Macro Risk_Est_PH_reg_prop_score User's Guide

Michael Crager
7 October 2022

Macro Risk_Est_PH_reg_prop_score

Macro Risk_Est_PH_reg_prop_score computes risk estimates from a proportional hazards regression with IPTW weighting based on propensity scores, accounting for the variability in propensity score estimation using logistic regression as described in Crager (2022). Multiple treatments are accommodated. The Cox regression may include "external" time-dependent covariates, stratified cohort sampling weights, and left truncation. The macro also provides regression parameter and hazard ratio estimates with appropriate variance estimation and diagnostics for covariate balance across treatments using the IPTW weights.

The macro is called as follows:

```
%Risk_Est_PH_reg_time_dep(
    /* Input Specification */  indsn=, byvar= , vars=, vars_logistic=,
                    time=, censor=, censorlist=,
                    entrytime=, weight=,
                    response=, stabilize=, truncate_pct=,
                    programming_statements=%str(), calc_vars=, covariate_dsn=,
    /* Analysis Parameters */  risk_time=, print_phreg=, print_logistic=,
                                alpha=, strata=, CI_method=
    /* Output Specification */ outdsn=, Risk=, Risk_LCL=, Risk_UCL=,
                      CumHaz, CumHaz_LCL=, CumHaz_UCL=,
                      LogCumHaz=, SE_LogCumHaz=,
                      Parameter_estout=, parameter_covout=,
                    std_diff_vars=,
                    std_diff_out=, std_diff_graph_name= ,graph_path=
            );
```

The macro parameters are defined in Table 1. The time dependence of the covariates, if any, is defined using programming statements (as are used for PROC PHREG). It is assumed that these programming statements, when applied to the input data set (and the covariate data set, if specified), will uniquely determine the covariate values at the time specified by the input data set

variable given by the macro parameter time. Note the programming statements will need to be enclosed in %str() so that the semicolon(s) will not cause a syntax error.

Table 1. Macro Risk_Est_PH_reg_prop_score Parameters

Parameter	Type	Required?	Default Value	Description
indsn	$	Yes	(at temporary library)	(Libname reference and) file name containing input data set.
byvar	$	No	—	Optional list of variables to do the analysis by.
vars	$	Yes	—	List of input data set variables to be used as the covariate in the Cox model used to estimate the risk. If the programming statements create the variables that are to be included in the model, list the variables thus created along with any time-invariant covariates.
vars_logistic	#	Yes	—	List of input data set variables to be used as the covariates in the logistic regression model for estimating the propensities.
time	#	Yes	—	Input data set variable containing the time to event (or censoring).
censor	#	Yes	—	Input data set variable indicating whether the observed time to event was censored.
censorlist	#	No	0	List of values of variable censor that indicate a censored observation. Default is the single value 0.
entrytime	#	No	—	Optional input data set variable containing a left truncation time for each observation.
weight	#	No	—	Input data set variable giving the observation's sampling weight if cohort sampling was used.
response	$/#	Yes	—	Input data set variables giving the response for which the propensity will be estimated (such as treatment or biomarker use). This may be a binary or multinomial outcome.
stabilize	$	No	yes	If this parameter is set to yet, stabilized propensity score weights will be used (multiplying the inverse probability by the proportion of patients with the response).
truncate_pct	#	No	0	If this parameter is set to a non-negative number, the propensity score weights will be truncated the &truncate_pct and 100-&truncate_pct percentiles. The percentage will be rounded to the nearest tenth of a percent. The percentage specified must be non-negative and less than 50. A typical value is 5 (truncating at the 5th and 95th percentiles).

