User Manual for RecoderPlus

# About RecoderPlus

RecoderPlus is an SPSS macro that reads in a set of variables from an SPSS system file, and processes them in one of 3 ways:

1. Creates contrast coded variables for the variables listed as CATVARS. Contrast coding can be done using dummy or effect coding by using the parameter CONTYPE.
2. Imputes missing values for a variable with the mean of the non-missing values for variables listed as IMPMEAN. It can impute using the overall mean across all cases, or within groups defined by the variables BYVARS. It also creates a dummy coded variable indicating those cases with missing values in the original variable, which values were imputed.
3. Recodes the values of the variable listed as USECRIT to the mean for the group on the criterion variable CRITVAR. The imputed value could be calculated across the entire sample, or within the group defined by the variables BYVARS.

In addition to recoding individual variables, it creates interaction variables between pairs of original variables, using the recoded vectors.

# Requirements

RecoderPlus requires a working copy of SPSS installed in the computer where the analysis is conducted. It also requires an SPSS system file with at least one variable to be recoded. When creating interaction variable, you will need at least 2 variables. Variables can be numeric of string variables.

# Using RecoderPlus

To use RecoderPlus to process variables, you will need to create an SPSS syntax file with an INCLUDE command, followed by a call to the macro. The syntax will look something like this:

include file = "C:\SPSS Macros\RecoderPlus\RecoderPlus.ieasps".

!RecoderPlus

indir = "C:\SPSS Macros\RecoderPlus" /

infile = RecoderPlusSampleData /

outdir = "C:\SPSS Macros\RecoderPlus" /

outfile = RecoderPlusSampleResults /

Contype = d /

Catvars = CatVar1 CatVar2 CatVar3 StringVar1 StringVar2/

ImpMean = ImpMean1 ImpMean2 /

UseCrit = UseCrit1 UseCrit2 /

CritVar = CritVar /

IntVarsA = CatVar1 CatVar2 UseCrit1 CatVar1 ImpMean1 ImpMean1 UseCrit1 /

IntVarsB = ImpMean1 StringVar1 CatVar2 UseCrit1 ImpMean2 UseCrit1 UseCrit2 /

IntVec = Cat1Imp1 Cat3Str1 Use1Cat2 Cat1Use1 Imp1Imp2 Imp1Use2 Use1Use2 /

viewcod = n /

qcstats = y /

clean = y .

Notice that the first line INCLUDEs the macro, and the following lines are the call to the macro. Some of the parameters are optional and can be left blank. You only need to include the macro once per interactive session.

When calling the macro, you need to use the exclamation point (!) before RecoderPlus, as in !RecoderPlus. Following !RecoderPlus you will have a series of parameters with the corresponding value(s). Each parameter is separated by a slash, much in the same way you would separate subcommands when writing SPSS syntax. The last parameter ends with a period (“.”). Some commands are mandatory, like the name and location of the input and output file, but others are optional. The section on parameters in this User Manual will specify the parameters, their use, and accepted and default values.

The parameters and corresponding values are not case sensitive. The order of the parameters is also not important. Neither is the alignment, other than they should not start on the first column. You can also have multiple parameters in a single line, but always separated by a slash. If you repeat a parameter in the call, the last instance of the parameter will be the one used by the program.

The alignment and sequence of the parameters in the syntax above is done as such for looks and ease of reading.

# Output Files

Depending on the setup and configuration, RecoderPlus creates several output files. There is a set of files that are always generated, and another set that is generated depending on the setup. A file with a log of the output is always generated and saved in SPV format. There are also files with the transformation instructions that are generated. These will be generated only if needed based on the parameters used for the run.

All files generated by RecoderPlus are written to the directory specified with the parameter OUTDIR, and use the text specified with the parameter OUTFILE as the root for the file name. What follows is a list of the files generated by RecoderPlus.

Please note that if a file with the same name already exists, the existing file will be automatically replaced with no warning to the user.

## [OUTFILE]\_Output.SPV

This file contains the entire SPSS output file in native SPSS SPV format. Depending on the options requested, the contents of this file will vary. But in general, this file will mostly have output that should be used for quality control of the data and the corresponding analysis. There is a single SPV file for each call to the macro submitted.

