Paper 231-2008

A Macro For Getting More Out Of Your ROC Curve

Jennifer Lambert, Eli Lilly and Company, Indianapolis, IN
Ilya Lipkovich, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

As a part of clinical decision making, a continuous measure reflective of a patient's condition (e.g. laboratory test result) is often dichotomized (or "cut") into two groups (e.g., abnormal and normal) for predictive or screening purposes. The decision of where to select the cut-off point is often governed by a reasonable trade-off between sensitivity and specificity for a particular test. A common method to help find this balance is to plot sensitivity versus (1-specificity) as a "ROC" (Receiver Operator Characteristic) curve. Although a useful tool, the ROC curve rarely displays the individual cut-off values. Because of this limitation, the user cannot visualize the impact of varying sensitivity and specificity against the cut-off values of the prediction variable.

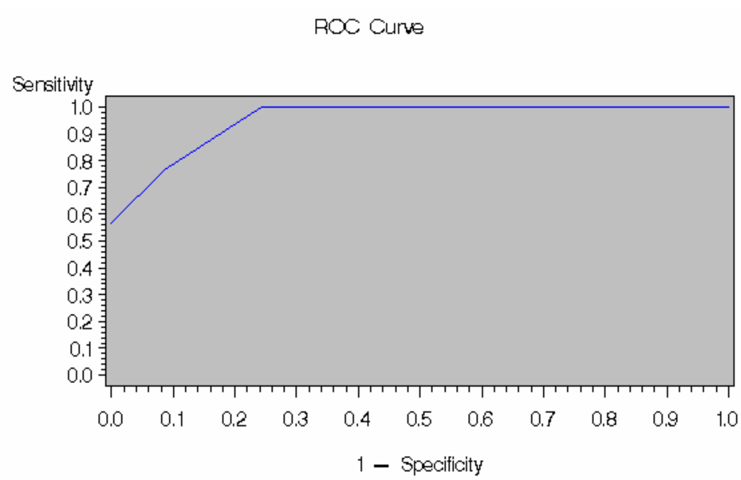
The macro presented in this paper creates a single graph that, for a given prediction variable and a binary outcome ("true condition"), simultaneously displays the following: sensitivity, specificity, Youden Index, and various user-defined measures of misclassification error against cut-off values of the prediction variable. The graph is a simple but highly informative visual tool that provides the user with greater functionality than a standard ROC curve. This macro is intended for an intermediate SAS user with PC SAS (SAS/STAT® and SAS/GRAPH®) version 8.2 or 9.1 capabilities.

INTRODUCTION

In diagnostic or predictive testing, sensitivity and specificity are used to describe how well a test discriminates between subjects (cases) with and without a certain condition. Usually, a continuous variable, *X*, is made categorical by selecting a cut-off point that results in reasonably high values for both measures. However, depending on the conditions of the test, historical considerations, or other non-data driven factors, it might be more desirable to select a higher or lower cut-off value.

The exploration to find an optimal cut-off value can begin with assessing the relative importance attached to false positives and false negatives. As sensitivity is increased, more cases with a certain condition can be identified. But as sensitivity increases, accuracy is sacrificed on identifying those without the condition (specificity). As part of the process of determining an optimal cut-off point, a Receiver Operating Characteristic curve (or ROC curve) is usually constructed (shown below). It is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for various cut-off values of *X*. The ROC curve provides a visual demonstration of:

- the trade-off between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity). The better the classification, the closer the curve will be to the upper left corner of the plot.
- the Area Under the Curve (AUC), which is a way to quantify the discriminatory ability of the test. An area of 1.0 corresponds to an ideal test since it achieves both 100% sensitivity and 100% specificity. On the other extreme, an area of 0.5 means that the test is no better than random classification. For example, tossing a coin with a fixed probability of heads (p), yields sensitivity =1-specificity = p (e.g. when p= 0.5 both sensitivity and specificity = 50%).



While each point on the ROC curve corresponds to a specific cut-off criterion for the prediction variable, it is not possible to see how the values of *X* vary as sensitivity (and specificity) change. The %SNSP_TRADEOFF macro presented here attempts to bridge that gap by providing a flexible graphical tool to visualize the impact of changing the cut-off values of a continuous prediction variable on the levels of sensitivity, specificity, and several additional popular operating characteristic measures.

