DESIGN OF FACTORIAL SURVEY EXPERIMENTS IN STATA

Maurizio Pisati and Livia Ridolfi

maurizio.pisati@unimib.it
l.ridolfi@campus.unimib.it

Department of Sociology and Social Research
University of Milano-Bicocca (Italy)

2015 Italian Stata Users Group meeting
Firenze, November 12-13, 2015

OUTLINE

1. Factorial survey experiments
2. `fsdesign`
3. Example
4. Conclusions
5. References

FACTORIAL SURVEY EXPERIMENTS

DEFINITION

- Factorial survey experiments (FSEs) are
  - multifactorial experiments
  - conducted within sample surveys of individuals
  - based on manipulated descriptions of some objects of interest (people, institutions, policies, goods, services, events, ...)
- The basic purpose of a FSE is to investigate the judgment principles that underlie respondents' values, attitudes or preferences toward the objects of study (Auspurg and Hinz 2014; Jasso 2006; Rossi and Nock 1982)
- FSEs belong in the family of methods for stated preference analysis and, therefore, are closely related to such techniques as conjoint analysis, discrete choice experiments and best-worst scaling (Aizaki *et al.* 2015; Louviere *et al.* 2000)

BASIC FEATURES

- In a FSE, sample respondents are asked to carry out a task or sequence of tasks
- Within each task, every respondent is presented with a set of objects, each described by a given profile – i.e., by a given combination of levels of a predefined set of attributes (factors)
- Respondents are then asked to choose among, rank or rate the objects in the set
- By varying the profiles of the objects presented to respondents according to a given experimental design, it is possible to estimate the impact of each attribute and its levels on respondents' preferences

- In a study considering d attributes A_j ($j = 1, \ldots d$), each taking k_j levels, the size of the profile population is:

$$N_p = \prod_{j=1}^{d} k_j$$

- When the size of the profile population is smaller than the planned number of respondents n – i.e., $N_p < n$ – a full factorial design can be used
- Often, however, $N_p \gg n$, so that one is forced to use a fractional factorial design, i.e., a subset of the full factorial design

fsdesign

- `fsdesign` is a novel user-written Stata command for designing both full and fractional FSEs
- The user is first required to specify each attribute's name and levels, and optionally its label and randomization weights
- Four additional options are required:
  - the number of tasks per respondent
  - the number of profiles per task
  - the number of unique blocks in the design (a block is the whole set of profiles evaluated by a respondent)
  - the number of respondents (must be a multiple of the number of unique blocks)

- Optionally, the user can request that:
  - profile duplicates be avoided (within task, block, or design)
  - profile restrictions be imposed (typically, to avoid illogical profiles)
  - the generated design be saved to a Stata dataset
- For fractional designs, profiles are selected using simple random selection (alternative selection methods might be added in future versions)
- Although it is not guaranteed to always generate a “good” fractional design (Mee and Dean 2015), simple random selection is fast, easy to implement, and works well when the number of selected profiles is sufficiently large

`fsdesign` *attribute_definition* [`||` *attribute_definition*] [`||` *attribute_definition ...*], `tasks(#) profiles(#) blocks(#) respondents(#)` [`nodups(`*string*`) restrictions(`*string*`) saving(`*filename* [`, replace`]`)`]

where the syntax for *attribute_definition* is

`name(`*newvarname*`) levels(`*string*`)` [`label(`*string*`) rweights(`*string*`)`]

EXAMPLE

A study of women's mating preferences

- Between January and June 2015, we conducted an exploratory FSE aimed at analyzing Italian young women's mating preferences
- The study population was defined as Italian heterosexual women, aged 23-25, unmarried, living, working or studying in Milano (Italy), with a high-school diploma or a university degree
- The study sample ($n = 100$) was a quota sample selected using random criteria

A STUDY OF WOMEN'S MATING PREFERENCES

- The objects of evaluation were written descriptions of potential male partners, each defined by a combination of 11 three-level traits
- Given the small sample size, the 3^{11} full factorial design was clearly impracticable ($N_p = 177,147$)
- Thus we opted for a fractional design where each respondent was asked to evaluate a unique set of 12 profiles, so that altogether we tested a fraction of $100 \times 12 = 1,200$ profiles
- The evaluation task consisted in first ranking and then rating – on a $[0, 10]$ scale – the 12 profiles (ties allowed)