This file will have the following sections:

1. The minimum and maximum for the recoded categories for each of the CATVARS. A value of 0 indicates there are missing values for that variable. The number of valid cases is also included in the table.
2. When specifying USECRIT variables, one table for each of these variables showing the mean on the CRIRTVAR for each response category. The value -999 corresponds to missing values in the original variable. Here you want to see that the mean for Critvar and [USECRIT]\_1 are the same and for [USECRIT]\_1 the standard deviation is zero.
3. When specifying IMPMEAN variables, one table for each of these variables showing the mean on the IMPMEAN variable for all valid responses (variable [IMPMEAN]\_n = -777), and the mean on the transformed variable ([IMPMEAN]\_1) for the missing values (variable [IMPMEAN]\_n = -888 or -999) equal to the mean for those with no missing [IMPMEAN] and standard deviation equal to zero. In addition, the variable [IMPMEAN]\_2 shows mean=1 and standard deviation=0 for all cases missing the variables [IMPMEAN] and mean=0 and standard deviation=0 for cases with valid values.
4. When specifying CATVARS variables, one table for each of these variables showing a cross between the categories for the original variable ([CATVARS]\_n), and the contrast coded variables. The table shows mean 1 or -1 for the category represented by the contract variables. Standard deviation for all cells should be equal to 0. Categories for the original variable have been recoded to 99 indicating the missing category.
5. When interaction vectors have been requested, the program will show tables for each interaction between two categorical variables.

## [OUTFILE]\_Contrasts.SAV and .CSV

This file will contain the transformed variables and contrast vectors created. When SELCRIT is used, it will only contain the records that meet the SELCRIT criteria. Otherwise it will contain all the records from the original file. The transformed variables included in this file are the following:

[CATVARS]\_[CONTYPE]#: there will be one variable for each contrast vector created for the categorical variables. The first contrast vector will be # = 2 representing the second category. There will be as many contrast vectors as there are [categories – 1] where the categories include missing values, if any were to exists in the file for the variables.

[IMPMEAN]\_1 and [IMPMEAN]\_2: These variables will be the original IMPMEAN variable transformed so that those with missing values receive the average value of the IMPMEAN variable for the other values in the corresponding group ([IMPMEAN\_1]) and an indicator variable identifying the records that had a missing value for the IMPMEAN variable ([IMPMEAN]\_2).

[USECRIT]\_1 and [USECRIT]\_2 These will be the transformed variable where the original values have been substituted with the mean on the criterion variable (CRITVAR) of all cases in the corresponding group and an indicator variable identifying the records that had a missing value for the USECRIT variable ([USECRIT]\_2).

[INTVEC]\_# These will represent the interaction vector between pairs of variables specific in the setup for the program. The number of interaction variables will depend on the number of transformed variables. For example, the interaction of 2 USECRIT variables will be captured by one variable. The interaction between a categorical variable with 4 categories and another one with 3 categories will require 6 interaction variables or (4-1) \* (3-1) interactions.

## [OUTFILE]\_RecodedVars.csv

This file contains one record for each transformed variable created from the original variables. The file contains one record for each variable created. The variables in the file are the following:

Varname The name of the original variable in the file

ContrastVar The name of the newly created variable.

Type The type of the variable (Use Criterion, Contrast coded or Impute mean).

VarSeq The sequence of the variable in the list as read by the program.

Category For contrast coded variables (CATVARS), the category sequence in the original data with the values sorted alphabetically.

CategoryV For contrast coded variables (CATVARS), the category sequence in the original data with the values sorted alphabetically, but with the last category set to 99, should there be missing data for the variable in the file.

## [OUTFILE]\_Interactions.csv

This file contains one record for each interaction variable created from the transformed variables. The file contains one record for each variable created. The variables in the file are the following:

VarNameA and VarNameB: The names for the two variables used for the interaction

GroupA and GroupB: When creating interactions with categorical variables, the corresponding category sequence for the categorical variable.

IntVar The name of the interaction variable created. These will use the name provide in the INTVEC parameter, followed by a suffix “\_#” where “#” begins with numeral 1.

NInteract When creating interaction terms between 2 categorical variables, the total number of interaction terms created.

IntSeq When creating interaction terms between 2 categorical variables, the sequence of interaction terms created.

RECLINE1 The first transformation applied to create the interaction variable.

RECLINE2 The second transformation applied to create the interaction variable. This is only necessary when interacting categorical variables.

