User Manual for the PPlus Macro

# About the PPlus Macro

The PPlus macro is a program that computes the proportion correct for each of the items provided in a list of items (ITEMSET), the standard error of the proportion correct for each item, and the proportion correct for the aggregate of the items within the groups defined by the groups specified by the variables specified with the parameter CVARS.

For dichotomous items, the program computes the proportion correct on the item. For polytomous items the program computes the proportion achieving each of the possible score points on the item. For example, for a 2-point item, the program will compute the proportion of respondents obtaining one point, and the proportion of respondents obtaining two points on the item.

The program assumes the item responses have already been scored to values 0, 1… etc., cases have sampling weights (WGT), and information of the assignment of cases to strata (JKZ) and replicates (JKR) for the calculation of jackknife standard errors is available in the file.

There are two versions of this macro, one for SAS and one for SPSS. These two versions have identical functionality.

NOTE: The pplus statistics computed with this program are not necessarily comparable between groups when adaptive designs are used. When adaptive designs are used, an equated pplus should be computed for between group comparisons.

# Requirements

PPlus requires a working copy of SAS or SPSS installed in the computer where the analysis is conducted. It also requires a SAS or an SPSS system file with response data to a test. The responses must be already scored using sequential numeric values starting at zero. Missing or omitted responses should be coded as system missing values. PPlus also requires sampling weights to be included in the data file (WGT) as well as information about the corresponding jackknife strata and replicate to be used for the calculation of the jackknife standard error.

# Using PPlus with SPSS

To use PPlus to process response data, 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:\Temp\TIMSS PValues\pplus.ieasps".

!pplus scale = DataSolveProblems /

indir = C:\Temp\TIMSS PValues /

infile = timss2019grade4\_scr /

outdir = C:\Temp\TIMSS PValues /

outfile = DataSolveProblems /

cvars = idcntry itsex /

selvar = /

selcrit = /

wgt = totwgt /

nrwgt = 150 /

jkz = jkzone /

jkr = jkrep /

itemset = MP51507B MP71175 MP51009 MP71147B MP71187B

MP61211B MP71134B MP71218 MP61069B MP71138B /

maxpts = 1 2 1 1 1 1 1 1 1 1 .

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 session.

When calling the macro, you need to use the exclamation point (!) before PPlus, as in !PPlus. Following !PPlus 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 parameters are mandatory, like the name and location of the input and output file, but many of them are optional and/or have default values. 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.

# Using PPlus with SAS

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

%include "C:\Temp\TIMSS PValues\pplus.mac";

\* PPlus macro call for overall results;

%pplus(scale = DataSolveProblems,

indir = C:\Temp\TIMSS PValues,

infile = timss2019grade4\_scr,

outdir = C:\Temp\TIMSS PValues,

outfile = DataSolveProblems ,

cvars = idcntry itsex,

selvar = ,

selcrit = ,

wgt = totwgt,

nrwgt = 150,

jkz = jkzone,

jkr = jkrep,

itemset = MP51507B MP71175 MP51009 MP71147B MP71187B

MP61211B MP71134B MP71218 MP61069B MP71138B ,

maxpts = 1 2 1 1 1 1 1 1 1 1);

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 session.

Following the call of the macro PPlus you will have a series of parameters within parenthesis, each parameter with the corresponding value(s). Each parameter is separated by a comma. The last parameter ends with a semi-colon (“;”). Some parameters are mandatory, like the name and location of the input and output file, but many of them are optional and/or have default values. 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

There are two output files created by PPlus. These are both created in CSV format and saved to the OUTDIR using the name OUTFILE as the root for the filename, plus a suffix indicating the contents.

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.

When working with SAS, the suffix BATCH or DMS will be added to the CSV files depending on whether the program is being used in batch or interactive mode, respectively.

