

TECHNIQUES FOR MODEL SCORING ESUG



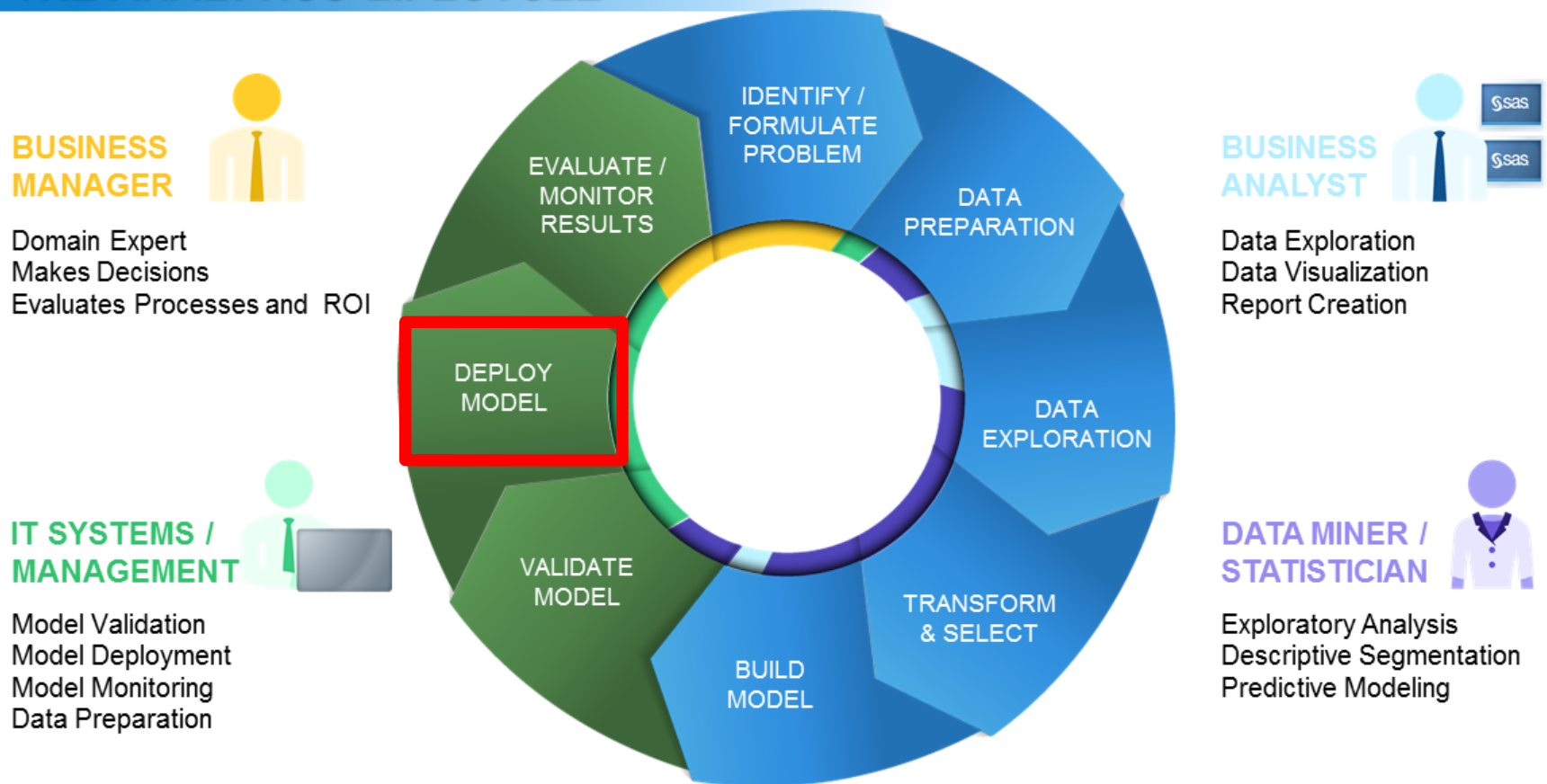
Sylvain Tremblay
SAS Canada

APRIL 15, 2015

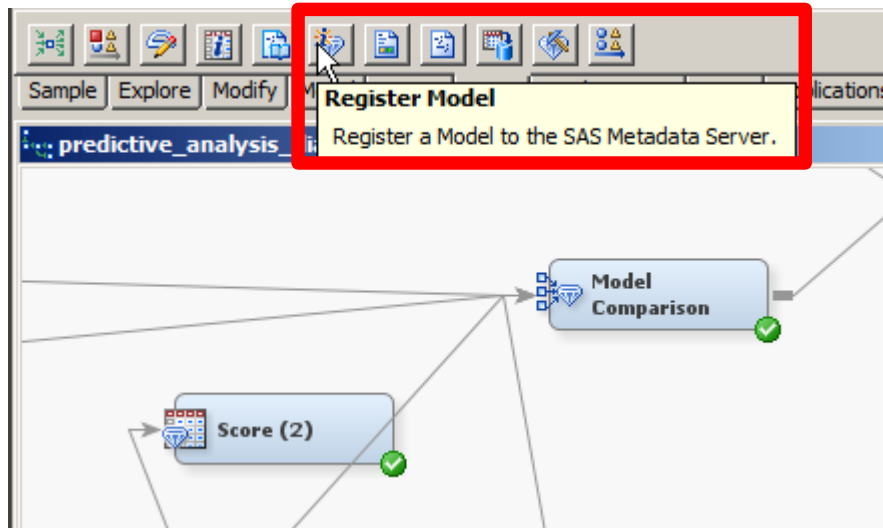
You are done and have a predictive model...

Now what? It's time to score

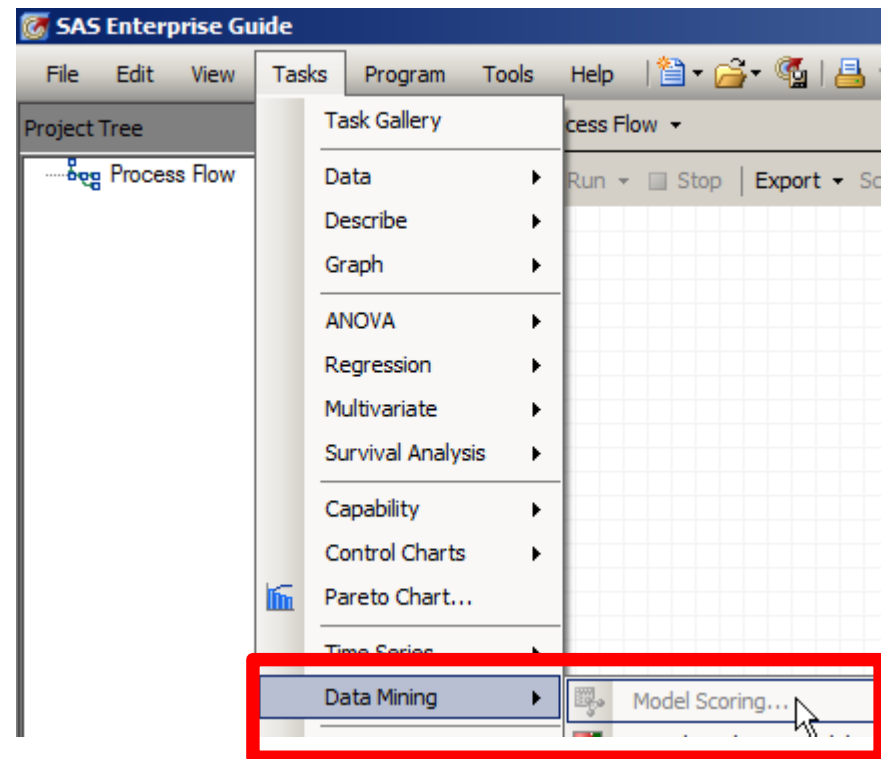
THE ANALYTICS LIFECYCLE



If you are using Enterprise Miner...



You can then do the scoring in EG



What if you are using SAS/Stat ?

Agenda

- Overview of model scoring
- Proc SCORE
- SCORE statement
- STORE statement and Proc PLM
- CODE STATEMENT
- References + Questions

Overview of Model Scoring in SAS/Stat

Many strategies are available:

- 1) Output parameter estimates for a model to a SAS dataset and use Proc SCORE
- 2) Use a SCORE statement in the Proc itself
- 3) Output your model information in a item store using the STORE statement and then use Proc PLM to score
- 4) Save score code using a CODE statement

Agenda

- Overview of model scoring
- Proc SCORE
- SCORE statement
- STORE statement and Proc PLM
- CODE STATEMENT
- References + Questions

Proc SCORE

Multiplies values from two SAS data sets:

- One containing the estimated coefficients
- the other containing raw data to be scored

$$\underline{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k$$

Proc SCORE

Syntax:

```
PROC SCORE DATA=SAS-data-set <options> ;  
  BY variables ;  
  ID variables ;  
  VAR variables ;
```

Statements:

- **ID** identifies which variables to be included in the output dataset
- **VAR** specifies the variables to be used in computing scores
- **BY** obtain separate analyses of observations in groups

Proc SCORE

Syntax:

```
PROC SCORE DATA=SAS-data-set <options> ;  
  BY variables ;  
  ID variables ;  
  VAR variables ;
```

Usefull options:

- SCORE= Names the data set containing the scoring coefficients
- TYPE= Specifies the obs. that contain scoring coefficients
- PREDICT To produce predicted values
- OUT= Specifies the name of the SAS data set created

Agenda

- Overview of model scoring
- Proc SCORE
- SCORE statement
- STORE statement and Proc PLM
- CODE STATEMENT
- References + Questions

SCORE Statement


Nonparametric regression procedures cannot output parameter estimates...

- You can use a SCORE Statement directly in the Modeling procedure itself to produce the scores
- Typically used in non parametric procedures

SCORE Statement

Syntax:

SCORE DATA=SAS-data-set OUT=SAS-data-set <keyword ... keyword> ;



Instructions:

- **DATA=** specifies the input SAS data set to be scored
- **OUT=** specifies the name of the SAS data set to contain the predictions

Agenda

- Overview of model scoring
- Proc SCORE
- SCORE statement
- STORE statement and Proc PLM
- CODE STATEMENT
- References + Questions

STORE Statement and Proc PLM

Efficient post-processing:

- The STORE statement in the modeling procedure saves all of the information needed to recreate and evaluate the model
- The saved information can be read by the PLM procedure, which includes a SCORE statement, as well as many other capabilities

STORE Statement and Proc PLM

Syntax

STORE <OUT=>item-store-name </ LABEL='label'> ;

```
PROC PLM RESTORE=item-store-specification <options> ;  
  CODE <options> ;  
  EFFECTPLOT <plot-type <(plot-definition-options)>> </ options> ;  
  ESTIMATE <'label'> estimate-specification <(divisor=n)><, ...<'label'> estimate-specification  
<(divisor=n)>> </ options> ;  
  FILTER expression ;  
  LSMEANS <model-effects> </ options> ;  
  LSMESTIMATE model-effect <'label'> values <divisor=n><, ...<'label'> values <divisor=n>>  
</ options> ;  
  SCORE DATA=SAS-data-set <OUT=SAS-data-set><keyword<=name>> ...  
<keyword<=name>> </ options> ;  
  SHOW options ;  
  SLICE model-effect </ options> ;  
  TEST <model-effects> </ options> ;  
  WHERE expression ;
```

STORE Statement and Proc PLM

Syntax

```
SCORE DATA=SAS-data-set <OUT=SAS-data-set>  
<keyword<=name>>  
<keyword<=name>> </ options>;
```

Usefull option of the SCORE statement:

ilink

Model $\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 \cdot x_1 + \hat{w}_2 \cdot x_2 = \text{logit}(\hat{p})$

ilink will create the score for $\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$

Agenda

- Overview of model scoring
- Proc SCORE
- SCORE statement
- STORE statement and Proc PLM
- **CODE STATEMENT**
- References + Questions

CODE statement

In SAS/STAT 12.1 (SAS 9.3) the CODE statement was added to several SAS/STAT regression procedures

- The CODE statement writes DATA step statements into a text file
- You can then use the %INCLUDE statement to insert those statements into a DATA step

Agenda

- Overview of model scoring
- Proc SCORE
- SCORE statement
- STORE statement and Proc PLM
- CODE STATEMENT
- References + Questions

References

Techniques for scoring a regression model in SAS

The DO Loop Blog, Rick Wicklin

<http://blogs.sas.com/content/iml/2014/02/19/scoring-a-regression-model-in-sas/>

SAS/STAT Documentation

SAS/STAT 13.2 User's Guide

<https://support.sas.com/documentation/onlinedoc/stat/>

Training

SAS Canada Training

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

SAS Global Certification program

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



SAS Advanced Analytics

SAS Certified Predictive Modeler Using SAS Enterprise Miner 7

Designed for SAS Enterprise Miner users who perform predictive analytics

SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling

Appropriate for professionals who solve business problems by performing statistical analyses and predictive modeling using SAS/STAT software

QUESTIONS?



**THE
POWER
TO KNOW.**

Thank you!