

Formatting data files for repeated-measures analyses in SPSS: Using the Aggregate and Restructure procedures

Guy L. Lacroix
Concordia University

Gyslain Giguère
Université du Québec à Montréal

In this tutorial, we demonstrate how to use the Aggregate and Restructure procedures available in SPSS (versions 11 and up) to prepare data files for repeated-measures analyses. In the first two sections of the tutorial, we briefly describe the Aggregate and Restructure procedures. In the final section, we present an example in which the data from a fictional lexical decision task are prepared for analysis using a mixed-design ANOVA. The tutorial demonstrates that the presented method is the most efficient way to prepare data for repeated-measures analyses in SPSS.

Specialized experiment generation and data collection software has greatly simplified the experimental psychologist's work during the last few decades. Commercially and freely available software such as E-Prime (Schneider, Eschman, & Zuccolotto, 2002a; 2002b) and PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) allow for the creation of basic experiments in a few minutes. Moreover, once the testing is finished, the experimenter possesses a complete digital record of the participants' responses and reaction times, which can then be submitted to statistical analyses.

Unfortunately, the raw data output files are rarely in a format to be immediately analyzed with SPSS. Because the typical data file usually includes one line per experimental trial (with participant identification, independent variables, and dependent variables appearing in separate columns), analyses involving repeated measures cannot be conducted without additional work. One must first calculate the average performance for each dependent variable at each level of each condition. Then, the data file must be reconfigured to show each participant's data on a single row (See Figure 1 for an illustration of the steps to be taken). It is only when these steps have been completed that the

researcher may conduct inferential statistics for experimental designs implicating repeated measures.

To address this difficulty, students and researchers use a wide variety of solutions. For instance, some will make the computations and modifications "by hand" using spreadsheet software such as Excel, while others will write programs on a wide range of platforms to accomplish the task in a more automatized fashion. Although these techniques ultimately achieve their goal, they have three important drawbacks: they are time consuming because they force the experimenter to simultaneously manipulate multiple programs; they are not very flexible, as, often, they must be customized for each analysis; and they are somewhat prone to error.

Since the release of SPSS (version 11), however, it is now possible to quickly and flexibly modify a raw data output file to prepare it for repeated-measures analyses using only SPSS functionalities. The goal of this tutorial paper is to demonstrate how these modifications may be accomplished using two SPSS procedures: Aggregate and Restructure. The intended readership for this tutorial is the upper-level undergraduate or graduate experimental psychology student who is learning to format and analyze data in SPSS, as well as the more experienced researcher who wishes to learn these data manipulation procedures quickly. In the first two sections of this paper, we will briefly describe the Aggregate and Restructure functions. In the final section, we

Starting July 1st, 2006, correspondence concerning this tutorial may be addressed to Guy L. Lacroix, Department of Psychology, Carleton University, B550 Loeb Building, 1125 Colonel By Drive, Ottawa, ON, Canada, K1S 5B6.

1 : subject 1

	subject	presrate	wordtype	rt
1	1	slow	word	1045
2	1	slow	nonword	1283
3	1	slow	word	1009
4	1	slow	nonword	1269
5	1	slow	word	1031
6	1	slow	nonword	1290
7	1	slow	word	1006
8	1	slow	nonword	1264
9	2	fast	word	1021
10	2	fast	nonword	1270

1. Imagine a lexical decision task experiment in which 10 participants must determine whether individually shown strings of letters are words or nonwords. Suppose that the participants are assigned to one of two groups: slow vs. fast stimulus presentation time. Also suppose that there are only four stimuli in the experiment: 2 words and 2 nonwords. Finally, assume that the only dependent variable of interest is the participants' response times. The resulting raw output file of this experiment would be similar to the one presented on the left. It can be seen that the data associated with each participant is presented on multiple rows and that the response time data are not averaged. Hence, in its present format, the data cannot be immediately submitted to a repeated-measures analysis.

1 : subject 1 Find

	subject	presrate	wordtype	rt_1
1	1	slow	nonword	1276.45
2	1	slow	word	1022.65
3	2	fast	nonword	1274.93
4	2	fast	word	1021.45
5	3	slow	nonword	1267.56
6	3	slow	word	1016.71
7	4	fast	nonword	1268.72
8	4	fast	word	1029.45
9	5	slow	nonword	1276.20
10	5	slow	word	1024.52

2. To remedy this situation, the research must first calculate, for each participant, the average response time at both levels of the within-subjects independent variable (words and nonwords). The result of this manipulation is shown on the left. Note that the data are still not amenable to a repeated-measures analysis in SPSS.

