



## SAS® Tech Report

July 8, 2008

**Dear Readers,**

I'm really excited about this month's *SAS Tech Report*. Instead of a summer lull, we've really turned up the heat. Your SAS NEWS section is full of information to help make your job easier. So don't miss these hot topics:

Tips for getting [your analytics project funded](#)

A sample about [creating a risk map](#)

A great installation [FAQ for SAS 9.1.3 SP4](#)

And finally, it's time to start planning which of the fall [users group conferences](#) you are going to attend. There are six great locations from which to choose. I hope to see you all there!

**Waynette Tubbs**

[techeditor@sas.com](mailto:techeditor@sas.com)

*Editor, SAS Tech Report*

## SAS News

### **Become a Wikipedian**

It's likely that you've joined the 684 million who've researched something on Wikipedia. But have you considered participating in how terms such as [business intelligence](#) , [predictive analytics](#) and [operations research](#) are defined? At SAS, we're [finding our voice](#) , and with a bit of [research](#) , you could too!

### **SAS® Abbreviations are Your Friends**

Is there a proc, function or block of code that you often use? If you answered yes to these questions and you use the Enhanced Editor on the Windows platform, learn about easy-to-use SAS abbreviations and save valuable time.

[http://analytics.ncsu.edu/sesug/2006/CC09\\_06.PDF](http://analytics.ncsu.edu/sesug/2006/CC09_06.PDF)

## **How to Sell Your Boss on the Power of Analytics**

### **Buy-in and payback and ROI – oh, my!**

*By Thornton May*

One would be hard-pressed to recall any major business intelligence and analytics conferences where keynote speakers and breakout presenters were not asked – frequently, repeatedly and quasi-desperately:

How do I sell BI/analytics projects to my senior managers?

How do I get executive buy-in?

How do I calculate the ROI on the projects I am proposing?

Researchers at the IT Leadership Academy were fascinated by the paradox of high-payback, reasonable-risk BI/analytics projects that were put on the back burner in many otherwise thought-to-be-clever enterprises. We undertook a quick analysis in an attempt to understand what is going on here.

#### **Understanding how money happens**

Some readers may recall the famous interchange between bank robber Willy Sutton and the policeman who captured him. “Why do you rob banks, Willy?” asked the policeman. “Cause that’s where the money is.” In many cases, the people asking for BI funding do not really know where the money is or how projects get funded. All organizations have a BUYING PROCESS – sometimes very formalized, sometimes more mood-based – detailing how money happens and how projects get funded. Very few BI/analytical professionals undertake more than a perfunctory analysis of these processes, deeming, I suppose, such bureaucratic procedures unworthy of their time or attention. This is a BIG mistake. A sidebar finding from the research was that most buying processes in most enterprises would benefit significantly if subjected to rigorous BI scrutiny. In the future, we predict BI applications in IT project governance will generate very high payback. Those wishing to “sell” BI projects to senior managers would be well-advised to familiarize themselves with the organizational norms of funding. At GE Capital Solutions, for instance, there has been a preference to fund small “popcorn stands,” or little businesses that will help the organization learn about bigger opportunities. At a major film studio, the BI project sponsor adopted the same language that show-biz people use to greenlight creative endeavors.

#### **Understanding mindsets and mental models**

In 2000, the CIO Posse – a group of transitioning and just-retired CIOs – was formed to offer the services of CIO tribal elders to vendor sales organizations seeking to improve their sales at the top of the house. The goal was to get vendor sales personnel inside the heads of potential buyers. To sell something, you must really understand how the person who is buying that thing thinks.<sup>1</sup> Step one toward project success is the recognition that on every issue, at any given point in time, multiple mental models may be at work. The analytical community can no longer merely observe, passively consume or merely relate to the mental models swirling around enterprises about what BI and high-end analytical resources can do. You will need to inventory the multiple mental models at work in your areas of engagement. You will need to intervene and remediate mental models toxic to BI’s ability to add value. Mental models are important. They can be dysfunctional. They should be monitored and managed – and in many cases upgraded.

#### **An exercise to try at work**

A fascinating exercise is to assemble your BI and analytics teams and have them draw a one-page picture of the universe that faces the executives who have been asked to fund the initiative. Next, arrange a lunch with those executives and ask them to draw their own one-page picture of the universe facing them. The differences in those “maps” will be important. The conversation about those differences will be helpful in achieving the shared mind that leads to project approval. As you are inventorying mental models, it is a good idea to get a feel for the general mental

model at work in the environment. Historians tell us that science proceeds by alternating between periods of normal science, during which investigators do research within a commonly accepted paradigm, and crises, during which investigators seek a new paradigm because of problems with the old one. The tone of your pitch should reflect the prevailing gestalt. You should know whether the executives you seek to work with are in an order-shattering mindset, an order-affirming mindset or an order-creating mode.<sup>2</sup> As a futurist, I believe most switched-on organizations have to understand that the world has changed, that existing modes of information management are not good enough and that new practices need to be deployed. As an on-the-ground worker in a living, breathing and frequently reluctant-to-change organism, you will need to fit your message into the mental model at work in your enterprise. Good luck!

