Exercise

Merging and Aggregating:in SPSS and STATA

Master-file: HH Ex.xls
Detail-file: Mem Ex.xls

One record from the master table is logical associated with n records in a detail table.

The logical connection is recorded in a key-field in each of the tables; here: Household Id (hhid)

The key-field in the master table must be unique i.e. each record in the master table has one and only one value.

Master-file
HH Ex.xls

hhid	h_locate	h_loan
100	A	1000
200	B	2000
300	C	3000
400	D	4000

Detail-file
Mem Ex.xls

hhid	memid	resp	m_name	m_ismale	m_age	m_inc
100	1	1	Mary	0	24	1000
200	1	0	Chris	1	35	1000
200	2	1	Anne	0	34	900
300	1	1	John	1	45	1000
300	2	0	Angelina	0	44	2000
300	3	0	Bob	1	14	0
400	1	0	Johanna	0	56	1000
400	2	1	Bill	1	55	2000
400	3	0	Walter	1	15	1000
400	4	0	April	0	16	0

1. **Merge** both file in one large detail file which contains all household information
2. **Aggregate** data to a new Master-file (one record per household)

   which contains the following new variables per household

   1. number of family members (h_N)
   2. number of family members age ≥ 18 (h_agege18_N)
   3. total household income (h_inc)
   4. number of female members (h_fem_N)
   5. number of male members (h_male_N)
   6. sum of female income (h_fem_inc)
   7. sum of male income (h_male_inc)
   8. female income as percentage from total household income (h_fem_perc_inc)

Result SPSS:

hhid	h_N	h_agege18_N	h_fem_N	h_male_N	h_inc	h_fem_inc	h_male_inc	h_fem_inc_perc
100	1	1,00	1,00	0,00	1000,00	1000,00	0,00	1,00
200	2	2,00	1,00	1,00	1900,00	900,00	1000,00	0,47
300	3	2,00	1,00	2,00	3000,00	2000,00	1000,00	0,67
400	4	2,00	2,00	2,00	4000,00	1000,00	3000,00	0,25

Result STATA:

hhid	h_agege18_N	h_fem_N	h_male_N	h_inc	h_fem_inc	h_male_inc	h_fem_inc_~c
100	1	1	0	1000	1000	0	1
200	2	1	1	1900	900	1000	.4736842
300	2	1	2	3000	2000	1000	.6666667
400	2	2	2	4000	1000	3000	.25

Syntax SPSS:

```
GET
  FILE='D:\@Guidos Eigene Dateien auf D\Desktop\ZEF BC\Day17 Sample 04 Patrick '+
    'Combining\Exercise\HH Ex.sav'.
DATASET NAME DataSet0 WINDOW=FRONT.

DATASET ACTIVATE $DataSet.

MATCH FILES /TABLE=*
  /FILE='D:\@Guidos Eigene Dateien auf D\Desktop\ZEF BC\Day17 Sample 04 Patrick '+
    'Combining\Exercise\Mem Ex.sav'
  /BY hhid.
EXECUTE.

RECODE m_age (Lowest thru 17=0) (18 thru Highest=1) (ELSE=Copy) INTO isagege18.
EXECUTE.

IF  (m_ismale=0) isfem=1.
EXECUTE.
IF  (m_ismale=1) isfem=0.
EXECUTE.
IF  (m_ismale=1) ismale=1.
EXECUTE.
IF  (m_ismale=0) ismale=0.
EXECUTE.

COMPUTE fem_inc  =m_inc*isfem.
EXECUTE.
COMPUTE male_inc =m_inc*ismale.
EXECUTE.


SORT CASES BY hhid.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /PRESORTED
  /BREAK=hhid
  /h_N=N
  /h_agege18_N =SUM(isagege18)
  /h_fem_N =SUM(isfem)
  /h_male_N =SUM(ismale)
  /h_inc_sum=SUM(m_inc)
  /h_fem_inc_sum =SUM(fem_inc)
  /h_male_inc_sum =SUM(male_inc).

DATASET DECLARE agg.
SORT CASES BY hhid.
AGGREGATE
  /OUTFILE='agg'
  /PRESORTED
  /BREAK=hhid
  /h_N=N
  /h_agege18_N =SUM(isagege18)
  /h_fem_N =SUM(isfem)
  /h_male_N =SUM(ismale)
  /h_inc=SUM(m_inc)
  /h_fem_inc =SUM(fem_inc)
  /h_male_inc =SUM(male_inc).

COMPUTE h_fem_inc_perc  = h_fem_inc/h_inc.
EXECUTE.
/ *
delete variables    respname respage memid memname memismale memage memedu memincome totinc isagege18 isfemale
ismale femincome maleincome. */
```

Syntax STATA:

```
use "HH Ex.dta", clear

merge hhid using "Mem Ex.dta", uniqmaster

save "HH Mem Ex.dta",replace



gen     isagege18 = 0 if m_age<.
replace isagege18 = m_age>=18 & m_age <.

gen isfem    = m_ismale==0
gen ismale   = m_ismale==1

gen fem_inc  = m_inc*isfem
gen male_inc = m_inc*ismale



collapse (sum)   h_agege18_N = isagege18                              ///
                 h_fem_N     = isfem            h_male_N = ismale     ///
                 h_inc       = m_inc                                  ///
                 h_fem_inc   = fem_inc        h_male_inc = male_inc ///
                 , by(hhid)

gen  h_fem_inc_perc  = h_fem_inc/h_inc


save "HH Mem Ex Solution.dta"
```