1 : subject 1

	subject	presrate	nonword	word
1	1	slow	1276.45	1022.65
2	2	fast	1274.93	1021.45
3	3	slow	1267.56	1016.71
4	4	fast	1268.72	1029.45
5	5	slow	1276.20	1024.52
6	6	fast	1258.20	1021.25
7	7	slow	1270.05	1031.47
8	8	fast	1264.27	1033.94
9	9	slow	1272.99	1033.13
10	10	fast	1260.24	1028.33

3. In a final step, the data file must be modified such that the data for each participant appear on a single row with each level of the within-subjects variable appearing as column header. Now, the researcher can conduct the repeated-measures analysis.

Figure 1. An example of the three typical steps taken to prepare a raw data file for a repeated-measures analysis. Step 1 shows the raw data output file generated by experimental software such as E-prime once it has been imported into SPSS. Step 2 shows data that has been aggregated. Step 3 shows data that has been restructured and that is ready for a repeated-measures analysis.

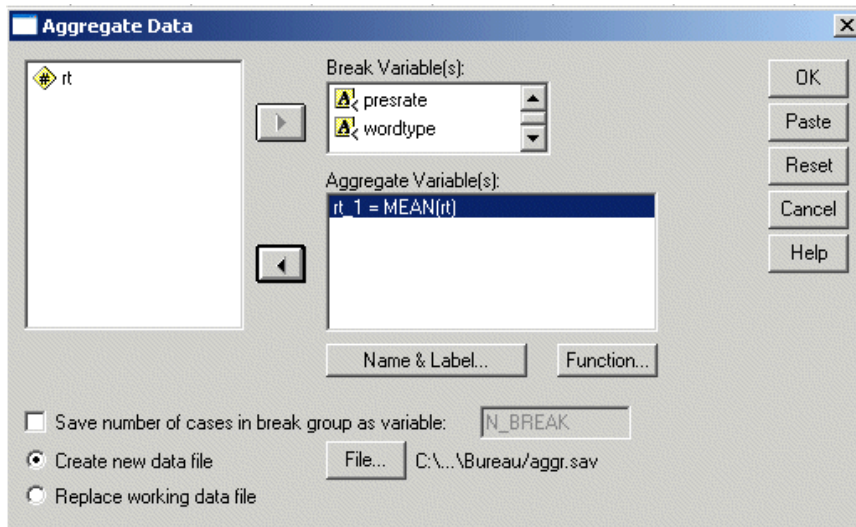


Figure 2. Screen capture of the opening Aggregate procedure window in SPSS (standard version 11).

will present a concrete example and explain how these procedures may be used in combination to modify raw data output files in preparation for repeated-measures analyses.

The Aggregate procedure

The SPSS Aggregate procedure allows the researcher to obtain summary statistics for any dependent variable that may be specified for each participant and each independent variable. When the Aggregate function is accessed in the Data menu, the window shown in Figure 2 appears.

All variables are listed in the box on the left of the

window and may be dragged into the *Break variable(s)* box or *Aggregate variable(s)* box. Break variables are variables for which a summary statistic is to be computed and aggregate variables are the variables to be summarized. Most commonly, the participant identifier variable and the independent variables are dragged into the *Break variable(s)* box and the dependent variables, in the *Aggregate variable(s)* box. By default, the Aggregate function calculates arithmetic mean for each specified break variable. Nevertheless, by clicking on the *function...* button below the *Aggregate variable(s)* box, the user may choose to compute other