<sup>1</sup> Gerald Zaltman, *How Customers Think: Essential Insights into the Mind of the Market* [Boston: Harvard Business School Press, 2003] and Paco Underhill, *Why We Buy: The Science of Shopping* [NY: Simon & Schuster, 2000]. <sup>2</sup> T.S. Kuhn, *The Structure of Scientific Revolutions* [Chicago: University of Chicago Press, 1962]. Stephen Skowronek, *The Politics Presidents Make: Leadership from John Adams to George Bush* [Cambridge, MA: The Belknap Press of Harvard University Press, 1993], 20.

**Bio:** Thornton May, Executive Director and Dean at the IT Leadership Academy, is one of the premier visionaries in the IT industry. [www.itleadershipacademy.com](http://www.itleadershipacademy.com)

## It's time for SAS® users to connect!

This fall, you won't need to travel far to give your SAS skills a boost. With six regional SAS users group conferences spread throughout the US, one is sure to be close to you. The conference season opens September 14 with the [NorthEast SAS Users Group \(NESUG\)](#), and the season ends November 5 with the [Western Users of SAS Software \(WUSS\)](#).

Each conference is unique, but there are some common benefits, including:

- Optional pre- and post-conference training courses
- Hands-on workshops where you can learn and network
- Presentations by expert SAS staff and industry peers
- Exhibits and demonstration rooms
- Network opportunities

There are conferences in locations as inviting as Seattle and St. Pete Beach, Florida, so you may want to add some personal time at the end to soak up the scenery. The complete regional SAS users group conference schedule is at

<http://support.sas.com/usergroups/namerica/allregconf.html>.

## **SAS® Global Forum Papers: The Best of the Best**

We are continuing our highlight of the top papers presented at SAS® Global Forum. You can read the [award-winning papers](#) or check out the [complete list](#) of papers that were presented at the forum.

## Great Forum Paper Series: Revisiting Dynamic Data Exchange

Nathaniel Derby argues that Dynamic Data Exchange outperforms the newer methods for exporting data into custom-formatted spreadsheets in some cases.

<http://www2.sas.com/proceedings/forum2008/259-2008.pdf>

## **Great Forum Paper Series: Using nmon to Monitor SAS® Applications on AIX Servers**

Learn to use nmon, a free downloadable tool, to monitor and determine the causes of performance problems that occur as a result of dynamic changes within the datacenter or the enterprise.

<http://www2.sas.com/proceedings/forum2008/386-2008.pdf>

## Creating a Risk Map

This sample uses the BAND statement in PROC SGPLOT to create a risk map.

```
/* Create sample data */
data work.bp;
  infile datalines;
  input dias losys hisys grp $ 13-33 refx refy;
datalines;
75 110 190 Severe Hypertension 75 140
115 110 190 Severe Hypertension 90 140
75 110 180 Moderate Hypertension 90 190
110 110 180 Moderate Hypertension . .
75 110 160 Mild Hypertension . .
100 110 160 Mild Hypertension . .
75 110 140 High/Normal . .
90 110 140 High/Normal . .
75 110 130 Normal . .
85 110 130 Normal . .
75 110 120 Optimal . .
80 110 120 Optimal . .
;
run;

/* Use PROC TEMPLATE to define the gradient colors to be
   used for the bands in the graph */
proc template;
  define style styles.bp;
  parent=styles.listing;
  style graphcolors from graphcolors /
    'gdata6'=CXFFFFFFF
    'gdata5'=CXF5D9D8
    'gdata4'=CXEBB3B1
    'gdata3'=CXE18D8A
    'gdata2'=CXD76763
    'gdata1'=CXBF1810;
  end;
run;

/* Use the STYLE created in PROC TEMPLATE */
ods listing style=bp;
title 'Stages of Hypertension';

/* Use the BAND statement to create the bands
   in the plot with an overlaid line from the
   SERIES statement */
proc sgplot data=bp;
  band x=dias lower=losys upper=hisys /
    group=grp name="bp";
  series x=refx y=refy / lineattrs=(thickness=2);
  xaxis values=(75 to 115 by 5) label='Diastolic Pressure (mmHg)';
  yaxis values=(110 to 190 by 10) label='Systolic Pressure (mmHg)';
  keylegend "bp" / position=right across=1 title='Stage';
  inset "Isolated" "Systolic" "Hypertension" / position=left
```

```
textattrs=graphlabeltext;  
run;  
run;
```

## Nonparametric Comparison of Areas Under Correlated ROC Curves

**NOTE: Beginning in SAS 9.2, comparison of areas under correlated ROC curves can be done using the ROC and ROCCONTRAST statements in PROC LOGISTIC. See the PROC LOGISTIC documentation for details and examples. NOTE: Indicated portions of the SAS/IML code in this sample program were provided, and are supported, by the author D.M. DeLong (see [REFERENCES](#) below).**

### PURPOSE:

Nonparametric comparison of areas under correlated ROC curves. Provides point and confidence interval estimates of each curve's area and of the pairwise differences among the areas. Tests of the pairwise differences are also given. Any contrast among the areas may be estimated and tested.