## [OUTFILE]\_dtl.csv

This file contains one row for each item by combination of groups defined by the CVARS, should these be used. If no CVARS are specified, there will be one row per item. The columns in this file will be the following:

K A constant for all records

[CVARS] When specified, the classification variables.

Item The name for the item followed by the number of score points represented in that record. For example, if there is an ITEM\_A worth up to two points, there will be an entry for ITEM\_A\_1 and one for ITEM\_A\_2. The first entry will contain the weighted pplus for scoring one or better on the item. The second entry will contain the weighted pplus for scoring to or better on the items.

PPlus The weighted pplus for the response represented in the ITEM column.

RPPlus1… RPPlus[NRWGT] The replicated pplus for the response represented in the ITEM column. These are compute using the replicate weights and there will be one entry for each replicate from 1 to NRWGT.

PPlus\_var and PPlus\_se The jackknife variance and jackknife standard errors for the pplus.

Depending on the software used, there will be a corresponding SAS or SPSS data file created with this same information.

## [OUTFILE]\_ttl.csv

This file contains one row for each combination of groups defined by the CVARS, should these be used. If no CVARS are specified, there will be one row with the summary pplus across all items in the ITEMSET. The columns in this file will be the following:

[CVARS] When specified, the classification variables.

PPlus The weighted pplus for the response represented in the ITEM column.

PPlus\_se The jackknifed standard errors for the pplus.

TotalPoints The number of points contributing to the calculation of the average pplus for the group defined by the CVARS.

TotalItems The number of items specified in the ITEMSET parameter.

MaxPoints The maximum number of points that could contribute to the calculation of the pplus for the group defined by the CVARS.

Scale The name of the scale as specified with the parameter SCALE.

In addition to one row for each combination of groups defined by the CVARS, there will also be a table average with the average pplus across all members of the groups defined by the first of the CVARS. For example, when specifying CVARS = COUNTRY SEX, there will be an entry for the simple average across all groups defined by the variable COUNTRY for each value of the variable SEX. If the variable COUNTRY identifies country records in the data, these records will represent a so-called international average for each SEX group.

When specifying CVARS = SEX COUNTRY, there will be an entry for the average across all groups defined by the variable SEX within each group defined by the variable IDCNTRY. These records are identified with a value of 1 in the column TblMn.

Depending on the software used, there will be a corresponding SAS or SPSS data file created with this same information.

## [OUTFILE]\_Output.spv or html

Depending on the software used, there will be an SPV file (when using SPSS) or an HTML file (when using SAS) that displays three sets of results.

The first set of results shows descriptive statistics for the original response variables. In this table you should check that the minimum value is always zero, and the maximum is the maximum possible for that item.

The second set of results shows descriptive statistics for the transformed variables. Since the original response variables are all transformed to dichotomous variables, the min and max should not be below 0 or exceed 1, respectively. When the maximum value exceeds 1, you should check the MAXPTS parameter to confirm the values are set correctly. When the maximum value is 0, you should also check the MAXPTS parameter to confirm the values are set correctly, although you would observe MAX=0 when that score value was not obtained by any respondent in the file. This is an unlikely, yet possible event.

# Parameters

What follows is a list of the parameter that can be used when calling PPlus. They are presented 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 response file. The directory location must be fully spelled out and enclosed in quotes.

Example:

INDIR = "c:\temp\itemanalysis\example" /

INFILE (not Optional; no default)

File with item responses. It does not need the SAV extension for the file name. Do not use quotes.

Example:

infile = SampleDataPIRLS /

OUTDIR (not optional; no default)

Directory where to write the results. The directory location must exist and be accessible from your computer, be fully spelled out and enclosed in quotes. This directory is also used to save many temporary files created by the program, therefore we recommend using a local directory so that the program will run with optimal performance.

Example:

outdir = "c:\Temp\ItemAnalysis\Examples" /

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/SAS variable naming convention.

Example:

outfile = PIRLSExample /

## Response Processing and Coding Parameters

SCALE (not optional; no default)

Name for the test or scale that defines the ITEMSET.