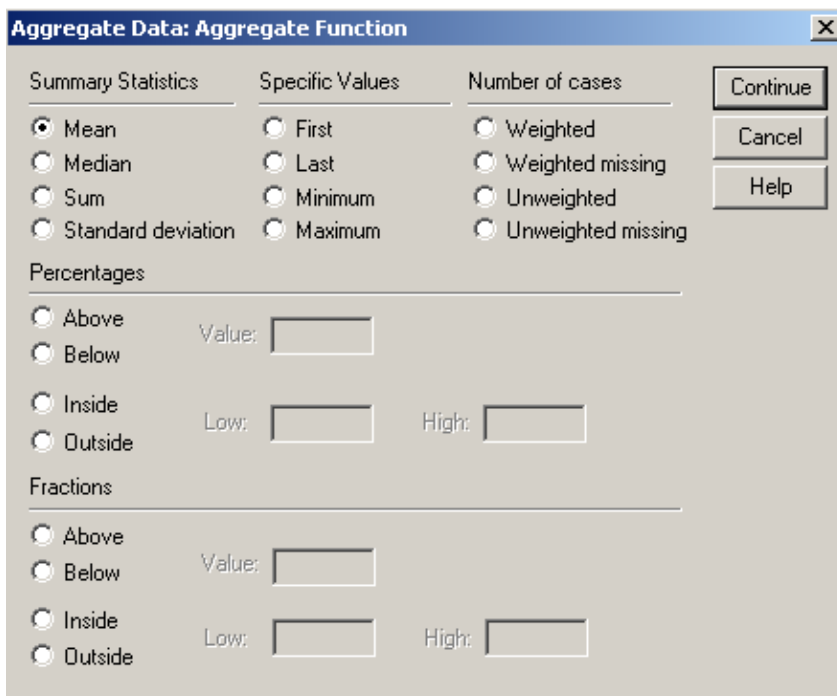


Figure 3. Screen capture of the options that are available when the *Function...* button on the opening Aggregate procedure is selected in SPSS (standard version 11).

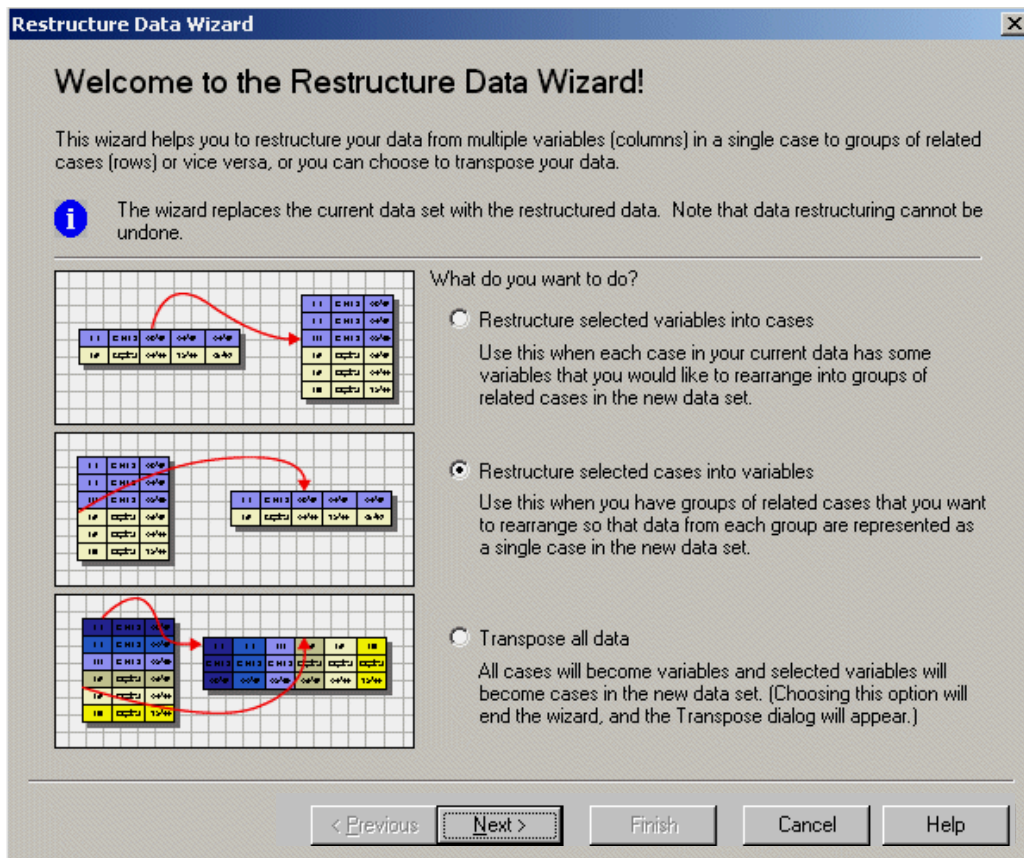


Figure 4. Screen capture of the opening Restructure procedure window in SPSS (standard version 11).

summary statistics such as the median, the sum, and the standard deviation (see Figure 3 for a screen capture of the *function...* window). Finally, the researcher may choose to save the newly created summary data in a different file by clicking on the *Create new data file* option or continue working in the same file by selecting the *Replace working data file* option. When the Aggregate function is executed by pressing the OK button, the new file will display the selected summary statistic for each specified break variable.