### HISTORY:

<i>Version</i>	<i>Update Notes</i>
1.7	Fixed row labeling when CONTRAST= option is specified.
1.6	Removed computation of estimated correlations to avoid problems with use of intercept-only model. Fix to automatic check for new version.
1.5	Fixed subscript error in RNAME with 15 or more data sets and XBETA variables. Added automatic check for new version.
1.4	Corrected errors with default contrast and only one variable specified in VAR=. The single area is now estimated and tested.
1.3	Added confidence intervals for each area. Added ALPHA= parameter to control confidence level. Default contrasts are now all pairwise differences. Added table of confidence intervals and tests of each contrast row.
1.2	Put original author's code in macro form. Allow input of several data sets to compare several models which may each have multiple predictors. Default contrast set generated. Added DETAILS= parameter to minimize default output. Creates macro variable with overall p-value.

### REQUIREMENTS:

Version 7 or later of Base SAS and SAS/IML.

### USAGE:

Follow the instructions in the Downloads tab of this sample to save the %ROC macro definition. Replace the text within quotes in the following statement with the location of the %ROC macro definition file on your system. In your SAS program or in the SAS editor window, specify this statement to define the %ROC macro and make it available for use:

```
%inc "<location of your file containing the ROC macro>";
```

Following this statement, you may call the %ROC macro. See the Results tab for examples.

You must first run the LOGISTIC procedure to fit each of the the models whose ROC curves are to be compared. Each model must be fit to exactly the same data set (with observations in the same order) and they must all be fit to the same response variable. The ROC macro requires that the response variable have only values of 0 and 1 even though this is not a restriction necessary for PROC LOGISTIC. Use the XBETA= or PRED= option in the OUTPUT statement in each run of LOGISTIC. The output data set names must be unique. Since the XBETA= or PRED= output variables are used to

identify and compare the models' ROC curves, the names must be unique and it is recommended that you select a name for each variable that adequately identifies the fitted model.

The following macro parameters can be specified. The DATA, VAR=, and RESPONSE= parameters are required.

**DATA=**

Specify the names, separated by spaces, of the OUT= data sets from PROC LOGISTIC to be analyzed. Each data set should contain a variable created by the XBETA= or PRED= option. This parameter is required.

**VAR=**

The names of the XBETA= or PRED= variables, one from each data set. No name may occur more than once. Separate variable names in the list with spaces. This parameter is required.

**RESPONSE=**

The name of the response variable which has values of 0 or 1 only. This parameter is required.

**CONTRAST=%str(hypothesis matrix)**

Specifies the desired hypothesis matrix, L, for testing the null hypothesis  $L \cdot R = 0$ , where R is the vector of ROC curve areas and the ROC curve areas are assumed to be in the order given in the VAR= variable. The hypothesis matrix, L, must be surrounded by %str(). An L matrix with more than one row may be specified by using commas to separate the rows of coefficients. If omitted, a default contrast testing the equality of all ROC curve areas jointly and pairwise is used.

**ALPHA= $\alpha$**

Specifies the level of significance  $\alpha$  for  $100(1-\alpha)\%$  confidence intervals. The value  $\alpha$  must be between 0 and 1. The default value is 0.05 which results in 95% intervals.

**DETAILS=YES | NO**

Requests additional output. Default is no. Any other value will produce the additional output.

The version of the %ROC macro that you are using is displayed when you specify **version** (or any string) as the first argument. For example:

```
%roc(version, data=a b, var=xba xbb, response=y)
```

**DETAILS**

Details of the methodology are given in DeLong, et. al. (1988). The method is closely related to the jackknife. The method of components is described in chapter 3 of Puri and Sen (1971).

The ROC macro assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select or exclude tables and to create output data sets. The table names are listed in the following table along with any macro parameters required to make the table available. When outputting a table to a data set, you can determine the names of variables in the data set by using PROC CONTENTS.

ODS table name	Description	%ROC macro option
T	Pairwise deletion Mann-Whitney Statistics	DETAILS=yes
V	Estimated Variance Matrix	DETAILS=yes
NX	X populations sample sizes	DETAILS=yes
NY	Y populations sample sizes	DETAILS=yes

L	Coefficients of Contrast	default
LT	Estimates of Contrast	default
LV	Variance Estimates of Contrast	DETAILS=yes
AREASTAB	ROC Area Std Error Confidence Limits	default
DIFFSTAB	Tests and 100(1- $\alpha$ )% Confidence Intervals for Contrast Rows	default
CTEST	Contrast Test Results	default

In addition to displayable output, the ROC macro creates a macro variable, **pvalue**, containing the p-value from the test of joint equality of ROC curve areas.