REVIEW OF CONCEPTS AND TERMINOLOGY

		TRUE Condition (Outcome)		
		Positive	Negative	*Total*
TEST Result (Predicted)	Positive	True Positive *(TP)*	False Positive *(FP)*	*(TP+FP)* = all positive tests
	Negative	False Negative *(FN)*	True Negative *(TN)*	*(FN+TN)* = all negative tests
	Total	*(TP+FN)* = all true positives	*(FP+TN)* = all true negatives	*N* (total sample size)

In the table above, the columns represent the categories of the true or actual condition. The rows represent the predicted categories. The terms *positive* and *negative* can be used interchangeably with *present/absent*, *yes/no*, *normal/abnormal*, *etc.*, as is applicable to the testing situation.

Sensitivity (SN) = $\frac{TP}{TP+FN}$. Ability of the test to correctly identify those cases with the condition.

Specificity (*SP*) = $\frac{TN}{FP+TN}$. Ability of the test to correctly identify those cases that do not have the condition.

Positive Predictive Value (PPV) = $\frac{TP}{TP+FP}$. The proportion of cases that have the condition among those classified with the condition.

Negative Predictive Value (NPV) = $\frac{TN}{FN+TN}$. The proportion of cases that do not have the condition among those classified without the condition.

Total Accuracy (TA) = $\frac{TP+TN}{N}$. The proportion of cases whose tests accurately predict the true outcome.

*Youden Index (J) = SN+SP-*1. Commonly used overall measure of test accuracy (+1 is perfect prediction).

Matthews Correlation Coefficient (MCC) = $\frac{TPTN - FPFN}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}}$.
Another overall measure of test accuracy (+1 is perfect prediction).

METHOD

Diagnostic measurements can be described in terms of maximizing accuracy (e.g., Youden Index, Matthew's Correlation Coefficient, and Total Accuracy) or minimizing misclassification errors. The flexibility of the %SNSP_TRADEOFF macro allows you to specify the output in terms of measures of accuracy, measures of misclassification errors, or combinations of both. This macro uses simple logistic regression (PROC LOGISTIC) to compute all operating characteristics. The misclassification errors used by this macro are based on two groupings, Total Misclassification Errors and Weighted Misclassification Errors. Their calculations are as follows:

- Total Misclassification Error (*TME*), expressed as

$$TME = FN{*}L_{FN} + FP{*}L_{FP} \tag{1}$$

where L_{FN} = the losses associated with each false negative
L_{FP} = the losses associated with each false positive.

- Weighted Misclassification Error (*WME*), expressed as

$$WME = \frac{FN + W_{FN}}{N} + \frac{FP + W_{FP}}{N} \tag{2}$$

where weights are based on the relative losses as follows

$$W_{FN} = \frac{L_{FN}}{L_{FN} + L_{FP}}, W_{FP} = \frac{L_{FP}}{L_{FN} + L_{FP}} \tag{3}$$

When the two losses ($L_{FN,}$ L_{FP}) are both proportional to the inverse of their respective true classification groups ($\frac{1}{TP+FN}$, $\frac{1}{FP+TN}$), the *TME* from (1) becomes

$$TME_0 = \frac{FN}{TP+FN} + \frac{FP}{FP+TN}$$, which is the inverse of the Youden Index.

Hence, a minimization of *TME* when losses due to false negative and false positive misclassifications are inversely proportional to their true classification group is equivalent to a maximization of the Youden Index. And when the two losses ($L_{FN,}$ L_{FP}) are proportional to the inverse of the overall sample size (1/*N*), the *TME* from (1) becomes

$$TME_1 = \frac{FN+FP}{N}$$, which is the inverse of Total Accuracy.

Hence, a minimization of *TME* when losses due to false negative and false positive misclassifications are equal and inversely proportional to the sample size is equivalent to a maximization of *TA*. The similar derivation can be used to calculate WME_0 and WME_1, by inputting the appropriate loss definition (inverse of classification group size or inverse of overall sample size) in (3) to define (2). Furthermore, both *TME* and *WME* can be expressed by a ratio of the two losses as follows:

$$\gamma = \frac{L_{FN}}{L_{FP}}$$

For example, if γ=2 then the loss due to misclassification of a true 'positive' outcome as a 'negative' is twice that of misclassifying a true 'negative' outcome as a 'positive' test result. Using the ratio method for loss, the *WME* becomes

$$WME_2 = \frac{FN}{N}\frac{\gamma}{1+\gamma} + \frac{FP}{N}\frac{1}{1+\gamma}$$

and the *TME* becomes

$$TME_2 = \frac{FN}{N} + \frac{FP}{N}\frac{1}{\gamma}$$, where the total loss, L_{FN}, is fixed at 1/*N*.

