RSt 3  Enter the Rbrul source code
      > source("http://www.danielezrajohanson.com/Rbrul.R")
RSt 4  Enter the Rbrul launch command
      > rbrul()
      Rbrul has successfully launched when the Main Menu appears.
```

Alternative launches

If the Main Menu does not show up after the source code and Rbrul launch command have been entered, try the following alternative launches - in this order.

```
AL 1.1 Close and Reopen R
AL 1.2 Enter the source code of the latest Rbrul version:
      > source("http://www.danielezrajohanson.com/Rbrul.R")
AL 1.3 Enter the Rbrul launch command:
      > rbrul()
      Rbrul has successfully launched when the Main Menu appears.
```

If it does, Rbrul has to be started with AL 1 all the time now: always use this alternative launch code.

It is possible that some functionalities are limited (e.g. The displaying multi-level scatterplots may not work). However, it is likely that Rbrul generates an error message requesting the download of the newest version of R. Proceed to AL 2.

```
AL 2.1 Close and Reopen R
AL 2.2 Download the .RData file from the website of the English Department, University of Bern:
      http://ens.unibe.ch/content/studies/corpus\_computer/index\_eng.html
```

```
AL 2.3 Go to the 'File' menu and click on 'Load Workspace'
```

```
AL 2.4 Select the .RData file
```

Note: By default, Mac operating systems do not show this type of files. If you want the .RData file to be displayed, follow the steps laid out in this guide:

```
http://macs.about.com/od/tipstricks/qt/hiddenfolder.htm
```

```
AL 2.5 Enter the Rbrul launch command:
```

```
> rbrul()
Rbrul has successfully launched when the Main Menu appears.
```

If it does, Rbrul has to be started with AL 2 all the time now: always use this alternative launch code.

It is possible that some functionalities are limited (e.g. The displaying multi-level scatterplots may not work). However, it is likely that Rbrul generates an error message requesting the download of the newest version of R. Proceed to AL 3.

AL 3.1 Close and Reopen R

AL 3.2 Enter the source code of an older Rbrul version, e.g.:

```
> source("http://www.danielezrajohanson.com/Rbrul_215.R")
```

or:

```
> source("http://www.danielezrajohanson.com/Rbrul_216.R")
```

AL 3.3 Enter the Rbrul launch command:

```
> rbrul()
```

Rbrul has successfully launched when the Main Menu appears.

If it does, Rbrul has to be started with AL 2 all the time now: always use this alternative launch code.

It is possible that some functionalities are limited (e.g. The displaying multi-level scatterplots may not work).

However, it is likely that Rbrul generates an error message requesting the download of the newest version of R. Proceed to AL 4.

AL 4.1 Close R

AL 4.2 Uninstall R

AL 4.2.W Windows users:

AL 4.2.W.1 Locate R in the Control Panel, uninstall it AND search the computer for any remaining folders and files (in most cases, there will be quite a few).

AL 4.2.W.2 Delete everything.

AL 4.2.M Mac users:

AL 4.2.M.1 Go to the location where R is saved and delete it from there.

Alternatively, install another version of R (step AL 4.3) without deleting the old one; follow the steps outlined here:

<https://support.rstudio.com/hc/en-us/articles/200486138-Using-Different-Versions-of-R>

Note: This leaves behind quite a number of files that are normally invisible.

If AL3 does not work the first time, try to find and delete those remainders by following the steps laid out in this guide:

<http://macs.about.com/od/tipstricks/qt/hiddenfolder.htm>

AL 4.3 Install the newest version of R

AL 4.3.1 Visit the following web site:

<http://stat.ethz.ch/CRAN/>

AL 4.3.2 Chose the correct Operating System and install the newest version of R with default settings

AL 4.4 Install the Rbrul packages according to the Installation guide

AL 4.5 Launch R with the normal source code or with one of the above Alternative launches

Scatterplots are not displayed

There might be display problems in the R versions between versions 2.15.3 and 3.2.2; the installation of a previous version (like R 2.15.3) may help.

ScPl 1 Uninstall the current R version 3.0.0 or higher

ScPl 2.W Windows users:

ScPl 2.W.1 Locate R in the Control Panel, uninstall it AND search the computer for any remaining folders and files (in most cases, there will be quite a few).

ScPl 2.W.2 Delete everything.

ScPl 2.M Mac users:

ScPl 2.M.1 Go to the location where R is saved and delete it from there.

Alternatively, install another version of R (step ScPl3) without deleting the old one; follow the steps outlined here:

<https://support.rstudio.com/hc/en-us/articles/200486138-Using-Different-Versions-of-R>

Note: This leaves behind quite a number of files, that are normally invisible.

If that does not work the first time, try to find and delete those remainders by following the steps laid out in this guide:

<http://macs.about.com/od/tipstricks/qt/hiddenfolder.htm>

ScPl 3 Install another R version (like 2.15.3).
ScPl 3.W Windows users:
ScPl 3.W.1 Visit the following web site:
ScPl 3.W.2 <http://cran.r-project.org/bin/windows/base/old>
ScPl 3.W.3 Choose an older version (like 2.15.3)
ScPl 3.W.4

ScPl 3.M Mac users:
ScPl 3.M.1 Visit the following web site:
ScPl 3.M.2 <http://cran.r-project.org/bin/macosx/old>
ScPl 3.M.3 Choose the .pkg file of an older version (like 2.15.3)
ScPl 3.M.4 Install it with default settings
ScPl 4 If ScPl 1-3 did not resolve the problem, read the following section on 'Crashes while modeling & scatterplots are not displayed contd.'

Crashes while modeling & scatterplots are not displayed contd.

Cr1 Reset R to factory defaults using the following command:
 `> system("defaults write org.R-project.R force.LANGen_US.UTF-8")`
If problems occur, manually also uninstall all packages from the library folder.
Cr2 Re-install these packages:
 `> install.packages(c("boot", "Hmisc", "lattice", "lme4", "Matrix"), dependencies = TRUE)`
Windows users:
 Pay attention to the messages displayed in R's main window during the installation process.
 The installation of a package may fail because another one needs to be installed before.
 Install all of these required packages, too.
Mac users:
 Make sure that the box for "Install Dependencies" is ticked for the installation of all the listed packages.

Normal shut down

`> 0` Command: Exit Rbrul
Then, R can be closed
An attempt to close R without having quit Rbrul may cause the following message to appear:
1: q()

Force shut down

Shutting Rbrul down when it crashed is mostly only possible via the Task Manager:
Ctrl + Alt + Del
End Task: R

Getting lost / Running wrong commands

Rbrul cannot go back to the Menu structure at all times.
Once a command for a specific process is entered (e.g. 1 - [choose variables](#)), the entire process needs to be completed before returning to the Menu is possible.
Attempt to complete processes or restart R and Rbrul after a force shut down.

Step-up and step-down mismatch

These error messages occur mostly in cases where Rbrul is to process too large calculations. Mismatches are calculation errors by Rbrul.
Break the data analysis into smaller parts, e.g. run Step-up/step-down models with individual IVs or small groups of IVs.

Error messages in step-up/step-down models

In some cases, error messages other than mismatches ("[1] - error in run X", or similar) may occur due to formatting of the Text (tab delimited) file.
When copy/pasting selected data from one excel file to another one which will be saved as Text (tab delimited) file and ultimately loaded into Rbrul, the pasting needs to be completed with the option "paste values only".

Results are not displayed in the preferred order

Rbrul displays results in the order of different numeric pieces of information (e.g. coefficients), or in alphabetical order (e.g. scatterplots).
Change the data (recommended: in excel) and create the preferred order with numbers or letters of the alphabet - and brackets.
E.g.: a) female, b) male; 1) young, 2) middle, 3) old

Change number categories in Excel

After importing data into Excel via the Text Import Wizard, it may occur that the data cannot be computed further, for calculations or the creations of graphs, etc.

This is most likely due to the number category.

Change it in the 'Home' tab, under the 'Number' section. Choose 'Number' here.

In addition, it may be necessary to replace the current comma symbol (",") with a different one to allow for such a conversion of category.