The %ROC macro attempts to check for a later version of itself. If it is unable to do this (such as if there is no active internet connection available), the macro will issue the following message:

ROC: Unable to check for newer version

The computations performed by the macro are not affected by the appearance of this message.

#### LIMITATIONS

The ROC macro cannot do BY-group processing and cannot handle output from PROC LOGISTIC which used BY-group processing. Limited error checking is done. Be sure that the data sets and variables exist and are correctly spelled. Only response values of 0 and 1 are allowed.

#### MISSING VALUES

Pairwise deletion of missing values is used prior to computing Mann-Whitney components.

#### SEE ALSO

The [%ROC PLOT macro](#) can plot an ROC curve with labelled points.

#### REFERENCES

E.R. DeLong, D.M. DeLong, and D.L. Clarke-Pearson (1988), "Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, 44, 837-845.

Puri M.L and Sen P.K. (1971), *Nonparametric Methods in Multivariate Analysis*, Wiley.

#### EXAMPLE 1: Comparing ROC curves of three diagnostics

From DeLong et. al. (1988). Three prognostic indices of post-operative success are compared.

Each ROC curve is plotted separately with labels and a final joint plot of all curves is produced. Statistical tests of the equality of the areas under all three curves and pairwise comparisons among the curves are given along with point and confidence interval estimates.

Before running this example, be sure to define both the %ROC and [%ROC PLOT](#) macros.

```
/* Define the ROC macro */
%include "<path to your copy of the ROC macro>";

/* Define the ROC PLOT macro */
%include "<path to your copy of the ROC PLOT macro>";

data roc;
input alb tp totscore popind;
```

```

totscore = 10 - totscore;
datalines;
3.05.8 10 0
3.26.3 5 1
3.96.8 3 1
2.84.8 6 0
3.25.8 3 1
0.94.0 5 0
2.55.7 8 0
1.65.6 5 1
3.85.7 5 1
3.76.7 6 1
. . 6 1
3.25.4 4 1
3.86.6 6 1
4.16.6 5 1
3.65.7 5 1
4.37.0 4 1
3.66.7 4 0
2.34.4 6 1
4.27.6 4 0
4.06.6 6 0
3.55.8 6 1
3.86.8 7 1
3.04.7 8 0
4.57.4 5 1
3.77.4 5 1
3.16.6 6 1
4.18.2 6 1
4.37.0 5 1
4.36.5 4 1
3.25.1 5 1
2.64.7 6 1
3.36.8 6 0
1.74.0 7 0
. . 6 1
3.76.1 5 1
3.36.3 7 1
4.27.7 6 1
3.56.2 5 1
2.95.7 9 0
2.14.8 7 1
. . 8 1
2.86.2 8 0
. . 7 1
. . 7 1
4.07.0 7 1
3.35.7 6 1
3.76.9 5 1
2.0 . 7 1
3.66.6 5 1
;

/* Albumin */
title "ROC plot for Albumin";
proc logistic data=roc;
    model popind(event='1') = alb / outroc=or roceps=0;

```

```

output out=albpred p=palb;
ods output association=assoc;
run;

data _null_;
  set assoc;
  if label2='c' then call symput("area",cvalue2);
run;

title2 "Approximate area under curve = &area";
%rocplot(out=albpred, outroc=or, p=palb, id=alb)
title2;

data joint;
  set _rocplot;
  length index $ 13;
  Index='Albumin';
run;

/* Total Protein */
title "ROC plot for Total Protein";
proc logistic data=roc;
  model popind(event='1') = tp / outroc=or roceps=0;
  output out=tppred p=ptp;
  ods output association=assoc;
run;

data _null_;
  set assoc;
  if label2='c' then call symput("area",cvalue2);
run;

title2 "Approximate area under curve = &area";
%rocplot(out=tppred, outroc=or, p=ptp, id=tp)
title2;

data tp;
  set _rocplot;
  length index $ 13;
  Index='Total Protein';
run;

data joint;
  set joint tp;
run;

/* K-G score */
title "ROC plot for K-G score";
proc logistic data=roc;
  model popind(event='1') = totscore / outroc=or roceps=0;
  output out=totspred p=ptots;
  ods output association=assoc;
run;

data _null_;
  set assoc;
  if label2='c' then call symput("area",cvalue2);
run;

title2 "Approximate area under curve = &area";
%rocplot(out=totspred, outroc=or, p=ptots, id=totscore)
title2;

data kg;