Table 1. Macro Risk_Est_PH_reg_prop_score Parameters

Parameter	Type	Required?	Default Value	Description
programming_ statements	$	No	—	%str()-enclosed text string including programming statements that will be inserted into proc PHREG and various data steps to compute the time-dependent covariate values. For example: programming_statements = %str(if time <= 3 then x_3 = 0; else x_3 = x;) If no programming statements are entered, the risk calculations will be made for covariates that are constant over time.
calc_vars	$	Yes, if time-dependent covariates are used	—	List of variables that are used in the calculation of the time-dependent covariate values. Include all non-time-dependent covariate values in this list, too. Leave this parameter blank if time dependent variables are not used.
covariate_dsn	$	No	(input data set)	(Libname reference and) the name of a data set that contains the covariate values for which the risk is to be estimated. The data set must have all the variables included in the model, or that are required to derive these variables if the model has time-dependent covariates derived using programming statements. The data set must also include the stratification variable if the model is stratified. If no covariate data set is specified, the risk will be estimated for every patient in the main input data set.
risk_time	#	Yes	—	This is the time at which the risk is assessed for each patient. That is, the risk is defined as the probability that the patient will have the event on or before risk_time.
print_phreg	$	No	yes	If this parameter is set to no, output from the PHREG model fit will not be printed. The variability estimates and confidence intervals will account for the weight estimation variability.
print_logistic	$	No	yes	If this parameter is set to no, the PROC LOGISTIC output will not be printed.
alpha	#	No	0.05	The macro will compute a 100(1-alpha)% confidence interval for the risk and cumulative hazard.
strata	$/#	No	—	Character string giving input data set variable by which the proportional hazards regression analysis will be stratified.

Table 1. Macro Risk_Est_PH_reg_prop_score Parameters

Parameter	Type	Required?	Default Value	Description
CI_method	$	No	loglog	Character string giving the method for computing the confidence intervals. If linear is specified, the confidence interval is computed on the risk scale. If log is specified, the confidence interval is computed on the cumulative hazard scale and transformed to the risk scale. If loglog is specified, the confidence interval is computed on the log cumulative hazard scale and transformed to the risk scale.
outdsn	$	Yes	(at temporary library)	(Libname reference and) the output data set name. This data set will contain all the records and variables of the covariate data set (or the input data set if no separate coavariate data set is specified) plus the variables named by the following eight macro parameters.
Risk	#	No	Risk	Name of output data set variable that will contain the risk estimate.
Risk_LCL	#	No	Risk_LCL	Name of output data set variable will contain the lower limit of a 1-alpha confidence interval for the risk.
Risk_UCL	#	No	Risk_UCL	Name of output data set variable will contain the upper limit of a 1-alpha confidence interval for the risk.
CumHaz	#	No	CumHaz	Name of output data set variable that will contain the cumulative hazard estimate.
CumHaz_LCL	#	No	CumHaz_LCL	Name of output data set variable will contain the lower limit of a 1-alpha confidence interval for the cumulative hazard.
CumHaz_UCL	#	No	CumHaz_UCL	Name of output data set variable will contain the upper limit of a 1-alpha confidence interval for the cumulative hazard.
LogCumHaz	#	No	LogCumHaz	Name of output data set variable that will contain the log cumulative hazard estimate.
SE_LogCumHaz	#	No	SE_LogCumHaz	Name of output data set variable that will contain the estimated standard error of the log cumulative hazard estimate.
Parameter_estout	$	No	(at temporary library)	Optional libname reference and data set name that will contain the Cox regression parameter estimates, standard errors, chi-square statitistics, p-values, hazard ratio estimates and confidence intervals, all computed using the covariance matrices accounting for variability in the propensity score-based weight estimates.
Parameter_covout	$	No	(at temporary library)	Optional libname reference and name of data set that will contain the Cox regression parameter estimate covariance matrix.

Table 1. Macro Risk_Est_PH_reg_prop_score Parameters

Parameter	Type	Required?	Default Value	Description
std_diff_vars	$	No	—	Optional list of variables for which the standardized differences among logistic| regression outcomes will be assessed.
Std_diff_out	$	No	—	Optional libname reference and name of data set that will contain the weighted and| unweighted standardized differences among the logistic regression response outcomes for each logistic regression covariate.
Std_diff_graph_name	$	No	—	Optional name for standardized difference graphs. The graphs will be placed in| the directory named in parameter graph_path and suffixes of the form i_j will be affixed to the graph name specifying the number logistic regression outcomes being compared in each graph.
Graph_path	$	No	—	Optional specification of folder in which to place the standardized difference graphs.

References

Crager MR (2022). Accounting for propensity score variability in IPTW weighted Cox proportional hazards regression and risk estimaton. *American Journal of Applied Mathematics*. In press.