For the ease of graphical presentation, TME_2 is rescaled to fit the range of 0 to 1 by dividing each individual TME_2 value by the maximal TME_2 value.

In summary, losses due to false negative and false positives are calculated by:

- Equal sizes, based upon the inverse of the total sample size (1/*N*).
- Inverse to the true classification group sizes: 1/(*TP*+*FN*) and 1/(*FP*+*TN*), respectively.
- User-defined ratio, $\gamma = \frac{L_{FN}}{L_{FP}}$.

EXAMPLE

The following example illustrates the data set structure, macro call, and graphical output produced by the %SNSP_TRADEOFF macro. In this example, researchers are studying a mentally ill patient population. They have obtained clinical test scores from the patients during the duration of treatment. The researchers are interested in how a change in test scores expressed early during treatment might help to predict successful long-term response to treatment. Specifically, they aim to identify a single value where a change in test score below this value suggests that a patient is not likely to respond to treatment and a change above this value suggests that a patient is likely to respond to treatment.

Below is the SAS data set used by the macro called "cutoff_search". Notice that the data is structured so that there is one row per patient. The only additional variables required are the binary outcome (RESULT) and the continuous prediction variable (SCORE_CHANGE).

PATIENT	SCORE_CHANGE	RESULT
101	-4	No Response
102	22	No Response
103	15	No Response
104	22	Response
105	28	Response
.	.	.
.	.	.
.	.	.
900	34	Response

The operating characteristics that are output by default are sensitivity, specificity, and AUC, as these are usually displayed in standard ROC output. For this example, the additional operating characteristics of total accuracy, the Youden Index, negative predictive value, positive predicted value, and Matthews Correlation Coefficient are presented. Additionally, the total and weighted misclassification errors expressed by the user-defined ratio (γ) have also been chosen (TME_2 and WME_2, respectively). The choice for the value of γ was selected to be < 1. The researchers decided to give more weight to the loss due to misclassification of false positives. Patients who are misclassified early as being successfully treated might fail to receive potentially beneficial interventions.

Below is the macro call used for this example along with a brief explanation of the macro parameters (full details and macro source code are given in the Appendix).