## [OUTFILE]\_CatRecodes.TXT [OUTFILE]\_ImpMeanRecodes.TXT [OUTFILE]\_UseCritRecodes.TXT

These are text files that contain the recoding instructions used to transform the variables in the original file. These can be used for quality control purpose to verify the adequacy of the recodes made to the original variables. Each of these files will only be created if there were these types of variable specified in the call.

## [OUTFILE]\_CatRecodesCheck.TXT [OUTFILE]\_ImpMeanRecodesCheck.TXT [OUTFILE]\_UseCritRecodesCheck.TXT

These are text files that contain the instructions to create summary statistics to check the transformation of the variables. These can be used for quality control purpose to verify the adequacy of the recodes made to the original variables. Each of these files will only be created if there were these types of variable specified in the call.

# Parameters

What follows is a list of the parameter that can be used when calling RecoderPlus. They are in the logical order in which you would want to consider them, and not in alphabetical order. But the order of the parameters in the call can be any order. Some parameters are optional, some are mandatory, and some have default values. This is all indicated below.

Depending on the parameter, the values assigned can be one or many. They are not case sensitive. The parameters are assigned a value or values using the equal sign. When more than one value is specified for a parameter, these need to be separated with spaces, and the values from one parameter to the next are always separated by a forward slash (‘/’).

## Data Input and Output

INDIR (not optional; no default)

Directory with the file that has the variables that will be processed. The directory location must be fully spelled out and enclosed in quotes.

Example:

Indir = "C:\SPSS Macros\RecoderPlus" /

INFILE (not Optional; no default)

File with the variables that will be processed. It does not need the SAV extension for the file name.

Example:

Infile = RecoderPlusSampleData /

OUTDIR (not optional; no default)

Directory where to write the output files. The directory location must exist and be accessible from your computer, be fully spelled out and enclosed in quotes. The program outputs several files described earlier in this manual.

Example:

Outdir = "C:\SPSS Macros\RecoderPlus" /

OUTFILE (not optional; no default)

Name used for the output files. Presented as [OUTFILE] in the descriptions above. This value cannot start with a number, cannot have special characters or spaces, and must comply with the SPSS variable naming convention.

Example:

Outfile = RecoderPlusSampleResults /

SELCRIT (optional; no default)

The selection criteria to apply when reading the INFILE. Only records that meet the specific selection criteria are read and accounted for in the processing and are written to the corresponding output file.

The selection criteria must be specified using SPSS IF/THEN logic, but no IF is needed. Any, and all operations must fit in a single line ending with a forward slash (“/”).

Example:

Selcrit = (var1 = 1 and var2 > 4) /

In the example above, only cases with VAR1 = 1 and VAR2 > 4 will be read from the INFILE. Parenthesis are used for readability but are not necessary.

WGTVAR (optional; no default; options = a single variable with the survey or sampling weight)

The name of the variable with the survey or sampling weight. When no WGTVAR is specified, RecoderPlus assigns all cases a weight of 1, therefore all response records contribute equally to the calculations. When a WGTVAR is specified, this is used to weight the response records. Cases with zero, missing and negative sampling weights are excluded from the analysis from the outset.

The sampling weights are used only when processing the IMPMEAN and USECRIT variables. These have no effect for the categorical variables that are contrast coded.

Example:

wgtvar = TOTWGT /

IDVars (optional; no default)

Variable(s) to be used as identification variable(s) for the records in the file. The program automatically generates a variable indicating the case sequence number in the input file (CaseSeq) and the output file is sorted in the same order as the input file.

Example:

idvars = idcntry idstud /

CONTYPE (optional; default = D; options = D or E)

Instructs the program on the type of contrast variables to create with the CATVARS provided. Options are to create dummy coded (“D”) or effect coded (“E”) contrast variables. Regardless of the type or contrast requested or the type of variable (string or numeric), the value for the first category sorted in alphabetical order will be used as the reference category. A single CONTYPE can be specified that will apply to all the categorical variables. Contrast coded variables are created using the original variable name, with the suffix “\_D#” or “\_E#” added to the name, where # corresponds to the sequential number of the value for the variable sorted in alphabetical order. Missing values will be last in the sort order.

Example:

Contype = d /

In the case where we have, for example, a variable CATVAR1 with 3 categories, the program will create the following contrast coded variables: CATVAR1\_D2, CATVAR1\_D3, and should these be missing data, CATVAR1\_D4. If CONTYPE had been specified as E, the program would have created variables CATVAR1\_E2, CATVAR\_E3, and CATVAR1\_E4.

