Rcbc 0.2: CBC in R with Individual Utilities & Survey Mockups

Chris Chapman, Steven Ellis
cchapman@google.com, elliss@google.com

Contribution: Improved tools to work with CBC in R

Rcbc 0.2 lets analysts: (1) mock up CBC surveys easily; (2) simulate CBC designs and responses; (3) estimate aggregate and individual-level Hierarchical Bayes utilities easily; (4) import designs and responses from commercial CBC software (e.g., Sawtooth Software [3]) and do additional and parallel analyses in R.

Rcbc [1] is Open-Source Software, available from the authors under the GNU General Public License.

A Working CBC Mockup in 5 Lines of R

Rcbc 0.2 adds the ability to tag attributes with friendly names, and to write them to a CSV file as specified by the design matrix. This allows easy visualization of the CBC format, and for testing of the CBC using a typical spreadsheet program.

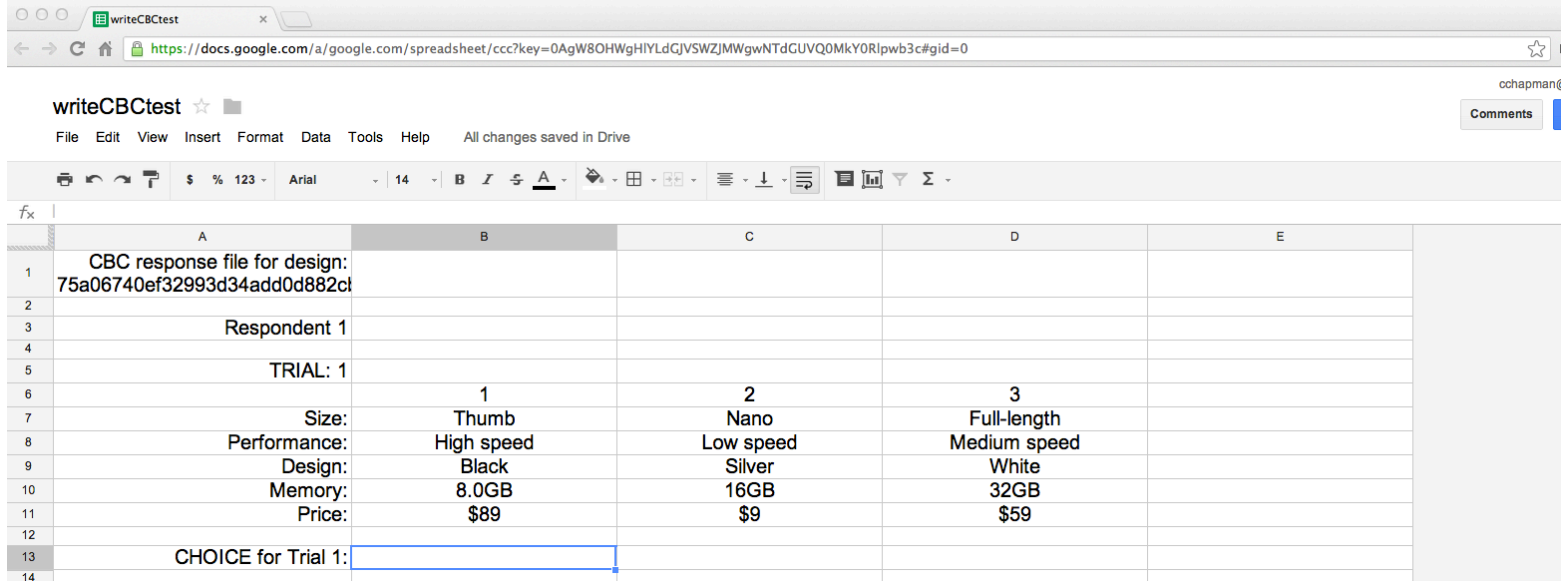
We demonstrate this using Google Spreadsheets, which allows simultaneous completion by multiple testers.

```
# This example imagines we're doing a "designer USB flash drive"
# Step 1: define the CBC
attr.list ← c(3, 3, 5, 5, 4) # defines CBC: 5 attributes, 3-5 levels each
tmp.tab ← generateMNLrandomTab(attr.list, resp=3, cards=3, trials=12) # design matrix

# Step 2: assign friendly names to the attributes and levels
attr.names ← c("Size", "Performance", "Design", "Memory", "Price")
attr.labels ← c( "Nano", "Thumb", "Full-length",
"Low speed", "Medium speed", "High speed",
"Tie-dye", "Silver", "Black", "White", "Prada",
"1.0GB", "8.0GB", "16GB", "32GB", "256GB",
"$9", "$29", "$59", "$89" )

# Step 3: Write the survey to a CSV
writeCBCdesignCSV(tmp.tab, attr.list=attr.list, lab.attrs=attr.names, lab.levels=attr.
    labels, filename="writeCBCtest.csv", delim=",")
```