```

```

        set _rocplot;
        length index $ 13;
        Index='K-G Score';
        run;

data joint;
    set joint kg;
    run;

/* Compare areas under the ROC curves of the indices using the
method presented
in DeLong, et. al. (1988).
*/
%roc(data=albpred tppred totspred,
var=palb ptp ptots,
response=popind)

/* Plot all indices and p-value of overall test comparing areas */
symbol1 i=join v=circle c=blue line=33;
symbol2 i=join v=dot c=green line=1;
symbol3 i=join v=triangle c=red line=4;

proc gplot data=joint;
    title "Post-Op Indices of surgical success";
    title2 "Test of H0: equal areas under curves -- p=&pvalue";
    footnote "From DeLong, et. al. (1988, Biometrics 44)";
    label index="Index";
    plot _sensit_ * _1mspec_ = Index /
        vaxis=0 to 1 by .1 haxis=0 to 1 by .1;
    run;
    quit;
title;
title2;
footnote;

```

## RESULTS

The default set of contrasts compares the three indices. The large p-value ( $p=.4423$ ) indicates no significant advantage of the scoring system. The results vary somewhat from the DeLong et.al. (1988) paper because this program computes components based on pairwise deletion of missing observations, whereas the original analysis used pairwise deletion after the components were calculated. The current strategy is felt to be more conservative with respect to possible association of response with the probability of the response not being observed.

For brevity, the PROC LOGISTIC results are not shown. The following plots are generated by the %ROC PLOT macro:

[ROC plot for Albumin \(alb\)](#)  
[ROC plot for Total Protein \(tp\)](#)  
[ROC plot for K-G score \(totscore\)](#)

### Comparative ROC plot

Following are the results of the %ROC macro.

The ROC Macro

**ROC Curve Areas and 95% Confidence Intervals**

	<b>ROC Area</b>	<b>Std Error</b>	<b>Confidence</b>	<b>Limits</b>
				<b>palb</b>
	0.7188			
	0.0928			
	0.5368			
	0.9007			
				<b>ptp</b>
	0.6478			
	0.1000			
	0.4518			
	0.8439			
				<b>ptots</b>
	0.6869			
	0.1051			
	0.4809			
	0.8929			

**Contrast Coefficients**

	<b>palb</b>	<b>ptp</b>	<b>ptots</b>	<b>Row1</b>
1				
-1				
0				
				<b>Row2</b>
1				
0				
-1				
				<b>Row3</b>
0				
1				
-1				

**Tests and 95% Confidence Intervals for Contrast Rows**

	<b>Estimate</b>	<b>Std Error</b>	<b>Confidence</b>	<b>Limits</b>	<b>Chi-square</b>	<b>Pr &gt; ChiSq</b>
						<b>Row1</b>
	0.0709					
	0.0560					
	-0.0389					
	0.1807					

1.6017	
0.2057	
	<b>Row2</b>
0.0318	
0.0980	
-0.1603	
0.2240	
0.1053	
0.7456	
	<b>Row3</b>
-0.0391	
0.1069	
-0.2485	
0.1703	
0.1338	
0.7145	
	<b>Contrast Test Results</b>
	<b>Chi-Square</b>
	<b>DF</b>
	<b>Pr &gt; ChiSq</b>
1.6314	
2	
0.4423	

### EXAMPLE 2: Comparing three competing models

The data are from an example in *Logistic Regression Using the SAS System: Theory and Application* by Paul Allison (book number 55770) in which students were asked if they would return a lost wallet and contents if found. The response, WALLETS, indicates if the wallet and contents would be returned (WALLETS=1) or kept (WALLETS=0). Predictors are gender, whether in business school, parental punishment at various ages, and parental explanation for punishment.

Three models are fit and their ROC curve areas tested for equality. A significant finding is followed by pairwise comparisons.

```

data wallet;
  input wallet male business punish explain @@;
  if wallet in (1,2) then wallet=0;
  if wallet=3 then wallet=1;
  datalines;
2 0 0 2 0   3 0 0 1 1   3 1 0 1 1   2 0 1 1 1   3 1 0 1 1
2 0 0 2 1   3 0 0 1 1   3 0 0 1 1   1 0 0 3 0   2 1 0 1 0
3 0 0 1 1   3 0 0 1 1   3 0 1 1 1   1 0 1 2 0   3 0 1 2 1
3 0 0 2 0   3 0 0 2 0   3 0 0 1 1   3 0 0 1 1   2 1 1 2 1
1 1 0 1 1   3 1 1 1 1   3 0 0 1 1   3 0 0 1 1   3 0 0 1 1

