Tabulation of Multiple Responses

Ben Jann

ETH Zurich, Sociology
ETH Zentrum SEW E 21
CH-8093 Zürich

jann@soz.gess.ethz.ch

Example of a multiple response question:

"Which narcotic substances did you consume during the last week?"

☞ *Tick all appropriate boxes*

- ☐ Cannabis
- ☐ Cocktail (cocaine/heroin)
- ☐ Heroin (alone)
- ☐ Cocaine (alone)
- ☐ Ecstasy
- ☐ Other substances: .....

  .....

  ..... ✍

Approaches to store multiple responses:

(Cox/Kohler 2003: Stata FAQ "How do I deal with multiple responses?")

- As a set of indicator or dummy variables, one for each possible answer.

id	d1_can	d2_mix	d3_her	d4_coc	d6_xtc	d5_oth
1	1	0	0	0	1	0
2	1	0	0	1	1	0
3	1	0	1	0	0	0
4	0	1	1	1	0	1
5	0	0	0	0	0	1

I will call this the **indicator mode** or **dichotomous mode**.

Approaches to store multiple responses:

(Cox/Kohler 2003: Stata FAQ "How do I deal with multiple responses?")

- As a set of polytomous variables, one for each response.

id	druguse1	druguse2	druguse3	druguse4
1	1	5	0	0
2	5	4	1	0
3	3	1	0	0
4	4	3	2	6
5	7	0	0	0

```
Labels:
1 "cannabis" 2 "cocktail"
3 "heroin" 4 "cocaine"
5 "ecstasy" 6 "other: morphine"
7 "other: LSD" ...
0 "no further"
```

id	druguse1	druguse2	druguse3	druguse4
1	cannabis	ecstasy		
2	ecstasy	cocaine	cannabis	
3	heroin	cannabis		
4	cocaine	heroin	cocktail	morphine
5	LSD			

I will call this the **polytomous mode**.

- Other approaches: ranked multiple responses, composite variables, long form

Approaches to analyze multiple responses:

- For example: Transform the data to indicator mode and use `tabstat`; there are also several user-written commands which might help in some cases.
- However, even simple descriptive analyses of multiple responses are usually a pain (especially if in polytomous mode).
- It would be nice to have a general and easy to use command for the tabulation of multiple responses (just like the official `tabulate` for "regular" variables).
- The new package `mrtab` tries to fill this gap (first released on SSC in January 2004, revised version to be published shortly).

Features of `mrtab`:

- One-way tables of frequency distribution of responses
- Two-way tables (tabulate a multiple response set against a "regular" variable)
- Row/column percents regarding to n. of responses and/or n. of respondents (or respondents with one response at least)
- Supported data structures:
  - indicator mode (also non-dichotomous)
  - polytomous mode: numeric or string; neglect repeated responses or count all
- Significance tests (two-way tables): overall χ^2- or LR-test, separate χ^2- or LR-tests for rows (multiple tests)

Features of `mrtab`:

- Display: sort responses in order of frequency, label response set, specify formats and max. width of response labels, turn on/off labels/names/values, turn on/off breaking wide tables, suppress freq. table, ...
- Misc: generate new indicator variables, fweights/aweights allowed, byable, if/in, exclude missings casewise, saves results in r(), ...