The Restructure procedure

The SPSS Restructure function, also found in the Data menu, allows the researcher to change the organization of data files. When Restructure is accessed, the window shown in Figure 4 appears. Three options are available: *Restructure selected variable into cases*, *Restructure selected cases into variables*, and *Transpose all data*. Because the goal of this tutorial is to prepare data for repeated-measures analyses, we will focus on the second option. In essence, the *restructure selected cases into variables* option allows the researcher to select the data found in rows and transpose them into columns. When selected (and the *Next* button is clicked), the window shown in Figure 5 appears. The box on the left contains all the variables. The variables may be

dragged into one of two boxes: *Identifier variable(s)* and *Index variable(s)*. The participant identifier variable and between-groups variables must be dragged into the *Identifier variable(s)* box. The repeated-measures variables must be dragged into the *Index variable(s)* box. Note that the variable or variables, which contain the dependent variables, remain unselected. When this step is completed, the experimenter will once again click the *Next* button and a third window entitled *Cases to variables: sorting data* will appear. By default, the option “Yes – data will be sorted by the identifier and index variables” will be selected and the *Finish* button will be enabled. To finish the Restructure procedure, the experimenter may immediately click this button. If the *Next* button is selected, two more windows will appear. The first allows the experimenter to sort the levels of the different independent variables by level or variable, and the last window executes the procedure or pastes the syntax for the procedure into a syntax window.

Preparing data for repeated-measures analyses using the Aggregate and Restructure procedures

In Figure 1, we present the three typical steps taken to prepare a raw data file that has been imported into SPSS for a repeated-measures analysis. In step 1, the output of a

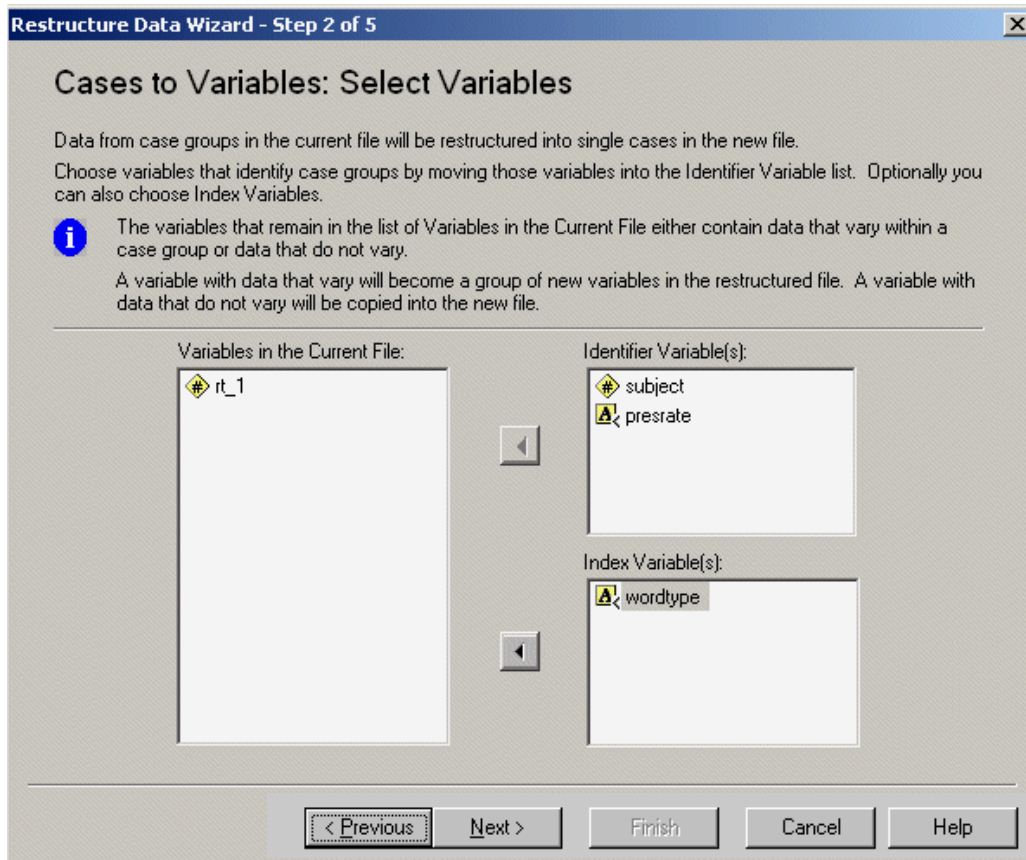


Figure 5. Screen capture of the second Restructure procedure window in SPSS (standard version 11).

fictional lexical decision task is given. There are two independent variables: one between-subject variable (presentation rate: slow vs. fast) and one within-subjects variable (stimulus type: word vs. nonword). The dependent variable is the participants' response time. The reader will observe that the data is found in four columns: subject (the participant identifier variable), presrate (the between-subject variable), wordtype (the within-subjects variable), and rt (the dependent variable).