```

```

3 0 0 1 1    3 1 0 1 1    3 0 0 2 1    3 0 0 1 1    1 1 0 1 1
3 0 0 1 1    3 0 0 1 1    2 1 1 1 0    3 1 1 1 1    3 1 1 3 1
3 1 0 1 1    3 0 0 1 1    3 1 0 1 1    3 0 0 1 0    2 1 0 1 1
3 1 0 1 1    3 0 0 1 0    3 0 0 1 1    3 0 0 1 1    1 0 0 1 0
3 0 0 2 1    3 0 1 3 0    2 1 1 1 1    3 1 0 1 1    2 1 0 3 0
2 0 1 1 1    3 1 0 2 0    3 0 0 1 1    3 1 1 1 1    3 0 0 1 1
3 1 1 1 1    2 0 0 2 1    3 1 1 1 1    3 1 0 1 1    2 1 1 1 1
3 1 0 1 1    3 0 0 1 1    3 0 0 1 1    2 1 1 1 1    3 0 0 1 1
3 1 0 1 1    3 1 0 1 1    3 0 0 1 1    2 0 0 1 0    2 0 0 2 0
3 0 0 1 1    3 1 0 1 1    3 0 0 1 1    3 0 0 1 1    2 1 0 1 1
3 0 0 1 0    3 1 0 1 0    3 0 1 1 1    2 1 1 2 0    1 1 1 2 0
3 0 0 2 1    3 1 1 1 0    3 0 0 2 1    3 1 0 1 0    2 1 0 2 0
2 0 0 3 0    3 1 0 1 0    1 1 1 1 1    2 1 0 1 0    2 0 0 1 1
1 1 1 3 0    2 1 1 3 1    1 1 0 2 0    3 0 0 2 1    1 0 1 3 0
2 0 0 1 1    3 0 0 1 1    2 1 0 1 1    3 1 1 1 1    2 1 0 1 1
2 1 0 2 0    3 0 0 3 1    1 1 1 1 1    1 0 0 3 0    2 1 0 1 1
3 0 0 1 1    3 0 0 1 1    2 1 0 1 1    3 0 0 1 1    3 1 0 1 1
3 1 0 1 1    3 0 0 1 0    1 1 1 3 0    3 1 1 1 1    3 0 0 1 1
3 1 1 1 0    3 0 0 1 1    2 1 1 1 1    3 0 0 1 0    2 1 0 1 1
3 1 0 1 1    1 0 1 1 1    1 1 1 3 1    3 0 1 1 0    1 1 0 2 0
2 0 0 1 0    3 0 0 1 0    3 0 0 3 1    2 0 0 1 1    2 0 1 2 1
2 1 0 3 0    3 1 0 1 1    3 0 0 1 1    3 1 0 1 1    3 0 0 2 0
1 1 0 2 0    1 1 0 1 0    2 0 0 1 1    2 1 0 1 0    2 1 1 1 1
3 1 0 2 0    3 1 0 3 1    3 0 0 1 1    3 0 0 1 0    3 0 1 2 1
3 1 0 1 1    2 1 0 3 1    1 0 1 3 0    2 1 0 1 1    1 0 0 3 0
3 0 0 1 1    1 1 1 2 1    3 0 0 1 1    3 1 0 1 1    2 1 1 1 0
3 1 0 2 1    1 1 0 2 1    2 0 0 1 0    3 1 0 2 0    3 0 0 3 1
1 1 0 1 1    3 0 0 1 1    3 0 0 1 1    3 1 1 2 1    2 1 0 2 1
3 0 0 2 0    3 0 0 3 0    3 0 0 1 1    3 1 0 1 1    3 0 0 1 1
3 1 0 1 0    2 1 0 1 1    2 1 0 1 1    2 1 0 1 0    3 1 0 3 1
3 1 1 2 1    2 0 0 1 1    3 1 0 1 1    3 0 0 1 1    3 0 0 1 1
3 0 0 2 0    2 1 0 1 1    3 0 0 1 1    3 1 0 1 1    3 1 1 1 1
3 0 0 1 1    3 0 0 1 0    1 1 0 3 0    2 1 1 1 0    3 1 0 1 1
3 1 0 1 1    3 1 1 1 1    2 1 0 1 1    3 0 0 1 1    2 1 0 1 1
;
/* Fit three models to the Pr(return wallet) */
proc logistic data=wallet;
  model wallet(event="1") = male business punish explain;
  output out=w1 xbeta=mbpe;
run;

proc logistic data=wallet;
  model wallet(event="1") = business punish explain;
  output out=w2 xbeta=bpe;
run;

proc logistic data=wallet;
  model wallet(event="1") = business punish ;
  output out=w3 xbeta=bp;
run;

/* Test equality of the three ROC curve areas. */
%roc(data=w1 w2 w3,
      response=wallet,
      var=mbpe bpe bp)

```

## RESULTS

Following are the results from the %ROC macro. The overall p-value ( $p=.0281$ ) indicates

that the three areas differ. The default pairwise comparisons show that this difference is due to a significant difference between the first and third models.