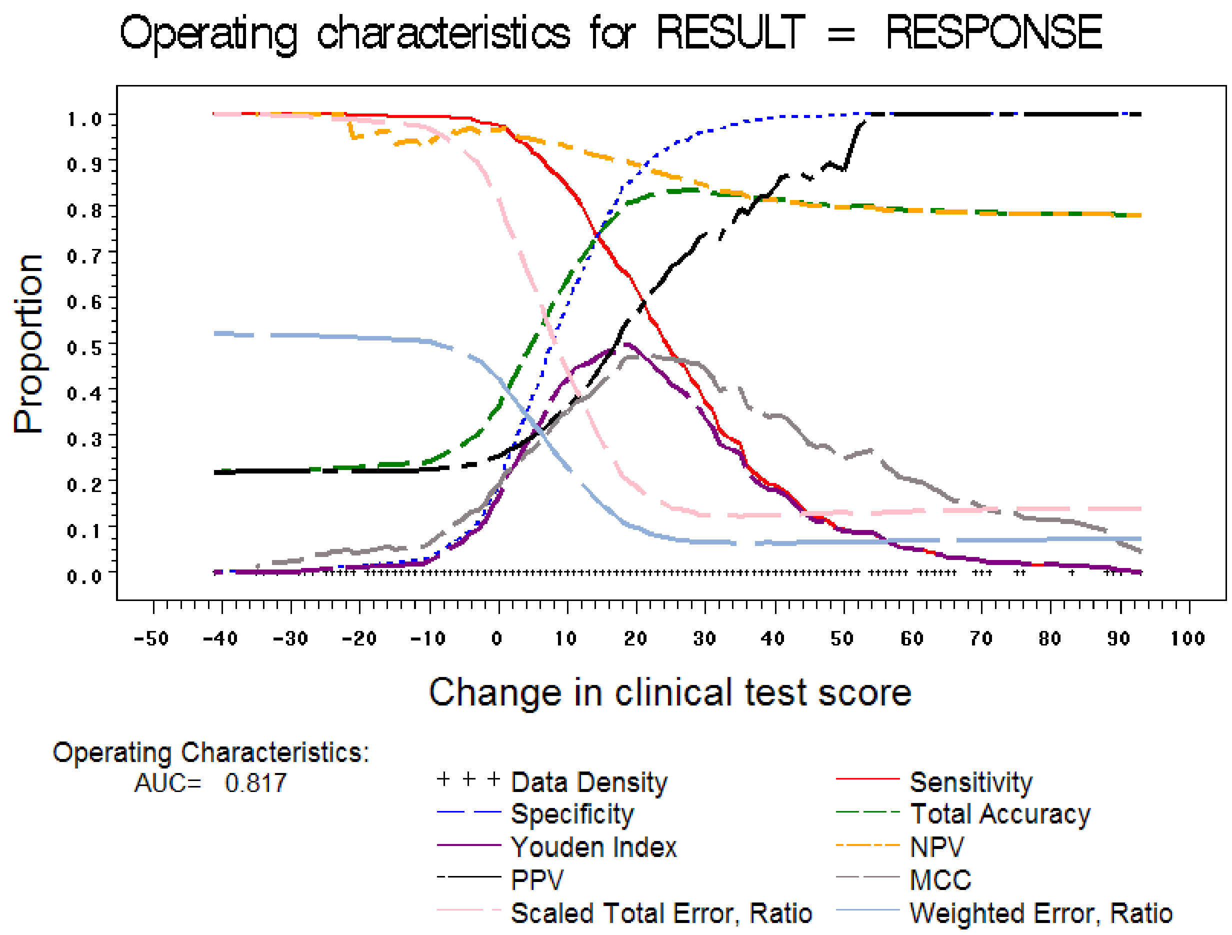
```
%SNSP_TRADEOFF(DATASET = cutoff_search, OUTCOME = RESULT, OUTCOME_LEV = RESPONSE,
XVAR = SCORE_CHANGE, FONT = ARIAL, XVAR_LABEL = Change in clinical test score, AUC =
Y, SN = Y, SP = Y, TA = Y, YI = Y, NPV = Y, PPV = Y, MCC = Y, TME = N, TME1 = N, TME2
= Y, WME = N, WME1 = N, WME2 = Y, GAMMA = .5);
```

The DATASET variable is simply the name of your SAS data set. The OUTCOME parameter refers to the name of your categorical outcome variable, and OUTCOME_LEV is the level of that outcome variable that defines your result. For this example, the RESULT variable is defined as either "Response" or "No Response", and since the interest here is response to treatment, the value of "Response" is passed to the macro. It should be noted that OUTCOME_LEV could also be a numerical value, e.g., 1 or 0. XVAR refers to the name of your continuous prediction variable, and XVAR_LABEL is the label that will be displayed along the x-axis of your graph. The remaining parameters represent the operating characteristics. All macro parameters with values of "Y" are displayed in your output.

MACRO OUTPUT

Each operating characteristic is displayed by a unique color and line style. The x-axis will be the range of values from your prediction variable. The y-axis is restricted to values from 0.0 to 1.0, i.e., all accuracy measures only assume values from 0.0 to 1.0, and the measure of total error when the ratio of losses is used is scaled to be in the same range. Additionally, to allow you to visualize the density of the data in different regions of your prediction variable, individual data points are displayed as black "+" symbols along the x-axis.

The intersection of sensitivity and specificity correspond to the point that is commonly used to define the "optimal" trade-off between these measures. However, as seen by this example, both accuracy and error measures can be used in conjunction with sensitivity and specificity in making the decision of an optimal cut-off value. Cut-off values could be chosen where accuracy measures (*J, MCC, TA*) are maximized and error measures (TME_0, TME_1, TME_2, WME_0, WME_1, WME_2) are minimized.



In this example, sensitivity (—) and specificity (- - -) curves intersect near a value of approximately 15. However, the accuracy measures of *J*, *MCC*, and *TA* all appear to be maximized at values closer to 20-24. Both the total and weighted error ratios are minimized at values also near 20-24. This suggests that a change in test score in the range of 20-24 points might be a desirable alternative choice of cut-off point, as there is agreement of all accuracy measures (maximized) and weighted error measures (minimized). Based on the above results, the researchers would want to consider weighing the cost of a loss of sensitivity against the gain of an increase in specificity. In the end, it is you, the user, who must weigh the data driven results with the specific objective(s) of the test and potential other factors such as cost, ethical considerations, and prevalence of disease in selecting the best cut-off point.

