# Preparing Data Sets for Growth Curve Analysis

## SPSS

Most longitudinal data sets are wide data sets (i.e., person-level, aggregated, one record per case, or within-subjects), but multilevel analysis in SPSS requires that the data be disaggregated so that each time period appears as a row in the data file, most commonly known as long format (i.e., person-period, disaggregated, one record per time period). The `VARSTOCASES` command makes this transition relatively simple.[1] Below, I illustrate `VARSTOCASES` with data on older widows collected by David Morgan and Margaret Neal. The variables are depression scores on the CES-D depression scale (`depress`), self-rated health (`healt`), the number of network members who are sources of negative social interactions (`negative`), and age at the beginning of the study (`age1`). I create a new variable called `time`, which is 1, 2, or 3, to designate the wave of the study (we will recode this to a more common and meaningful values of 0, 1, and 2). `rid` is the respondent's ID number.

```
VARSTOCASES
 /MAKE health FROM hihlth1 hihlth2 hihlth3
 /MAKE negative FROM ntot1 ntot4 ntot7
 /MAKE depress FROM cesdtot1 cesdtot2 cesdtot3
 /INDEX = time
 /KEEP = rid age1.

LIST rid age1 time health negative depress
   /CASES=FROM 1 TO 30.
```

The resulting data file looks like this:

```
 rid age1 time health negative depress

1416  61     1    3       3     22.000
1416  61     2    4       6     20.000
1416  61     3    3       6     29.000
1421  70     1    5       3     22.416
1421  70     2    3      13     25.000
1421  70     3    2      13     14.000
1427  80     1    3       3      7.000
1427  80     2    4       9     14.000
1427  80     3    2      11     10.000
1434  63     1    5      10      4.000
1434  63     2    5      14      6.000
1434  63     3    4      16       .000
1444  73     1    5       3     39.000
1444  73     2    5       3       .000
1444  73     3    5       3       .000
1464  70     1    3       5      6.746
1464  70     2    3       5     20.000
1464  70     3    3       6     28.000
1467  69     1    4       4     26.000
1467  69     2    4       5     30.000
1467  69     3    5       6      7.000
1496  81     1    3       5     30.000
1496  81     2    3       8     17.000
1496  81     3    4       8      6.000
1499  73     1    4       3     16.000
1499  73     2    5       4     17.000
1499  73     3    3       4     14.000
1501  77     1    4       3     16.000
1501  77     2    2       3     23.000
1501  77     3    3       3      7.000


Number of cases read:  30    Number of cases listed:  30
```

[1] There are a number of webpages that will walk you through restructuring a data set using the SPSS menus, see for example this page: https://kb.iu.edu/d/bbqj

## R

```
#clear active frame from previous analyses
rm(d)
rm(widedata)

library(lessR)
widedata = Read("c:/jason/spsswin/mlrclass/growth.sav",quiet=TRUE)

#make hlth variables numeric
widedata$hihlth1 <- as.numeric(widedata$hihlth1)
widedata$hihlth2 <- as.numeric(widedata$hihlth2)
widedata$hihlth3 <- as.numeric(widedata$hihlth3)

widedata$time1=1
widedata$time2=2
widedata$time3=3

#see that data from this file are in wide form (repeated measures) format
View(widedata)

#transform wide data format to long data format
d <- reshape(widedata, idvar="rid",varying=list(c("hihlth1","hihlth2","hihlth3"),
                 c("cesdtot1","cesdtot2","cesdtot3"),c("time1","time2","time3")),
                  v.names=c("health","depress","time"),direction="long",sep="")

#change time codes to 0, 1, and 2
d$time <- d$time - 1

#recheck data, now in long form
View(d)
```

## HLM

Preparing data for growth curve analysis in HLM is relatively simple. Although the system expects two files to be constructed, the level-1 data set (measurements at each time point) and a level-2 data set (person-level measures, such as gender), you can fool HLM by just creating one long format data file and then browsing for the same file for the level-1 and level-2 files in the dialog box used to create the MDM file. You then choose the ID and time-varying variables for the level 1 file and choose the ID and the person-level variables for the level-2 file. The first step is the VARSTOCASES command or array statements in SPSS or SAS as above. The second step is simply to create a new long format data sets (or separate level-1 and level-2 data sets both with person ID).

When creating the .mdm file which merges the two level variables from the SPSS file(s) (*Make new MDM file → stat package input → HLM2*), there is a radio button for choosing *persons within groups* or *measures within persons*. Statistically, there will be no differences if you use "persons within groups." The primary difference is that the output for the "persons within groups" choice will use for $\beta$s and $\gamma$s, whereas "measures within persons" will produce $\pi$s and $\beta$s. I tend to choose "persons within groups" even for longitudinal data, because I am more used to the group-based notation.