

## Remark

The files as described here are not always the files that are actually used in the various chapters. In many chapters one or more specific files are used to analyze the same data, for example using a different order for the items or with different design factors and additional external variables. However, the person-by-item data are always the same.

It will be explained in the part on the different chapters which file is used, and if it is not a file from those listed here, then the file in question is added.

## Example data set on verbal aggression

For a description of the original data, see Section 1.2 of Chapter 1.

The data stem from 316 persons who have responded to 24 items on a three-point scale (0, 1, 2).

The items are always ordered as indicated in Table 1.1.

The data are used in all Chapters but Chapter 4.

The data are available in four pairs of files, and they either have a txt extension or a xls extension.

### 1. Three-point scale data in vector format

- data verbal aggression vector poly.txt

- data verbal aggression vector poly.xls

In both these files, the item responses are ordered in one long vector. Each pair of a person and an item response defines a new row.

On each row one finds in this order:

- person ID
- observed response: 0, 1, or 2
- item indicator value (1 when it is a response to the corresponding item, 0 if not)
- Trait Anger score
- Gender (male=1, female=0).

The file has  $316$  (# respondents)  $\times$   $24$  (# items) =  $7,584$  rows, and  $1 + 1 + 24 + 2 = 28$  columns. The response is given in the second column.

### 2. Dichotomized data in vector format

- data verbal aggression vector dichot.txt

- data verbal aggression vector dichot.xls

Except for a dichotomization, the data are the same as in the previous two files. In order to dichotomize, 2-responses are transformed into 1-responses.

### 3. Three-point scale data in matrix format

- data verbal aggression matrix poly.txt

- data verbal aggression matrix poly.xls

In these two files, the original three-point scale item responses are ordered in a matrix, with one row per person. On each row, the responses of the person are preceded by an ID and followed by the Trait Anger score and the Gender value (male=1, female=0).

#### 4. Dichotomized data in matrix format

- data verbal aggression matrix dichot.txt
- data verbal aggression matrix dichot.xls

## CTB Mathematics and Science data set

For a description of the original data, see Section 5.6 of Chapter 5.

The data are used in Chapters 5 and 6.

The data stem from 1500 (from 35 schools) students who have responded to 76 items. Of these 56 are multiple-choice items (1 thru 31, and 42 to 66). The responses to these 56 items are coded with 0 (incorrect) and 1 (correct). The original responses are not given, and neither are the responses to the remaining items.

Binary Mathematics and Science data

- MASC is a SAS data set

The coded item responses are ordered in one long vector. Each pair of a person and an item response defines a new row, with the following elements, in this order.

- person ID
- school ID
- item number (not a dummy item indicator)
- coded response (1=correct, 0=incorrect)
- Gender (1=female, 0=male)
- Type (type of, 1=catholic, 2=other private, 3=public)
- Size (number of students in the school, in hundreds)
- Bachelor (transformed and standardized percentage of adults with BA degree or higher in area with school zip code, further indicated as the school area)
- BornInState (transformed and standardized percentage of adults in the school area who were born in the state where they now reside)
- Mortgage (transformed and standardized median of the monthly mortgage in the school area)
- OtherLanguage (transformed and standardized percentage of foreign language households in the school area)
- Math (1=mathematics item, 0=science item)
- item indicator value (for the items 1 to 31 and 42 to 66, 1 when it is a response to the corresponding item, 0 if not)

All zip coded derived variables come from the 1990 Census information.

The file has  $1500$  (# respondents)  $\times$   $56$  (# items) =  $84,000$  rows, and  $1 + 1 + 1 + 1 + 8 + 56$ . The response is given in the fourth column.

## **Guilt data set**

For a description of the original data, see Section 9.1 of Chapter 9.

The data are used in Chapter 9.

The data stem from 268 students (130 males, 138 females) who have responded to 40 binary items.

The guilt data set

- MIRID.sas7bdat and RWMIRID.sas7bdata are SAS data sets, and MIRID.txt versus RWMIRID.txt are text files which have a similar content as the two previously mentioned SAS data sets.

The item responses are ordered in one long vector. Each pair of a person and an item response defines a new row, with the following elements, in this order.

- person ID
- Gender, coded as 1 for males and 0 for females
- A: the discrimination value of the current item (for estimating the OPLM-variants)
- Y: The response
- Co: a dummy variable indicating whether the current item is a composite item (1) or a component item (0)
- 30 dummy variables denoting which component item is associated with the current item.

The file has 268 (# respondents) x 40 (# items) = 10,720 rows. The response is given in the fourth column.