

“Rectangularizing” a Data File in SPSS¹

It is sometimes necessary to reshape data so that you can do the kind of analysis that you want to do. If the following example, there are multiple data records for individuals describing events in their lives. An example of this type of file might be arrests. But, you may need to have detailed information about each arrest on one record. This document will demonstrate a couple of ways of creating a one record per case from many. The method you would use would depend on what kind of data you wanted to end up with.

The first method demonstrated will capture all of the information from each individual record. The process is a little cumbersome, and the final product is not always what you need. The second method will show how to create summary data for the individual. Some of the details from the first file will not be carried into the second.

For these examples, I will use the following data - seven arrests for two individuals, and only four variables.

ID	Year	Crime	convicted
1	1988	burglary	1
1	1992	murder	0
1	1993	assault	1
1	1992	DUI	1
2	1989	DUI	1
2	1992	DUI	1
2	1994	Cocaine	1

Method 1

The following program creates new variables for each case. The result will be two cases - one for each ID, with information for up to four crimes (since four is the maximum cases per ID in this example). You may copy this program into an SPSS syntax window and run it to see exactly how it works.

```
* rectangle1.sps .
DATA LIST FREE / id * year * crime (a15) convict * .
BEGIN DATA .
1 1988 burglary 1
1 1992 murder 0
1 1993 assault 1
1 1992 DUI 1
2 1989 DUI 1
2 1992 DUI 1
2 1994 Cocaine 1
END DATA .
sort case by id year .

* Create an indicator of the N of case within ID .
if missing(lag(id)) countvar = 0 .
if lag(id) ne id countvar = 0 .

* Use "Leave" statement so that variables will retain values across cases.
** !!! Do not put an execute or procedure statement between creation of a variable
** and the leave statement .
leave countvar .
compute countvar = countvar + 1 .
* Find out what the maximum number of cases is within ID .
freq var = countvar .
```

¹Prepared by Patty Glynn, University of Washington. 7/22/02, updated 8/12/02 C:\all\help\helpnew\rectangle_spss.wpd

```

* Initialize variables as strings .
string crim_1 crim_2 crim_3 crim_4 (a15) .
* Set character variables to blank and missing for first case within ID .
do repeat
  for1 = crim_1 /
  for2 = crim_2 /
  for3 = crim_3 /
  for4 = crim_4 .
  if countvar = 1 for1 = '' .
  if countvar = 1 for2 = '' .
  if countvar = 1 for3 = '' .
  if countvar = 1 for4 = '' .
end repeat .
do repeat
  for1 = year_1 conv_1 /
  for2 = year_2 conv_2 /
  for3 = year_3 conv_3 /
  for4 = year_4 conv_4 .
  if countvar = 1 for1 = -9 .
  if countvar = 1 for2 = -9 .
  if countvar = 1 for3 = -9 .
  if countvar = 1 for4 = -9 .
end repeat .
missing values year_1 to conv_4 (-9) .

* Assign values to new variables, depending on "countvar" .
do repeat
  orig = year convict crime /
  for1 = year_1 conv_1 crim_1 /
  for2 = year_2 conv_2 crim_2 /
  for3 = year_3 conv_3 crim_3 /
  for4 = year_4 conv_4 crim_4 .
  if countvar = 1 for1 = orig .
  if countvar = 2 for2 = orig .
  if countvar = 3 for3 = orig .
  if countvar = 4 for4 = orig .
end repeat .

* Use "Leave" statement so that variables will retain values across cases.
** !!! Do not put an execute or procedure statement between creation of a variable
** and the leave statement .
leave year_1 conv_1 crim_1
      year_2 conv_2 crim_2
      year_3 conv_3 crim_3
      year_4 conv_4 crim_4 .

* Sort cases and create a variable that indicates the last case of each ID .
* Only the last case for each ID will have values for all of the variables .
sort cases by id (d) year (d) .
if lag(id) gt id keepcase = 1 .
if missing(lag(id)) keepcase = 1 .
sort case by id year .

* There should be as many valid cases of keepcase as there are individuals in file .
freq var = keepcase .

* Select only the last case for each ID - which will have values for all variables.
select if keepcase = 1 .
* Save file and keep only variables needed.
save outfile = 'c:\trash\rectangle1.sav' / keep id CRIM_1 to CONV_4 .

```

Method 2

This method uses proc summary to create a data set with summarized information for each ID. For this example, I have added a variable "time_in", which is time incarcerated.

```
* rectangle2.sps .
DATA LIST FREE / id * year * crime (a15) convict * time_in * .
begin DATA .
1      1988      burglary      1      12
1      1992      murder        0      -9
1      1993      assault       1      15
1      1992      DUI           1      1
2      1989      DUI           1      1
2      1992      DUI           1      2
2      1994      Cocaine       1      3
END DATA .
missing values time_in (-9) .

* Use code like the following to determine values that need to be categorized .
string upcrime (a15) .
compute UPCRIME = UPCASE(CRIME) .
freq var = upcrime .

if UPCRIME = 'BURGLARY' and convict = 1 propconv = 1 .
if UPCRIME = 'MURDER'   and convict = 1 violconv = 1 .
if UPCRIME = 'ASSAULT'  and convict = 1 violconv = 1 .
if UPCRIME = 'DUI'     and convict = 1 alconv  = 1 .
if UPCRIME = 'COCAINE' and convict = 1 drugconv = 1 .
if year                lt 1990  pre1990 = 1 .
if year                ge 1990  post1990 = 1 .

* use Aggregate to get sums, means, etc, across cases .
* The "break" value asks SPSS to get values within IDs .
Aggregate outfile = *
/ break = id
/ nconvict = sum(convict )
/ npropcon = sum(propconv )
/ nviolcon = sum(violconv )
/ nalcccon = sum(alconv )
/ ndruggcon = sum(drugconv )
/ npre90   = sum(pre1990 )
/ npost90  = sum(post1990 )
/ ttimein  = sum(time_in )
/ narrest  = N .

variable labels
nconvict 'Number of convictions, total'
npropcon 'Property crime conviction'
nviolcon 'Violent crime conviction'
nalcccon 'Alcohol crime conviction'
ndruggcon 'Drug crime conviction'
npre90   'Number of convictions, pre-1990'
npost90  'Number of convictions, 1990 +'
time_in  'Months in prison'
narrest  'N of arrests for person'

* change missing values to zero .
do repeat
a = nconvict to narrest .
if missing(a) a = 0 .
end repeat .
freq var = nconvict to narrest .
```