CATVARS (optional; no default)

This parameter provides the names of the categorical variables that will be used to create the contrast coded variables. The variables must be specific one by one. They will be processed in sequential order as they appear in this list. The categorical variables can be string or numeric. They can also have missing values. The first category will be used as the reference category for the creation of the contrast coded variables. If missing values are found in the variable, the program will also create a contrast coded variable for it. The name of the categorical variable will be used as the root for the name of the contrast coded variable.

Even though this parameter is optional, the program requires at least one CATVARS, or one IMPMEAN, or one USECRIT variable in order to operate.

Example:

Catvars = CatVar1 CatVar2 CatVar3 StringVar1 StringVar2 /

IMPMEAN (optional; no default)

This is the list of variables that will have missing values replaced with the mean for the variable across the remaining cases. By the nature of the processing of these variables, these variables must be numeric. When BYVARS is specified, the mean used for the imputation is calculated for each of the groups defined by the unique combinations of the BYVARS. For example, when using BYVARS = COUNTRY SCHTYPE, the missing values will be imputed using the mean of the non-missing values within each school type by country. When there are no other cases with non-missing values within a group, the value imputed will be taken from the mean of the group defined by the next level. For example, if there are no non-missing values within a school type within a country, the value imputed will be the mean for the country. Variables with the imputed mean will be named using the name for the original variable, with the suffix “\_1”. In addition, a second variable will be created indicate the cases where the missing value was replaced by the mean. This second variable is named using the suffix “\_2”.

Even though this parameter is optional, the program requires at least one CATVARS, or one IMPMEAN, or one USECRIT variable in order to operate.

Example:

ImpMean = ImpMean1 /

In this example the program will create a new variable called ImpMean1\_1 where the missing values have been replaced with the mean of the variable ImpMean1, and a second variable called ImpMean1\_2 where cases with missing values for ImpMean1 receive value of 1, otherwise 0.

USECRIT (optional; no default)

This is the list of variables that will be criterion scaled. Criterion scaled variables are those that are transformed by replacing the values with the mean on a criterion variable for those with that values on the variable. The criterion variable is specified with the parameter CRITVAR. These variables can be string or numeric variables. The missing categories are replaced with the mean on the criterion variable for the existing cases. If BYVARS is specified, the means are calculated separately for each of the groups defined by the BYVARS.

The program will create one variable using the name of the USECRIT variables with the suffix “\_1”.

Even though this parameter is optional, the program requires at least one CATVARS, or one IMPMEAN, or one USECRIT variable in order to operate.

Example:

UseCrit = UseCrit1 UseCrit2 /

CRITVAR (optional; no default)

This is the criterion variable used to transform the USECRIT variables. This is a single variable and because of how it is used, it must be numeric. These must be valid values of the CRITVAR for each of the distinct values for the USECRIT variables. Missing values in the USECRIT variables will be replaced with the mean CRITVAR for those with missing cases in the USECRIT variable.

Example:

CritVar = CritVar /

INTVARSA, INVARSB and INTVEC (optional; no default)

These parameters are used to specify the sets of interactions to be created by the program. The program will create as many sets of interaction vectors as specified in the parameter INTVEC. The parameters INTVARSA and INTVARSB list the pairs or variables that will be used to create the interactions. These will be paired first with first, second with second, etc. The variables specified in INTVARSA and INTVARSB must have been included as one of the CATVARS, IMPMEAN or USECRIT variables. When creating the set of interaction vectors, the program will use the transformed variables. Any type of variable combination is possible. Depending on the type of variables combined to create the interaction set, there could be as few as one interaction variable in the set.

The variables capturing the interaction will use the value in INTVEC followed by the suffix “\_#” where # is the sequence for the interaction. As many combinations of variables are possible, please check the file listing of the interactions to confirm the interactions are created as expected, and which corresponding pairing they represent.

Example:

IntVarsA = Var1 Var2 Var3 /

IntVarsB = VarA VarB VarC /

IntVec = Int1A Int2B Int3C /

In this example 3 sets of interaction variables will be created. Int1A will represent the interactions between variables Var1 and VarA, Int2B will represent the interactions between variables Var2 and VarB, and Int3C will represent the interactions between variables Var3 and VarC.

BYVARS (optional)

This option specifies the classification variable(s) to use when replacing missing values with the mean or recoding the values with the mean on a criterion variable.

