

# Documentation

## sas-twophase-package:

### Macro twophase\_bw

Author: D. Ripke, D. Enders, W. Schill  
Date: 04.02.2016  
Version: 1

#### Contents

<b>1</b>	<b>General</b>	<b>2</b>
<b>2</b>	<b>Usage</b>	<b>2</b>
<b>3</b>	<b>Arguments</b>	<b>2</b>
<b>4</b>	<b>Variable Selection</b>	<b>5</b>

## 1 General

The SAS-macro `twophase_bw` is an extension of the macro `twophase`. It conducts a two-phase logistic regression analysis with a variable selection by backwards elimination. The significance level for removing effects and a list of variables to be omitted from the selection process can be specified.

## 2 Usage

The usage of the `twophase_bw`-macro is as follows:

```
%twophase_bw{ folder          = ,
               path_ph1       = ,
               path_ph2       = ,
               methods         = pl_bc,
               suffix          = ,
               compare         = 1,
               outproc         = 0,
               outcorr         = 0,
               outest          = 0,
               caco            = ,
               svar            = ,
               counts_ph1      = ,
               case            = 1,
               control         = 0,
               weights_ph2     = ,
               regr            = ,
               maxit           = 1000,
               epsil           = 1e-10,
               slpr            = 0,
               bw_select       = 1,
               bw_sl           = 0.05,
               bw_omit         = ,
               bw_out          = 1 }
```

## 3 Arguments

The following parameters can be set:

Parameter	Description
<code>folder</code>	Path of the folder containing all macros called by <code>twophase</code> . <b>Value:</b> string <b>Default:</b> none

<code>path_ph1</code>	Path of the phase one dataset. The value is of the form <code>path_ph1=libname.filename</code> . If the dataset is located in the workspace, <code>libname</code> can be omitted. <b>Value:</b> string <b>Default:</b> none
<code>path_ph2</code>	Path of the phase two dataset. See <code>path_ph1</code> for details. <b>Value:</b> string <b>Default:</b> none
<code>methods</code>	Specifies the estimation methods. Valid names are <code>m1_em</code> , <code>m1_sw</code> , <code>p1_bc</code> , <code>p1_sch</code> , <code>w1</code> and <code>s2</code> . Separate different names by blanks. The corresponding estimation methods are: <code>m1_em</code> - Maximum Likelihood via EM algorithm <code>m1_sw</code> - Maximum Likelihood (Profile Likelihood, Scott/Wild) <code>p1_sch</code> - Pseudo Likelihood (Schill) <code>p1_bc</code> - Pseudo Likelihood (Breslow-Cain) <code>w1</code> - Weighted Likelihood <code>s2</code> - Sample 2 (complete case) analysis <b>Value:</b> string <b>Default:</b> <code>p1_bc</code>
<code>suffix</code>	String to append on name of output dataset. <b>Value:</b> string <b>Default:</b> none
<code>compare</code>	If not set to zero, the estimated regression parameters and standard errors of all chosen methods are directed to the SAS-output. <b>Value:</b> numeric <b>Default:</b> 1
<code>outproc</code>	If not set to zero, parameter estimates, standard errors and confidence intervals of every method are directed to the SAS-output. <b>Value:</b> numeric <b>Default:</b> 0
<code>outcorr</code>	If not set to zero, the correlation matrix of every estimation method is directed to the SAS-output. Does only have an impact if <code>&amp;outproc=1</code> . <b>Value:</b> numeric <b>Default:</b> 0
<code>outest</code>	If not set to zero, the estimates and standard errors of the phase one parameters are directed to the SAS-output, additional to the regression parameters. Does only have an impact if <code>&amp;compare=1</code> . <b>Value:</b> numeric <b>Default:</b> 0
<code>caco</code>	Name of outcome variable. <b>Value:</b> string <b>Default:</b> none