To prepare these data for analysis, the researcher will first use the Aggregate procedure to calculate the average response times for each level of the within-subjects variable (for each stimulus type). Once the Aggregate procedure is selected in the Data menu, the researcher will drag the subject, presrate, and wordtype variables into the *Break variable(s)* box and the rt variable into the *Aggregate variable(s)* box. When the procedure is executed, the resulting data file will be that shown in the second step of Figure 1. At this point, the data are still not ready for a repeated-measures analysis because the format of the file only allows for a between-subject ANOVA (using, for instance, the GLM univariate command found in the Analyze menu). Clearly, this design is inappropriate to analyze the data shown in the Figure 1 example.

To finish preparing the file for analysis, the researcher will now choose the Restructure procedure in the Data menu and select the *Cases to variables* option. Then, the researcher will drag the subject and presrate variables into the *Identifier variable(s)* box and the wordtype variable into the *Index variable(s)* box. Finally, the researcher must click on the *Next* button before executing the procedure. The resulting data file is shown in third step of Figure 1. The data are ready for a mixed-design ANOVA using the GLM repeated measures command found in the Analyze menu. Note that the same data manipulation steps may be executed in a SPSS syntax window. The syntax is explained in Figure 6.

Summary

In this tutorial, we have addressed a problem that the experimental psychologist often encounters, namely formatting data files for repeated-measures analyses. Yet, the functionalities offered in SPSS (versions 11 and up) allows researchers to prepare their data files for many types of analyses involving repeated-measures including repeated-measures t-tests, repeated-measures or mixed design ANOVAs, regressions, and factor analyses. Considering the ease of use and efficiency of the Aggregate

and Restructure procedures to reformat data, they are the method of choice for the researcher who uses SPSS for statistic analyses.

If the reader wishes to practice the formatting process demonstrated in this paper, the data file used as an illustration in Figure 1 is available on the Tutorials for Quantitative Methods for Psychology website at <http://www.tqmp.org>. The reader will find the SPSS data file by choosing the *Content* menu and then, by clicking on the word “sample” that follows the title of this paper.

References

- Cohen J. D., MacWhinney B., Flatt M., Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257-271.
- Schneider, W., Eschman, A., Zuccolotto, A. (2002a). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.
- Schneider, W., Eschman, A., Zuccolotto, A. (2002b). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools Inc.

Manuscript received march 31, 2006

Manuscript accepted april 17, 2006

The same data manipulation steps may be executed in a SPSS syntax window with the use of the AGGREGATE and CASESTOVARS commands. To aggregate the data, the AGGREGATE command is used in combination with the two following subcommands: OUTFILE, which identifies the path and the name of the aggregated file (between single quotes), and BREAK, which consists of the list of variables for which a summary statistic is to be computed. A third subcommand is used to define the aggregated variable. In the Figure 1 example, the MEAN function with the name of the variable that must be transformed (in parentheses) would be used to create the new variable “rt_1”. The listing below details the syntax related to the Figure 1 example.

```
AGGREGATE
/OUTFILE='C:\aggregate.sav'
/BREAK=subject presrate wordtype
/rt_1 = MEAN(rt).
```

Once the Aggregate command has been executed, the user must open the new data file to restructure the data. The data must first be sorted on the basis of the identifier and index variables with the SORT CASES BY command followed by the list of relevant variables. Then, the CASESTOVARS command is used to restructure the cases into variables with two subcommands: ID, a list of the identifier variables (participants and between-groups variables); and INDEX, a list of index variables (the within-groups variables). The listing below shows the syntax related to the Figure 1 example.

```
SORT CASES BY subject presrate wordtype.
CASESTOVARS
/ID = subject presrate
/INDEX = wordtype.
```

Figure 6. A description of the commands AGGREGATE and CASESTOVARS that may be executed to prepare data files for repeated-measures analyses using a SPSS syntax window.