If the resulting CSV is used to gather pilot data, the responses can be read and used to estimate utilities:

```
tmp.win ← readCBCchoices(tmp.tab, filename="writeCBCtest - Sheet 1.csv")
tmp.win.exp ← expandCBCwinners(tmp.win) # expand to Rcbc style
tmp.pws ← estimateMNLfromDesign(tmp.des, tmp.win.exp) # aggregate MNL utilities
```

Resulting Mockup: CBC in a Spreadsheet


writeCBCtest
https://docs.google.com/a/google.com/spreadsheet/ccc?key=0AgW8OHWgHIYLdGJVSWZJMWgwNTdGUVQ0MkY0RIpwb3c#gid=0
writeCBCtest
File Edit View Insert Format Data Tools Help
All changes saved in Drive
cchapman@
Comments
$ % 123 Arial 14 B I S A Σ
fx

	A	B	C	D	E
1	CBC response file for design: 75a06740ef32993d34add0d882ct				
2					
3	Respondent 1				
4					
5	TRIAL: 1				
6		1	2	3	
7	Size:	Thumb	Nano	Full-length	
8	Performance:	High speed	Low speed	Medium speed	
9	Design:	Black	Silver	White	
10	Memory:	8.0GB	16GB	32GB	
11	Price:	$89	$9	$59	
12					
13	CHOICE for Trial 1:				
14					

Individual-level HB Utilities Made Even Easier

Rcbc 0.2 makes Hierarchical Bayes estimation easy in R for CBC studies that have rectangular designs (each respondent has the same number of trials, and each trial has the same number of concepts). This is typical, for instance, in CBC designs from Sawtooth Software SSI/Web [3].

New functions in Rcbc 0.2 for working with HB estimation are:

`estimateMNLfromDesignHB()` : estimate utilities and saved draws from a given design file + responses
`extractHBbetas()` : get the individual-level mean betas from the above model

These functions use the R package `ChoiceModelR` [4] (which builds on and updates `bayesm` [2]). Even when Rcbc's assumptions don't fit a project, our code may be a starting point to work with `ChoiceModelR`.

Example: HB for CBC in 7 Lines of R

We assume that you have fielded a CBC study using Sawtooth Software SSI/Web, and saved the resulting "TAB" file with the Sawtooth-generated design and responses to `"MyCBCtabFileFromSawtooth.tab"`.

```
# Step 1: Import the data
tmp.raw ← read.csv("~/somedir/MyCBCtabFileFromSawtooth.tab") # load the data
tmp.tab ← tmp.raw[,15:24]) # get the design matrix from the relevant columns
tmp.attrs ← findSSIattrs(tmp.tab)# infer the CBC structure
tmp.win ← tmp.raw[,25] # get the winners from the relevant column

# Step 2: Estimate the HB model
tmp.logitHB ← estimateMNLfromDesignHB(tmp.tab, tmp.win, kCards=3, kTrials=8, kResp=200)

# Step 3: Get the aggregate mean beta utilities and individual-level mean betas
tmp.HBmeanbeta ← apply(tmp.logitHB$betadraw, 2, mean) # means across draws/respondents
tmp.HBindbetas ← extractHBbetas(tmp.logitHB, tmp.attrs) # mean of draws per respondent
```

Options include the MCMC chain length, number of draws saved, and the skip interval for saving draws.

Using a MacBook Pro (2011 15", 2Ghz i7, 8GB RAM, OSX 10.8.3, R 2.15.2) and simulated CBC data (5 attributes, 22 levels, N=200, 8 trials of 3 concepts), convergence of the HB model takes 20-30 seconds.

Limitations and Future Work

Rcbc is not a substitute for best-of-breed commercial software for CBC; it is a supplement.

Primary limitations:

- Only rectangular CBC designs
- Slower estimation than commercial software
- No checks on data quality; good data assumed

Future plans: (1) Do attribute impact estimation from HB models [1]. (2) Handle data from mockup surveys more robustly. (3) Refactor design structure to handle respondent IDs and meta-data in a smart way instead of assuming rectangular blocks.

References

[1] Chapman, C.N., and Alford, J.L. Rcbc: Choice-Based Conjoint Models in R. Poster presented at *Advanced Research Techniques Forum (ART Forum) 2010*. San Francisco, CA, 2010.

[2] Rossi, P. bayesm: Bayesian Inference for Marketing/ Microeconometrics. R package version 2.2-5. Available at `http://CRAN.R-project.org/package=bayesm`, 2012.

[3] Sawtooth Software. SSI/Web 8.2.2. Available at `http://www.sawtoothsoftware.com`. Orem, UT, 2013.

[4] Sermas, R. ChoiceModelR: Choice Modeling in R. R package version 1.2. Available at `http://CRAN.R-project.org/package=ChoiceModelR`, 2012.