<code>svar</code>	Name of stratum variable. <code>svar</code> must take on all the values <b>1, 2, . . . , J</b> where <b>J</b> is the number of strata. <b>Value:</b> string <b>Default:</b> none
<code>counts_ph1</code>	Variable containing the observation counts in phase one. <b>Value:</b> string <b>Default:</b> none
<code>case</code>	Value of <code>&amp;caco</code> (Case). <b>Value:</b> numeric <b>Default:</b> 1
<code>control</code>	Value of <code>&amp;caco</code> (Control). <b>Value:</b> numeric <b>Default:</b> 0
<code>weights_ph2</code>	Optional variable in phase two containing weights for each observation. <b>Value:</b> string <b>Default:</b> none
<code>regr</code>	Names of the regression variables. The names must be separated by blanks. <b>Value:</b> string <b>Default:</b> none
<code>maxit</code>	Maximum number of iterations (whenever iterations occur in the estimation methods). <b>Value:</b> positive integer <b>Default:</b> 1000
<code>epsil</code>	Accuracy of calculations in the estimation methods. <b>Value:</b> numeric <b>Default:</b> $10^{-10}$
<code>s1pr</code>	Indicates whether sampling in phase one is prospective or retrospective: 1-prospective, 0-retrospective. <b>Value:</b> 0/1 <b>Default:</b> 0
<code>bw_select</code>	If not set to zero, a backwards elimination variable selection is conducted. <b>Value:</b> numeric <b>Default:</b> 1
<code>bw_sl</code>	Significance level for removing effects. <b>Value:</b> numeric <b>Default:</b> 0.05
<code>bw_omit</code>	Lists the variables that are to be omitted from the selection process. Separate different names by blanks. Must be a subset of <code>&amp;regr</code> . <b>Value:</b> string <b>Default:</b> none

---

<b>bw_out</b>	Indicates whether the eliminated variables and their respective p-values at elimination are directed to the SAS-output: 1=yes. <b>Value:</b> numeric <b>Default:</b> 1
---------------	--

---

## 4 Variable Selection

The `twophase_bw`-macro extends the `twophase`-macro by combining the regression analysis for every estimation method specified in `&methods` with a backwards elimination, a stepwise variable selection, in order to exclude non-influential effects from the analysis. We focus on the features expanding the basic macro while a detailed explanation of the `twophase`-macro itself may be found in its documentation `TPDocu.pdf`.

Whether the selection process is actually conducted or not is controlled by the macro parameter `&bw_select`. At `&bw_select=0` the macro jumps directly to the regression analysis done by the `twophase`-macro using its original arguments.

The backwards elimination starts with the full model, i.e. the `twophase`-macro conducts a regression analysis with all variables listed in `&regr`. Its estimated parameters are used in calculating for every variable the p-value for a Wald test with the null hypothesis that the variable has no effect on the outcome. The purpose is to identify the variable with the maximum p-value. As the variables listed in `&bw_omit` are chosen to stay within the analysis under any circumstance and as such are exempted from the selection process, their p-values are excluded when the largest p-value is identified. Naturally, `&bw_omit` has to constitute a subset of `&regr`.

If the maximum p-value is greater than or equal to `&bw_sl`'s value, its respective variable is added to the list of eliminated variables and removed from the model.

In the next step of the backwards elimination, the `twophase`-macro uses this reduced model to once again conduct a regression analysis and the procedure is repeated until the maximum p-value drops below `&bw_sl`, at which point the selection process stops and the `twophase`-macro performs the final regression analysis with those variables remaining in the model.

Finally, the parameter `&bw_out` controls whether the list of eliminated variables and their respective p-values at elimination is directed to the SAS-output. In case of `&bw_out=1`, depending on the number of estimation methods in `&methods` and the values submitted for `&outproc` and `&compare`, the lists are either displayed separately after the results of each method's analysis or together in a combined table at the end.

If no variables are eliminated in any method, a notification to that effect is directed to the SAS-output instead.