CONCLUSION

In this paper, we presented a SAS macro that enhances the functionality of a standard ROC curve when the objective for test screening or prediction is to obtain a categorical result based on a single quantitative variable. This macro provides additional information that can be used during the decision making process of dichotomizing a continuous variable at an optimal cut-off point. We presented the methodology and the usability of the output generated by the macro through the use of an example.

REFERENCES

Bewick Viv, Liz Cheek, and Jonathan Ball. 2004. “Statistics review 13: Receiver operating characteristic curves”. *Critical Care* 8:508-512.

Matthews, B.W. 1975. “Comparison of the predicted and observed secondary structure of T4 phage lysozyme”. *Biochim. Biophys. Acta.*, 405:442-451.

SAS Institute Inc. 1999. SAS OnlineDoc®, Version 8. http://v8doc.sas.com/sashtml/stat/chap39/sect49.htm. Accessed 2008 Jan 02. Copyright 1999, SAS Institute Inc., Cary, NC, USA. All Rights Reserved. Reproduced with permission of SAS Institute Inc., Cary, NC.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Jennifer Sniadecki
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285
Work Phone: 317.277.7285
E-mail: sniadecki_jennifer@lilly.com

Ilya Lipkovich
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285
Work Phone: 317.651.6095
E-mail: lipkovichia@lilly.com

