WUSS 2019 Paper Proposal

Title (250 characters or less):

Automating a Summary Report of PSM Model and Match Results

Abstract (250 words or less):

When performing propensity score modeling and matching, various output is produced and located either in separate reports or datasets. As you iterate the process after adjusting the model or match inputs, it is helpful to have the key results in one report for easy comparison with later iterations. This paper demonstrates a method to automate the creation of a report to summarize key results using the ODS output from PROC LOGISTIC and PROC PSMATCH. The intended audience are those with basic familiarity with SAS/STAT software.

Authors names, affiliations, email addresses, short biographies (75 words or less):

- Desiree Hall, Optum, desiree.hall@optum.com
  Desiree is a principal data analyst within Optum Analytics. She has worked as an analyst using SAS products for 20 years in the healthcare and financial industries. She is a SAS Certified Professional v6 and a SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling.
- Blake Krebs, Optum, blake.krebs@optum.com
  Blake is a principal data analyst within Optum Analytics. He has extensive experience as an analyst in the healthcare and financial industries. His current projects involve the integration of unstructured data from electronic health records with health claims data to gain deeper insight to the patient journey.

Working draft or outline:

During propensity score matching, the analyst runs a logistic regression and uses the resulting score (propensity score) to perform a match of the most similar patients between their treated and control groups. Usually the analyst will iterate through these steps multiple times in order to build and fine-tune their model and the match. In order to help streamline the iterative process, this paper focuses on summarizing the output created during the PROC LOGISTIC and PROC PSMATCH using ODS. Reports can then be compared between iterations as the analyst fine-tunes their analysis and weighs the merits of each model and match combination.

The following reports are created by PROC LOGISTIC and PROC PSMATCH. *<Show examples of the output reports>*. By using the ODS OUTPUT statement, we will ask SAS to hold onto these tables for the course of our work session, and we will read them into a report dataset.

We will build a summary report for this run of our propensity score model and match with the following report sections:

Model Build

Prior to your PROC LOGISTIC step, specify the ODS OUTPUT statement. Include the following output tables: ParameterEstimates, FitStatistics, ModelInfo, ResponseProfile, Association, OddsRatios, GlobalTests, and NObs. Calling out these tables will keep them active within your work session.

The ResponseProfile table provides your treated and comparison cohort labels. The ModelInfo table provides the model method. The NObs table provides the number of observations read and used.

Model Fit and Test of Global Null

The FitStatistics table provides the AIC, SC, and -2 Log L values of your model. The GlobalTests table provides the results of the tests of the global null.

Model Metrics

The Association table provides your logistic model performance metrics such as concordant/discordant/ties, and the c-stat.

Control Group Match

The DataInfo and MatchInfo tables are created during PROC PSMATCH. These will show the beginning and final matched set counts, the matching method, the caliper, and the absolute final difference.

Model Variables

The final report shows the combined estimates and odds ratios for each covariate used within your model. These are taken from the PROC LOGISTIC ParameterEstimates and OddsRatios tables, and then merged together in one dataset for easy viewing.

The code example at the end of this paper shows how all of these tables are combined to create two datasets, one containing the model and match information, and the other containing the model covariates. These two datasets can then be simply printed using PROC PRINT as your summary report. Once you have reviewed your model and match thoroughly, you can then repeat the process, produce this report again, and compare reports between the iterations.

Next Steps

An important next step in evaluating a matched group is balance assessment. See this PharmaSUG white paper for a method to evaluate balance:
http://www.pharmasug.org/proceedings/2014/SP/PharmaSUG-2014-SP07.pdf