Select the data in question, search for (Windows: CTRL + H; Mac: SHIFT+CMD + H) the current comma symbol (","), and replace it with a different one (most likely: ";").

Glossary

Variable	<p>Any piece of information given for each token E.g. what is said; who is the speaker; how old is the speaker; what sex is the speaker; in which linguistic setting does the utterance occur; etc. By definition, the variable has variable values, i.e. there several possibilities of what is said, several speakers, several ages, etc. - the different values are variable responses or variants</p>
Variable response/variant	<p>Different values of a variable E.g. there several possibilities of what is said, several speakers, several ages, etc.</p>
Dependent variable/DV	<p>The feature in the centre of attention. How does X (the DV) behave in different contexts? How is it realised in these contexts? (context: the speaker of X is male/female; the speaker of X young/old; X occurs before/after Y; etc. - those are IVs)</p>
Independent variable/IV	<p>Context variable, e.g. sex, age, linguistic setting etc.</p>
Random vs fixed effect	<p>A variable that is characterised by significant heterogeneity in terms of the number of occurrences of its responses is to be treated as random effect A variable that is not characterised by significant heterogeneity in terms of the number of occurrences of its responses is to be treated as fixed effect E.g. 1: 540 tokens for speaker A, 660 tokens for speaker B - relatively similar number of occurrences, i.e. fixed effect; E.g. 2: 1100 tokens for speaker A, 100 tokens for speaker B - significantly dissimilar number of occurrences, i.e. random effect.</p>
Step-up/step-down model	<p>Calculations of: token numbers percentages factor weights significances etc.</p>

Factor weight

Is a certain IV response favouring or disfavouring a certain realisation of the DV?

How strong is this favouring/disfavouring effect?

Key value: 0.5 (< 0.5: disfavouring; > 0.5: favouring)

It is not the percentage that allows such a conclusion!

Percentages can be way below 50% ("The number of tokens realised with the DV response X is lower than that of other DV responses."), but the effect of the IV can still be favouring.

E.g. A speaker may only very rarely add a [g] to a word ending in <ing> (orthographic script), resulting in the pronunciation [ɪŋg], e.g. [sɪŋɪŋg] - DV response -, but it may be that the linguistic environment (IV) "the following sound is a vowel" (IV response) invites the speaker to make this addition much more than any other IV.

Fact I: there is a number of tokens that the speaker realises with a [g].

Fact II: the number of these tokens is small.

Conclusion I: most IVs disfavour the realisation of [ɪŋg].

Conclusion II: some IV(s) promote, i.e. favour, the realisation of [ɪŋg]. If no IV did so, there would not be any tokens at all with the DV response [ɪŋg].

Of the tokens that fall into the category of IV response "the following sound is a vowel", the majority may be realised without the addition of a [g], but the factor weight that Rbrul calculates is above 0.5, telling us that this is a/the IV that favours the realisation of [ɪŋg].

R²

Multiple R-squared

Does the IV, e.g. speaker, have a strong or weak effect on the DV?

How much of the variance between DV values can be explained due to IV effect?

Unlike for binary DVs, where it is investigated whether IV responses favour or disfavour certain DV responses, the question for continuous DVs is how much a certain IV can explain the variance between the DV values:

Does an IV trigger DV values that deviate little or much from the intercept?

Small R² values indicate that a specific IV causes little deviation,

high R² values indicate that a specific IV causes much deviation.

Intercept

The mean of the means of the DV values within given IV responses.

E.g. The effect of the IV speaker on the DV formant frequency is investigated.

Every speaker (IV response) is likely to end up with different DV values and thus a different DV mean value.

The intercept is the calculation of the mean of every speaker's individual mean DV value:

mean formant frequency for speaker A: 100

mean formant frequency for speaker B: 200

mean formant frequency for speaker C: 300

intercept: $(100 + 200 + 300) / 3 = 600 / 3 = 200$

Significance

Is this favouring or disfavouring effect accidental or the manifestation of a clear pattern?

Key values:

0.05 (< 0.05: significant; > 0.05: not significant)

0.01 (< 0.01: significant; > 0.01: not significant)

0.001 (< 0.001: significant; > 0.001: not significant)

(e = 10^x, e.g. 3.46e-15 = 3.46 * 10⁻¹⁵ = 0.00000000000000346)

Statistical significance is a means to assess how much confidence researchers can have in their findings.

I.e. it is a means of evaluation of findings.

Colqohoun ([2014](#)) suggests, that a significance threshold of 0.001 be set. With a threshold of 0.05, there is at least a 30% likelihood that conclusions about effects are wrong! Colqohoun therefore advices:

- to set a threshold of 0.001, which reduces the likelihood of false discoveries to less than 5%;
- to treat findings with $p \sim 0.05$ merely as "worth another look".

Non-significance is often a consequence of a low token number in particular response variable (IV).

If token numbers are high, it is worth reporting that a particular response variable (IV) has no statistically significant effect on the dependent variable (DV) - this is a valid and important finding!

If token numbers are low, it is not possible to draw conclusions for a particular response variable (IV):

too low token numbers do not mean that a particular response variable does or does not exert an influence on the dependent variable (DV), it simply means that there is not enough data to make an assessment. In a particular context, it is worth pointing this out, and, with caution, make tentative statements.

E.g. "Low token numbers for IV X render an interpretation of its influence on the DV impossible. However, the factor weight seems to suggest that, if there was more data, XYZ effect may be apparent."

Significance tests:

likelihood-ratio Chi-squared test

for runs with binary DV responses (logistic regression)

F-test

for runs with continuous DV responses (linear regression)