APPENDIX

```
%MACRO SNSP_TRADEOFF(
/*********************************************************************
MACRO NAME         : SNSP_TRADEOFF
PURPOSE            : CREATES A GRAPH WITH MULTIPLE ROC MEASURES
SAS VERSION        : VERSION 8.2, 9.1
PARAMETERS         :
---------------------------------------------------------------------
NAME         TYPE    DEFAULT     DESCRIPTION AND VALID VALUES
---------   -------- --------- -----------------------------------------
DATASET     REQUIRED          : SOURCE DATASET
OUTCOME     REQUIRED          : RESPONSE VARIABLE (CATEGORICAL)
OUTCOME_LEV REQUIRED          : EVENT OF INTEREST. A LEVEL OF THE OUTCOME
                                PARAMETER THAT DEFINES THE EVENT
                               (E.G., OUTCOME_LEV=1 OR OUTCOME_LEV=RELAPSE)
XVAR        REQUIRED          : CONTINUOUS PREDICTOR
FONT        OPTIONAL  ARIAL   : FONT TYPE OF GRAPHICAL OUTPUT
XVAR_LABEL  OPTIONAL          : LABEL OF PREDICTOR IN GRAPHICAL OUTPUT
AUC         OPTIONAL    Y     : Y/N. DISPLAY AREA UNDER CURVE
SN          OPTIONAL    Y     : Y/N. DISPLAY SENSITIVITY
SP          OPTIONAL    Y     : Y/N. DISPLAY SPECIFICITY
TA          OPTIONAL          : Y/N. DISPLAY TOTAL ACCURACY
YI          OPTIONAL          : Y/N. DISPLAY YOUDEN INDEX
NPV         OPTIONAL          : Y/N. DISPLAY NEGATIVE PREDICTIVE VALUE
PPV         OPTIONAL          : Y/N. DISPLAY POSITIVE PREDICTIVE VALUE
MCC         OPTIONAL          : Y/N. DISPLAY MCC
TME         OPTIONAL          : Y/N. DISPLAY TOTAL MISCLASSIFICATION
                                     ERROR USING INVERSE OF TARGET GROUP
                                     AS LOSS
TME1        OPTIONAL          : Y/N. DISPLAY TOTAL MISCLASSIFICATION
                                     ERROR USING INVERSE OF OVERALL
                                     SAMPLE SIZE AS LOSS
TME2        OPTIONAL          : Y/N. DISPLAY TOTAL MISCLASSIFICATION
                                     ERROR USING RATIO OF LOSSES
WME         OPTIONAL          : Y/N. DISPLAY WEIGHTED MISCLASSIFICATION
                                     ERROR USING INVERSE OF TARGET GROUP
                                     AS LOSS
WME1        OPTIONAL          : Y/N. DISPLAY WEIGHTED MISCLASSIFICATION
                                     ERROR USING INVERSE OF OVERALL
                                     SAMPLE SIZE AS LOSS
WME2        OPTIONAL          : Y/N. DISPLAY WEIGHTED MISCLASSIFICATION
                                     ERROR USING RATIO OF LOSSES
```

```
GAMMA       REQUIRED IF      : Y/N. POSITIVE NUMERIC VALUE OF RATIO OF
LFN/LFP
            TME2 OR WME2=Y          (I.E., LOSS DUE TO
                                    MISCLASSIFICATION OF A TRUE
                                    POSITIVE EVENT IS "GAMMA" TIMES
                                    LARGER THAN MISCLASSIFICATION
                                    OF A TRUE NEGATIVE EVENT)
-------------------------------------------------------------------------
NOTES :THE FOLLOWING DATASETS ARE CREATED IN THE WORK LIBRARY AND DELETED
         BY THE MACRO PRIOR TO ALL CALCULATIONS.
        _SNSP_EST, _SNSP_ROC1, _SNSP_AUC, _SNSP_OUTPUT.
***************************************************************************/
DATASET = , OUTCOME = , OUTCOME_LEV = , XVAR = , FONT = ARIAL,
XVAR_LABEL = , AUC = Y, SN = Y, SP = Y, TA = , YI = , NPV = ,
PPV = , MCC = , TME = , TME1 = , TME2 = , WME = , WME1 = ,
WME2 = , GAMMA = );

%PUT MACRO SNSP_TRADEOFF IS NOW EXECUTING...;

%*****************************************************************;
%*   CHECKING THAT REQUIRED PARAMETERS ARE SPECIFIED             *;
%*****************************************************************;

%IF %SYSFUNC(EXIST(&DATASET)) = 0 %THEN %DO;
      %PUT ERROR: DATA SET &DATASET IS MISSING;