Code:

```
*SET MODEL PARAMETERS*;
%let TBRAND          =AMD_TREATED;                    *** treatment group name;
%let CBRAND          =AMD_UNTREATED;                  *** control group name;
%put _user_;

*** Classification variables for the model;
%let in_class= pre_ind_visit_er ind_male ind_ins_commercial ind_race_caucasian
ind_region_northeast pre_charlson_metastatic;

*** Continuous variables for the model;
%let in_cont=age PRE_OFFICE_VISITS PRE_CHARLSON_SCORE;

*** Model variables you want assessed by PROC PSMATCH;
%let key_match= ind_male ind_ins_commercial age pre_charlson_score pre_ind_visit_er;

*** Logistic Regression Model;

ods output ParameterEstimates=estimates FitStatistics=FitStats ModelInfo=ModelInfoRpt
ResponseProfile=ResponseProfileRpt Association=AssociationRpt OddsRatios=OddsRatios
GlobalTests=GlobalTests NObs=NObs;

proc logistic data=pat_model_vars  outmodel=model_v1 outest=scores_est namelen=30;
  class index_prod &in_class ;
  model index_prod (event="&TBRAND.")= &in_class &in_cont / ;
  score out=scores_v1;
run;

data propensityscores;
  set scores_v1 (rename=(P_&TBRAND. = P_1));
  if index_prod="&TBRAND." then test_patient=1;
  else test_patient=0;
  prob=p_1;
run;

*** Propensity Score Match;

ods graphics;
ods output MatchInfo=matchinfo PSInfo=psinfo DataInfo=datainfo;

proc psmatch data=propensityscores region=treated;
    class index_prod &in_class ;
    psdata ps=prob treatvar=index_prod(treated="&TBRAND.");
    match method=greedy(k=1) caliper(mult=stddev)=0.2;
    assess lps var=(&key_match)/ weight=none stddev=pooled(allobs=NO);
    output out(obs=match)=out_psmatch_v1 matchid=_MatchID;
run;
ods graphics off;


*** Model and Match Report;
```

```
proc sql;
create table model_input_vars as
        select distinct _NAME_ as cValue1
        from model_v1
        where _TYPE_ in ('G','Z');
quit;

*** set estimates and odds ratios next to each other for each variable;

data OddsRatios;
  set OddsRatios;
  Effect=upcase(Effect);
run;
proc sort data=OddsRatios;
  by Effect;
run;
data estimates;
  set estimates;
  Variable=upcase(Variable);
run;
proc sort data=estimates;
  by Variable;
run;
data rpt_variables;
  merge estimates (in=ina drop=DF _ESTTYPE_)
        OddsRatios (in=inb rename=(Effect=Variable));
  by Variable;
  if ina or inb;
run;

*** set the model and match metrics into one dataset;

data model_rpt;
  length label1 rptSection $50 cValue1 cValue2 $80;
  set
    ResponseProfileRpt (in=ina keep=Outcome rename=(Outcome=Label1 ))
    ModelInfoRpt (in=ind keep=Description Value rename=(Description=Label1 Value=cValue1)
                      where=(Label1 in ('Model','Optimization Technique')))
    NObs (in=inj keep=Label N)
    FitStats (in=inc
        rename=(Criterion=Label1 InterceptOnly=nValue1 InterceptAndCovariates=nValue2))
    GlobalTests (in=ini)
    AssociationRpt (in=ine keep=Label1 cValue1)
    AssociationRpt (in=inh keep=Label2 cValue2)
    datainfo (in=inf keep=Label1 cValue1
        where=(label1 in ('All Obs (Treated)','All Obs (Control)')))
    matchinfo (in=ing keep=label1 cValue1
        where=(label1 not in ('Order','Matched Obs (Treated)','Matched Obs (Control)')));
  *** reformat rows to align properly in the report;
  if ina and label1="&TBRAND" then do;
      rptSection='Model Build';
      label1='Treated Cohort';
      cValue1="&TBRAND";
  end;
  else if ina and label1="&CBRAND" then do;
      rptSection='Model Build';
      label1='Comparison Cohort';
      cValue1="&CBRAND";
  end;
  else if inc then do;
      cValue1=put(nValue1,8.3);
```

```
        cValue2=put(nValue2,8.3);
        rptSection='Model Fit';
  end;
  else if inh then do;
        cValue1=cValue2;
        label1=label2;
        cValue2=.;
        rptSection='Model Metrics';
  end;
  else if ini then do;
    label1=Test;
        cValue1=put(ChiSq,8.3);
        cValue2=put(DF,8.0);
        rptSection='Model Test Global Null';
  end;
  else if inj then do;
        label1=Label;
        cValue1=put(N,8.0);
        rptSection='Model Build';
  end;
  else if ind then rptSection='Model Build';
  else if ine then rptSection='Model Metrics';
  else if inf or ing then rptSection='Control Group Match';
run;

proc print data=model_rpt noobs;
  var rptSection label1 cValue1 cValue2 ProbChiSq;* StdErr ClassVal0 Estimate WaldChiSq
OddsRatioEst LowerCL UpperCL;
  where label1 ne 'All Candidate Variables';
  title 'Model and Match Report';
run;

proc print data=rpt_variables noobs;
  title 'Model Variables';
run;
```