The ROC Macro	
ROC Curve Areas and 95% Confidence Intervals	
	ROC Area Std Error Confidence Limits mbpe
0.7412	
0.0355	
0.6717	
0.8107	
	<b>bpe</b>
0.6873	
0.0383	
0.6122	
0.7624	
	<b>bp</b>
0.6508	
0.0379	
0.5766	
0.7250	
Contrast Coefficients	
	<b>mbpe</b>
	<b>bpe</b>
	<b>bp</b>
	<b>Row1</b>
1	
-1	
0	
	<b>Row2</b>
1	
0	
-1	
	<b>Row3</b>
0	
1	
-1	
Tests and 95% Confidence Intervals for Contrast Rows	
	<b>Estimate Std Error Confidence Limits Chi-square Pr &gt; ChiSq Row1</b>

0.0539	
0.0291	
-0.0032	
0.1110	
3.4215	
0.0644	
	<b>Row2</b>
0.0904	
0.0340	
0.0239	
0.1570	
7.0884	
0.0078	
	<b>Row3</b>
0.0365	
0.0276	
-0.0175	
0.0906	
1.7530	
0.1855	
	<b>Contrast Test Results</b>
	<b>Chi-Square</b>
	<b>DF</b>
	<b>Pr &gt; ChiSq</b>
7.1417	
2	
0.0281	

### EXAMPLE 3: Testing the area under the ROC curve

You can test the null hypothesis that the area under an ROC curve is 0.5 by comparing the model of interest to an intercept-only model. The intercept-only model has area equal to 0.5, so the default contrast (1 -1) performs this test. These additional statements test the area of the last model in the previous example. The PROC LOGISTIC step fits the intercept-only model.

```
proc logistic data=wallet;
  model wallet(event="1") = ;
  output out=w4 xbeta=int;
run;
```

```
%roc(data=w3 w4,
      response=wallet,
      var=bp int)
```

### RESULTS:

The test shows that the area under the ROC curve for this model is significantly different from 0.5 ( $p < 0.0001$ ).

The ROC Macro	
ROC Curve Areas and 95% Confidence Intervals	
	ROC Area Std Error Confidence Limits bp
0.6508	
0.0379	
0.5766	
0.7250	
	int
0.5000	
0.0000	
0.5000	
0.5000	
Contrast Coefficients	
	bp int Row1
1	
-1	
Tests and 95% Confidence Intervals for Contrast Rows	
	Estimate Std Error Confidence Limits Chi-square Pr > ChiSq Row1
0.1508	
0.0379	
0.0766	
0.2250	
15.8648	
<.0001	
Contrast Test Results	
15.8648	
1	
<.0001	

**Type:** Sample

**Topic:** SAS Reference ==> Procedures ==> IML Analytics ==> Regression Analytics ==> Matrix Programming SAS Reference ==> Procedures ==> LOGISTICAnalytics ==> Descriptive Statistics

**Date Modified:** 2007-08-14 03:03:08

**Date Created:** 2005-01-13 15:03:44

<b>Product Family</b>	<b>Product</b>	<b>Host</b>	<b>Starting Release</b>	<b>Ending Release</b>
SAS System	SAS/STAT	All	8 TS M0	n/a
SAS System	SAS/IML	All	8 TS M0	n/a

## Create Variable Labels from Data Set Values

This sample uses the variable values from the first observation in a data set to create variable labels.

This technique is often helpful when you are reading external files whose first record is meant to be variable names. If the values are not valid SAS variable names, you can use them as variable labels instead. See Example 2 for more details.

```
/* Example 1: Use the values from one data set to create labels for
another */
/*          data set
*/

/* Create a macro to generate labels from values in one data set and
apply */
/* the new labels to variables in another data set.
*/

%macro label(datain= );
  %local dsid getvalue getvarname close i ;

  /* Open dataset whose values in the first observation will become new
labels */
  %let dsid=%sysfunc(open(&datain));

  /* ATTRN and NVAR will return the number of variables in &datain */
  %do i=1 %to %sysfunc(attrn(&dsid,nvars));

    /* Retrieve each variable name in &datain */
    %let getvarname=%sysfunc(varname(&dsid,&i));

    /* FETCHOBS reads the specified observation from &datain */
    %let rc=%sysfunc(fetchobs(&dsid,1));

    /* Retrieve the value of each variable */
    %let getvalue=%qsysfunc(getvarc(&dsid,&i));

    /* Build the syntax for the LABEL statement that will be generated
*/
    &getvarname = "&getvalue"
  %end;

  /* Close the dataset */
  %let close=%sysfunc(close(&dsid));

%mend label;

/* Create a sample data set whose values will later be used as labels
for */
/* a second data set.
*/

data one;
  infile datalines dsd ;
  input (var1-var3) (: $13.);
datalines;
label's one,label's two,label's three
```

```

;

/* Call the macro %LABEL while creating WORK.TWO. %LABEL will generate
a */
/* LABEL statement based upon the variables and values from the
specified */
/* data set, in this case, WORK.ONE. Note the variable names must be
the */
/* same in both data sets.
*/

data two;
  input (var1-var3) ($1.,+1);
  label %label(datain = one);
datalines;
a b c
d e f
;

proc print data=two label;
run;

/* Example 2: Applying the macro %LABEL when reading from one file
*/
/*
*/
/*
/*          Read a flat file whose first record is meant to be
variable    */
/*          names but are invalid SAS variable names. Create labels
from    */
/*          the first record instead.
*/

/* Create sample test file to be read. Modify your FILE statement as
needed. */

data _null_;
  file "c:\temp\sample1727.txt";
  put "2005,2006,2007,Total Revenue";
  put "1,2,3,25000";
  put "4,5,6,50000";
run;

/* Read only the first record from SAMPLE1727.TXT using OBS=1 on the
INFILE */
/* statement.
*/

data one_2;
  infile "c:\temp\sample1727.txt" dsd obs=1;
  input (var1-var3) (:$4.) var4 :$13.;
run;

/* Read the rest of SAMPLE1727.TXT starting from the second