      %GOTO EXIT;
%END;
%IF %LENGTH(&DATASET) = 0 %THEN %DO;
      %PUT ERROR: VALUE FOR PARAMETER DATASET IS MISSING;
      %GOTO EXIT;
%END;
%IF %LENGTH(&OUTCOME) = 0 %THEN %DO;
      %PUT ERROR: VALUE FOR PARAMETER OUTCOME IS MISSING;
      %GOTO EXIT;
%END;
%IF %LENGTH(&OUTCOME_LEV) = 0 %THEN %DO;
      %PUT ERROR: VALUE FOR PARAMETER OUTCOME_LEV IS MISSING;
      %GOTO EXIT;
%END;
%IF %LENGTH(&XVAR) = 0 %THEN %DO;
      %PUT ERROR: VALUE FOR PARAMETER XVAR IS MISSING;
      %GOTO EXIT;
%END;
%IF &TME2 = Y OR &WME2 = Y %THEN %DO;
      %IF %SYSEVALF(&GAMMA <= 0) %THEN %DO;
            %PUT ERROR: VALUE FOR GAMMA IS MISSING OR INCORRECT;
            %GOTO EXIT;
      %END;
%END;

  %LOCAL COUNT;
  %LOCAL CURDATA;
  %LET DATASETS = _SNSP_EST _SNSP_ROC1 _SNSP_AUC _SNSP_OUTPUT;
  %IF %LENGTH(&DATASETS) > 0 %THEN %DO;
        %LET COUNT=1;
        %LET CURDATA  =%SCAN(&DATASETS,&COUNT,' ');
        %DO %WHILE(&CURDATA NE);
            %IF %SYSFUNC(EXIST(&CURDATA)) %THEN %DO;
                      PROC DATASETS NOLIST;
                          DELETE &CURDATA;
```

```
                        RUN; QUIT;
              %END;
            %LET COUNT=%EVAL(&COUNT+1);
            %LET CURDATA  =%SCAN(&DATASETS,&COUNT,' ');
         %END;
   %END;

%LET AUC=%UPCASE(&AUC);
%LET SN=%UPCASE(&SN);
%LET SP=%UPCASE(&SP);
%LET TA=%UPCASE(&TA);
%LET YI=%UPCASE(&YI);
%LET NPV=%UPCASE(&NPV);
%LET PPV=%UPCASE(&PPV);
%LET MCC=%UPCASE(&MCC);
%LET TME=%UPCASE(&TME);
%LET TME1=%UPCASE(&TME1);
%LET TME2=%UPCASE(&TME2);
%LET WME=%UPCASE(&WME);
%LET WME1=%UPCASE(&WME1);
%LET WME2=%UPCASE(&WME2);
%LET OUTCOME_LEV = %SYSFUNC(TRANWRD(&OUTCOME_LEV,%STR(%"),));
%LET OUTCOME_LEV = %SYSFUNC(TRANWRD(&OUTCOME_LEV,%STR(%'),));

%*****************************************************************;
%*  GETTING ROC OUTPUT FROM MODEL                                *;
%*****************************************************************;

ODS LISTING CLOSE;
PROC LOGISTIC DATA=&DATASET OUTEST=_SNSP_EST;
    MODEL &OUTCOME (EVENT = "&OUTCOME_LEV") = &XVAR/EXPB OUTROC=_SNSP_ROC1;
    ODS OUTPUT ASSOCIATION=_SNSP_AUC (WHERE = (UPCASE(LABEL2) = "C"));
RUN;
ODS LISTING;
DATA _SNSP_ROC1;
  SET _SNSP_ROC1; INDEX =1;
RUN;

DATA _SNSP_EST (KEEP=INTERCEPT &XVAR INDEX);
   SET _SNSP_EST; INDEX =1;
RUN;

DATA _SNSP_OUTPUT;
      MERGE _SNSP_ROC1 _SNSP_EST;
      BY INDEX;
      SPEC = 1-_1MSPEC_;
      YI = (_SENSIT_ + SPEC) - 1;
      N = _POS_+ _NEG_+ _FALPOS_+ _FALNEG_;
      TA = (_POS_ + _NEG_)/N;
      X_VALUE = (LOG(_PROB_/(1-_PROB_))-INTERCEPT)/&XVAR;
      IF (_NEG_+_FALNEG_) NE 0 THEN DO;
                  NPV = _NEG_/(_NEG_ +_FALNEG_);
      END;
      IF (_POS_+_FALPOS_) NE 0 THEN DO;
            PPV = _POS_/(_POS_ +_FALPOS_);
      END;
      IF
((_POS_+_FALPOS_)*(_POS_+_FALNEG_)*(_NEG_+_FALPOS_)*(_NEG_+_FALNEG_)) NE 0
THEN DO;
            MCC = ((_POS_*_NEG_)-(_FALPOS_*_FALNEG_))/
```

```
      SQRT(((_POS_+_FALPOS_)*(_POS_+_FALNEG_)*(_NEG_+_FALPOS_)*(_NEG_+_FALNEG
_)));
      END;
%******************************************************************;
%* FOR TOTAL ERRORS                                                *;
%******************************************************************;
   %***TARGET GROUP;
      LFN = 1/(_POS_ + _FALNEG_);
      LFP = 1/(_NEG_ + _FALPOS_);
      TLN = _FALNEG_*LFN;
      TLP = _FALPOS_*LFP;
      TME = TLN+TLP;
      %***EQUAL FPR/FNR;
      LFN1 = 1/N;
      LFP1 = 1/N;
      TLN1 = _FALNEG_*LFN1;
      TLP1 = _FALPOS_*LFP1;
      TME1 = TLN1+TLP1;

%******************************************************************;
%* FOR WEIGHTED ERRORS                                             *;
%******************************************************************;
      %***TARGET GROUP;
      WFN = LFN/(LFN+LFP);
      WFP = LFP/(LFN+LFP);