Sample Report Output:

Model and Match Report				
rptSection	**label1**	**cValue1**	**cValue2**	**ProbChiSq**
Model Build	Treated Cohort	AMD_TREATED		.
Model Build	Comparison Cohort	AMD_UNTREATED		.
Model Build	Model	binary logit		.
Model Build	Optimization Technique	Fisher's scoring		.
Model Build	Number of Observations Read	6011		.
Model Build	Number of Observations Used	6011		.
Model Fit	AIC	8225.804	7774.711	.
Model Fit	SC	8232.505	7841.724	.
Model Fit	-2 Log L	8223.804	7754.711	.
Model Test Global Null	Likelihood Ratio	469.093	9	<.0001
Model Test Global Null	Score	425.263	9	<.0001
Model Test Global Null	Wald	380.755	9	<.0001
Model Metrics	Percent Concordant	64.9		.

rptSection	label1	cValue1	cValue2	ProbChiSq
Model Metrics	Percent Discordant	35.1		.
Model Metrics	Percent Tied	0.0		.
Model Metrics	Pairs	8869410		.
Model Metrics	Somers' D	0.299	.	.
Model Metrics	Gamma	0.299	.	.
Model Metrics	Tau-a	0.147	.	.
Model Metrics	c	0.649	.	.
Control Group Match	All Obs (Treated)	2601		.
Control Group Match	All Obs (Control)	3410		.
Control Group Match	Distance Metric	Logit of Propensity Score		.
Control Group Match	Method	Greedy Matching		.
Control Group Match	Control/Treated Ratio	1		.
Control Group Match	Caliper (Logit PS)	0.12182		.
Control Group Match	Matched Sets	2582		.
Control Group Match	Total Absolute Difference	160.1091		.

Model Variables

Variable	ClassVal0	Estimate	StdErr	WaldChiSq	ProbChiSq	OddsRatioEst	LowerCL	UpperCL
AGE		0.0476	0.00390	148.8802	<.0001	1.049	1.041	1.057
IND_INS_COMMERCIAL	0	0.1462	0.0443	10.9086	0.0010	1.340	1.126	1.594
IND_MALE	0	0.0173	0.0281	0.3776	0.5389	1.035	0.927	1.156
IND_RACE_CAUCASIAN	0	-0.1430	0.0459	9.7025	0.0018	0.751	0.628	0.899
IND_REGION_NORTHEAST	0	0.2359	0.0365	41.7816	<.0001	1.603	1.389	1.850
INTERCEPT		-4.1240	0.3136	172.9093	<.0001	.	.	.
PRE_CHARLSON_METASTATIC	0	-0.3869	0.1222	10.0187	0.0015	0.461	0.286	0.745
PRE_CHARLSON_SCORE		-0.0657	0.0174	14.3414	0.0002	0.936	0.905	0.969
PRE_IND_VISIT_ER	0	0.0926	0.0309	8.9699	0.0027	1.204	1.066	1.359
PRE_OFFICE_VISITS		0.0173	0.00212	66.8423	<.0001	1.017	1.013	1.022