```

```

observation. Use */
/* the macro %LABEL created above to use the values from WORK.ONE_2 as
the new */
/* variable labels.
*/

data two_2;
  infile "c:\temp\sample1727.txt" dsd firstobs=2;
  input (var1-var4) (:8.);
  label %label(datain = one_2);
run;

proc print data=two_2 label;
run;

```

**Example 1 Results**

Obs	label's one	label's two	label's three
1	a	b	c
2	d	e	f

**Example 2 Results**

Obs	2005	2006	2007	Total Revenue
1	1	2	3	25000
2	4	5	6	50000

## Support and Installation of SAS® 9.1.3 SP4 on Microsoft Windows Vista

- I. Which Microsoft Windows Vista(TM) operating systems does SAS support?
- \* SAS supports the following Windows Vista(TM) 32-bit editions:
    - Enterprise
    - Business
    - Ultimate
  - \* SAS does NOT support Windows Vista(TM) 32-bit Home Editions:
    - Premium
    - Basic
  - \* SAS does NOT support Windows Vista(TM) 64-bit editions.
- II. How do you install SAS Foundation on the Microsoft Windows Vista(TM) operating system?
- \* [#020595](#)
- III. Which SAS products does SAS support on the Microsoft Windows Vista(TM) operating system?
- \* [#020402](#)
- IV. Where are the SAS support statements for Microsoft Windows Vista(TM)?
- \* <http://support.sas.com/techsup/pcn/vista.html>
  - \* <http://support.sas.com/documentation/hosts/pc/>
- V. SAS supports SAS 9.1.3 SP4 with Internet Explorer 7 on the Windows Vista(TM) operating system.
- \* [#020480](#)
- VI. SAS 9.1.3 installs on Microsoft Windows Vista(TM) by starting with the SAS Setup Disk. If you try to install SAS using the SAS Software Navigator, the installation will not work:
- \* [#020380](#)
- VII. When installing a hotfix for SAS under the Microsoft Windows Vista(TM) operating system, you will be prompted three times to make sure you really want to install the hotfix.
- \* [#020443](#)
- VIII. Instructions for installing SAS/IntrNet CGI Tools for SAS 9.1.3 on the Microsoft Windows Vista(TM) operating system
- \* [#020469](#)

IX. Internet Explorer (IE7) crashes when running a "live" Java applet on the Microsoft Windows Vista(TM) operating system.

\* [#020465](#)

## Webcasts

### Cost and Profitability Optimization – On Demand

<http://www.bettermanagement.com/seminars/seminar.aspx?l=14782>

### Past, Present and Future of Forecasting

<http://www.sas.com/events/cm/413078/index.html>

Tuesday, July 15  
11:30 a.m. – 12:30 p.m. ET

### Sensing and Shaping Demand with Demand-Driven Forecasting

<http://www.bettermanagement.com/seminars/seminar.aspx?l=14734>

Tuesday, July 16  
11:30 a.m. – 12:30 p.m. ET

## Events

### TDWI World Conference

<http://www.tdwi.org/education/conferences/sandiego2008/index.aspx>

Business intelligence and data warehousing conference  
August 17 – 20  
San Diego

### KDD08

<http://www.sigkdd.org/kdd2008/>

International data mining and knowledge discovery conference  
August 24 – 27  
Las Vegas

### M2008

<http://www.sas.com/events/dmconf/index.html>

Data mining conference  
October 27 – 28  
Las Vegas

## Tech Resources

### FASTats: Frequently Asked-For Statistics

<http://support.sas.com/kb/30/333.html>

## **SAS® 9.2 Highlights and Resource Center**

<http://support.sas.com/software/>

Provides an overview of SAS 9.2 Phase 1 and links to other resources available on the site.

## **SAS® Maps Online**

<http://support.sas.com/rnd/datavisualization/maponline/>

This application is directed to a wide range of map data users: SAS/GRAPH® customers, educators, students and general map users.

## **Multimedia**

### **e-Learning & Back to School**

<http://www.support.sas.com/selfpaced>

With this special discount, you'll save 30% and SAS will donate 10% to the Communities in Schools organization.

### **Certification Exam Fair**

<http://support.sas.com/certify/promo.html>

Save 25% on certification fees when you take SAS programming and predictive modeling exams in July.

### **SAS® Training site redesigned!**

<http://support.sas.com/training>

Check out SAS Training's new video portal, improved search navigation, enhanced feedback tools and Starting Points pages.