Stata code for generating the experimental design

```
set seed 432183764

fsdesign                                                              ///
   name(age)                                                          ///
   levels("25" "30" "35")                                             ///
   label("Age")                                                       ///
   ||                                                                 ///
   name(origin)                                                       ///
   levels("Lower class" "Middle class" "Upper class")                 ///
   label("Origin class")                                              ///
   ||                                                                 ///
   name(educ)                                                         ///
   levels("Middle school" "High school" "University degree")          ///
   label("Level of education")                                        ///
   ||                                                                 ///
   [...]
   ,                                                                  ///
   tasks(1) profiles(12) blocks(100) respondents(100)                 ///
   nodups(wd) saving("Design.dta", replace)
```

First 24 rows and selected columns of the `Design.dta` dataset

fs_rid	fs_bid	fs_tid	fs_pid	age	origin	educ
1	1	1	1	30	Middle class	University degree
1	1	1	2	30	Middle class	University degree
1	1	1	3	30	Lower class	Middle school
1	1	1	4	25	Middle class	Middle school
1	1	1	5	30	Upper class	High school
1	1	1	6	35	Lower class	High school
1	1	1	7	35	Upper class	High school
1	1	1	8	35	Lower class	University degree
1	1	1	9	35	Middle class	High school
1	1	1	10	30	Middle class	University degree
1	1	1	11	30	Lower class	University degree
1	1	1	12	25	Lower class	University degree
2	2	1	1	25	Upper class	University degree
2	2	1	2	35	Upper class	High school
2	2	1	3	35	Middle class	University degree
2	2	1	4	25	Upper class	Middle school
2	2	1	5	25	Middle class	High school
2	2	1	6	25	Middle class	High school
2	2	1	7	25	Middle class	High school
2	2	1	8	25	Upper class	High school
2	2	1	9	35	Middle class	High school
2	2	1	10	25	Lower class	University degree
2	2	1	11	25	Upper class	High school
2	2	1	12	25	Lower class	Middle school

- After generating the experimental design, we wrote a Stata do-file for composing – via LaTeX– the verbal descriptions of the 1,200 profiles
- The do-file takes the `Design.dta` dataset generated by `fsdesign` as input and outputs a PDF file containing the verbal descriptions of the 1,200 profiles, ready to be printed and submitted to respondents
- Profile descriptions were composed using between-respondent randomization of trait presentation order, so as to minimize carry-over and primacy effects (Auspurg and Hinz 2014).

Example profile

rid: 1 pid: 1

P.F. è circa 5 cm più basso di te e, secondo la maggior parte delle tue amiche, ha un aspetto fisico molto attraente. Tende a essere dominante con le altre persone, ma non lo è con la propria compagna; inoltre è ugualmente premuroso e leale con la propria compagna e con le altre persone. Per quanto riguarda la sua vita sentimentale, finora si è difficilmente lasciato coinvolgere in relazioni di qualsiasi tipo. Le sue credenze religiose sono molto simili alle tue, mentre la sua posizione politica è un po' diversa dalla tua. È laureato e, attualmente, svolge un lavoro di basso livello socioeconomico. È nato nel 1985 da una famiglia di classe media.

Ordine ☐☐ Voto ☐☐

CONCLUSIONS

CONCLUSIONS

- `fsdesign` is still at alpha stage:
  - option `restrictions(`*string*`)` has yet to be implemented
  - further options might be added
  - the help file has to be written
  - more testing needs to be done
- We are also working on a companion command (`fsdiag`) aimed at carrying out several kinds of diagnostics on the designs generated by `fsdesign`
- A longer-term project would be to extend `fsdesign`'s capabilities by adding alternative methods for generating fractional designs

REFERENCES

REFERENCES

- Aizaki, H., Nakatani, T. and Sato, K. (2015) *Stated Preference Methods Using R*, Boca Raton, FL, CRC Press.
- Auspurg, K. and Hinz, T. (2014) *Factorial Survey Experiments*, Thousand Oaks, CA, Sage.
- Jasso, G. (2006) "Factorial survey methods for studying beliefs and judgments", *Sociological Methods & Research*, 34, 3, pp. 334-423.
- Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000) *Stated Choice Methods: Analysis and Applications*, Cambridge, Cambridge University Press.
- Mee, R. and Dean, A. (2015) "Regular fractional factorial designs", in *Handbook of Design and Analysis of Experiments*, ed. by A. Dean, M. Morris, J. Stufken and D. Bingham, Boca Raton, FL, CRC Press, pp. 279-319.
- Rossi, P. H. and Nock, S. L. (eds.) (1982) *Measuring Social Judgments: The Factorial Survey Approach*, Thousand Oaks, CA, Sage.