Example:

scale = reading /

ITEMSET (not optional; no default)

The list of variable names for the items to be processed. These need to be listed one after another, with spaces in between. You can use multiple lines, but none should begin in the first column. The last variable name should be followed by a forward slash in the case of SPSS, or a comma in the case of SAS. The items will be read and processed based on the order in which they appear on this list, regardless of the order of the items in the INFILE. Item responses can only be numeric or set to missing, and must be sequential, beginning with zero.

The variable names have to be all written explicitly. No “…TO…”, “… - …” or “… --…” shortcuts to reference variables are allowed.

Example:

ItemSet = /

MAXPTS (optional; default = 1 for all ITEMS)

The maximum number of points achievable for each item. When omitted, it assumes 1 for all items. When specified, there should be one entry per variables specified in the ITEMSET. The value of MAXPTS is used to calculate the p-plus for the response to the item.

Example:

Maxscrs = 1 1 1 1 1 1 1 2 1 3 2 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 /

To verify that the maximum number of points for an item has been entered and applied correctly, you can check the minimum and maximum values for the items listed in the descriptive statistics tables presented in the output.

When the value of MAXPTS has not been specified correctly for an item, the proportion correct for the item will show as 0 or as a very large number in the descriptive statistics table.

CVARS (optional; no default; options = any variable(s) in the INFILE)

A single variable, or a list, that will be used to report item statistics calculated with PPlus. When multiple CVARS are specified, the item statistics are computed for each of the possible groups available in the data crossing the values of the CVARS.

Example:

cvars = idcntry itlang /

SELVAR (optional; no default)

A variable or list of variables used to subset the data read from the INFILE. The variable(s) used to subset the data can be one of the items of the classification variables. Therefore, you only need to specify the SELVAR when this is not specified anywhere else.

Example:

selvar = var1 var2 /

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.

Depending on the platform used, the selection criteria must be specified using SPSS IF/THEN logic, of the SAS WHERE logic, but no IF or WHERE statement is needed. Any, and all selection criteria must fit in a single line.

The selection criteria is applied whenever SELCRIT is not blank, regardless of whether the SELVAR parameter if left blank or not.

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.

## Weighting and Variance Estimation Parameters

WGT (optional; default = TOTWGT)

The name of the variable with the survey or sampling weight. When no WGT is specified, PPlus uses the variable TOTWGT as the sampling weight.

Example:

wgt = TOTWGT /

JKZ (optional; default = JKZONE)

The name of the variable with stratification information to use to create the replicate weights for computing the jackknife variance. This variable should take on values from 1 to the number of sampling strata used for computing the variance for the estimates.

JKR (optional; default = JKREP)

The name of the variable with replication information to use to create the replicate weights for computing the jackknife variance. This variable should take on values of 0 or 1.

JK2TYPE (optional; default = FULL; options = FULL or HALF)

Whether to implement the FULL estimation of the jackknife variance estimation or use the HALF approach. Under the FULL implementation, two replicate weights are created for each jackknife strata. Under the HALF implementation, one replicate weight is created for each jackknife strata. When specifying FULL, the number of replicate weights specified with the parameter NRWGT should be 2x the maximum of the variable specified for JKZ.

NRWGT (optional; default = 150)

The number of replicate weights to use for the estimation of the variance of estimates. When using JK2TYPE = FULL this should be 2x the number of variance estimation strata. When using JK2TYPE = HALF this should be the number of variance estimation strata.

## Additional Parameters

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; only for SPSS)

Instructs the program to delete (CLEAN = Y) the temporary files created during processing. During processing in SPSS, 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. When CLEAN = Y, all temporary files are deleted.

In SAS all temporary files are stored in the WORK library and are deleted when the SAS session is closed.

Example:

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 process the syntax. Accompany your email with the corresponding \*.sps and \*.spv files, as well as any additional information you might consider relevant.

# Examples