      WLN = (_FALNEG_*WFN)/N;
      WPN = (_FALPOS_*WFP)/N;
      WME = WLN+WPN;
      %***EQUAL FPR/FNR;
      WFN1 = LFN1/(LFN1+LFP1);
      WFP1 = LFP1/(LFN1+LFP1);
      WLN1 = (_FALNEG_*WFN1)/N;
      WPN1 = (_FALPOS_*WFP1)/N;
      WME1 = WLN1+WPN1;
    DENSITY = 0;
RUN;

%******************************************************************;
%* RATIO FOR TOTAL-AND-WEIGHTED ERRORS                             *;
%******************************************************************;
DATA _SNSP_OUTPUT;
      SET _SNSP_OUTPUT;
      %IF %SYSEVALF(&GAMMA > 0) %THEN %DO;
                  G=&GAMMA;
                  TME2_NSCALE = TLN1+TLP1*(1/G);
                  WME2 = TLN1*(G/(1+G))+TLP1*(1/(1+G));
            RUN;
            PROC SQL NOPRINT;
                  SELECT MAX(TME2_NSCALE) INTO: MAX_TME2
            FROM _SNSP_OUTPUT;
            QUIT;
            DATA _SNSP_OUTPUT;
                  SET _SNSP_OUTPUT;
            TME2 = TME2_NSCALE/&MAX_TME2;
            RUN;
      %END;
      %ELSE %DO;
                  TME2 = .;
            WME2 = .;
```

```
                        %LET WME2 = N;
                        %LET TME2 = N;
            %END;
      RUN;

      %IF &AUC = Y %THEN %DO;
            DATA _NULL_;
            SET _SNSP_AUC;
            CALL SYMPUT("AUCVALUE", PUT((NVALUE2),8.3));
            RUN;
      %END;

      %LET FIRST = / &SN/ &SP/ &TA/ &YI/ &NPV/ &PPV/ &MCC/ &TME/ &TME1/ &TME2/
      &WME/ &WME1/ &WME2;
      %LET LIST =  _SENSIT_ SPEC TA YI NPV PPV MCC TME TME1 TME2 WME WME1 WME2;
                 %DO I = 1 %TO 13;
                        %IF %SCAN(&FIRST, &I, "/") = Y %THEN %DO;
                              %LET TT&I = %SCAN(&LIST, &I, "
      ")*X_VALUE=%EVAL(&I+1);
                        %END;
                        %ELSE %LET TT&I = ;
                        %END;
      TITLE "Operating characteristics for &OUTCOME = &OUTCOME_LEV";
      AXIS1 W=1 OFFSET=(3 PCT) LABEL=(F=&FONT H=2 "&XVAR_LABEL");
      AXIS2 W=1 OFFSET=(3 PCT) LABEL=(F=&FONT H=2 A=90 R=0 "Proportion") ORDER = (0
      TO 1 BY 0.1);
      LEGEND1 VALUE=(F=&FONT H=1.5) LABEL=(F=&FONT H=1.5 "Operating
      Characteristics:");
            %IF &AUC = Y %THEN %DO;

                  LEGEND2 VALUE=(F=&FONT H=1.5)
                  LABEL=(F=&FONT  H=1.5 JUSTIFY=C "Operating Characteristics:"
                                          H=1.3 JUSTIFY=C "AUC=&AUCVALUE");
            %END;
      SYMBOL1   H=0.3 I=NONE    W=1 C=BLACK V="|";
      SYMBOL2   H=1 I=JOIN L=1  W=2 C=RED;
      SYMBOL3   H=1 I=JOIN L=2  W=2 C=BLUE;
      SYMBOL4   H=1 I=JOIN L=4  W=2 C=GREEN;
      SYMBOL5   H=1 I=JOIN L=6  W=2 C=PURPLE;
      SYMBOL6   H=1 I=JOIN L=8  W=2 C=ORANGE;
      SYMBOL7   H=1 I=JOIN L=10 W=2 C=BLACK;
      SYMBOL8   H=1 I=JOIN L=12 W=2 C=RGR;
      SYMBOL9   H=1 I=JOIN L=14 W=2 C=LIME;
      SYMBOL10  H=1 I=JOIN L=16 W=2 C=BROWN;
      SYMBOL11  H=1 I=JOIN L=22 W=2 C=PINK;
      SYMBOL12  H=1 I=JOIN L=32 W=2 C=MAGENTA;
      SYMBOL13  H=1 I=JOIN L=33 W=2 C=YELLOW;
      SYMBOL14  H=1 I=JOIN L=24 W=2 C=VLIGB;
      PROC GPLOT DATA = _SNSP_OUTPUT;
        PLOT DENSITY*X_VALUE=1
             &TT1 &TT2 &TT3 &TT4 &TT5 &TT6 &TT7 &TT8 &TT9 &TT10 &TT11 &TT12 &TT13
             /OVERLAY VAXIS=AXIS2 HAXIS=AXIS1
              %IF &AUC = Y %THEN %DO;
                LEGEND=LEGEND2;
              %END;
               %ELSE %DO;
                LEGEND=LEGEND1;
              %END;
          LABEL DENSITY = "Data Density" TA = "Total Accuracy" YI= "Youden Index"
```

```
          SPEC = "Specificity" TME = "Total Error, Group" TME1 = "Total
Error, Overall"
          TME2 = "Scaled Total Error, Ratio"  WME = "Weighted Error, Group"
          WME1 = "Weighted Error, Overall" WME2 = "Weighted Error, Ratio";
RUN;
QUIT;

%EXIT:
TITLE;

%MEND SNSP_TRADEOFF;
```