Example:

BYVARS = SEX /

In this case missing values for the IMPMEAN will be replaced with the overall mean for boys or girls, depending on the categorization of the respondent.

QCSTATS (optional; default = Y; options = Y or N)

Using this option instructs the program to create a series of tables that can be used to check the proper transformation of the variables. Unless you are absolutely sure about the quality of the data, the way in which the program operates and the quality of the resulting transformations, we recommend using the default (Y) and checking the tables to confirm the variables have been transformed as intended.

Example:

QCStats = Y /

VIEWCOD (optional; default = N; options = Y or N)

Expands and displays in the output file the code generated by the macro. Useful for troubleshooting the processing and identifying where exactly problems or warnings occur, if any.

Example:

viewcod = N /

CLEAN (optional; default = Y; options = Y or N)

Instructs the program to delete (CLEAN = Y) the temporary files created during processing. During processing, several temporary files are created and stored in the OUTDIR. At times, preserving and viewing these files could be useful for troubleshooting the code or identify problems with the data.

Example:

clean = Y /

## Reading a List of Variables from an External Source

The list of variables to process can be read from an external .xlsx file. This is done when specifying the parameter !VFILE and the optional parameter !SHEET. When reading the list of variables from an external source, all values for CATVARS, IMPMEAN and USECRIT are ignored.

VFILE (optional)

This parameter is used to specify the name of an .xlsx file containing the specifications for the variables. When a VFILE is specified, the list of variables will be read from this external file. This file must be located in the INDIR directory, must be in .XLSX format, must have one entry per variable in the analysis, and also must contain the following four columns with the corresponding header:

**Variable**: A list of the variable that will be recoded.

**CatVar**: An indicator (Y) whether the variable should be treated as a CATVAR (See description above).

**UseCrit**: An indicator (Y) whether the variable should be treated as an USECRIT (See description above).

**ImpMean:** An indicator (Y) whether the variable should be treated as an IMPMEAN (See description above).

Each variable can only be marked with Y once on each row. Variables not marked with Y in any of the columns above will be ignored. Additional columns in this file will also be ignored by the program. Variables marked with a Y in more than one column will likely cause an error and provided incorrect results.

An example of a VFILE looks as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | CatVar | Impmean | UseCrit |
| CatVar1 | Y |  |  |
| CatVar2 | Y |  |  |
| StringVar1 | Y |  |  |
| Catvar3 | Y |  |  |
| StringVar2 | Y |  |  |
| ImpMean1 |  | Y |  |
| ImpMean2 |  | Y |  |
| UseCrit1 |  |  | Y |
| UseCrit2 |  |  | Y |

This file indicates that there are nine variables to be recoded. The first five variables will be contract coded (CATVAR), the next two will have missing values imputed with the mean, and the last two variables will be criterion coded.

SHEET (optional; default)

This is the Excel worksheet within the VFILE where the information is stored. If none is specified, the first worksheet in the Workbook is used. Additional sheets in the file will be ignored.

### Example

\* Encoding: windows-1252.

\* Check background data.

include file = "C:\SPSS Macros\RecoderPlus\RecoderPlus.ieasps".

!RecoderPlus

indir = "C:\SPSS Macros\RecoderPlus" /

infile = RecoderPlusSampleData /

outdir = "C:\SPSS Macros\RecoderPlus" /

outfile = RecoderPlusSampleResults /

Contype = d /

vfile = ExampleList /

sheet = Example /

CritVar = CritVar /

IntVarsA = CatVar1 CatVar2 UseCrit1 CatVar1 ImpMean1 ImpMean1 UseCrit1 /

IntVarsB = ImpMean1 StringVar1 CatVar2 UseCrit1 ImpMean2 UseCrit1 UseCrit2 /

IntVec = Cat1Imp1 Cat2Str1 Use1Cat2 Cat1Use1 Imp1Imp2 Imp1Use2 Use1Use2 /

viewcod = y /

qcstats = Y /

clean = y .

# Troubleshooting and Bug Report

To report bugs, or troubleshoot the program, please send details in an email to [egonzalez@ets.org](mailto:egonzalez@ets.org) or [eugene.gonzalez@iea-hamburg.de](mailto:eugene.gonzalez@iea-hamburg.de). When troubleshooting, please set VIEWCOD = Y / and submit to process the syntax. Accompany your email with the corresponding \*.sps and \*.spv files, as well as any additional information you might consider relevant.