Incremental Response Modeling with SAS® EM (14.1)

Also called Net Lift Modeling

Bruce Lund
Magnify Analytic Solutions,
Division of Marketing Associates

MSUG Meeting Feb 11, 2016

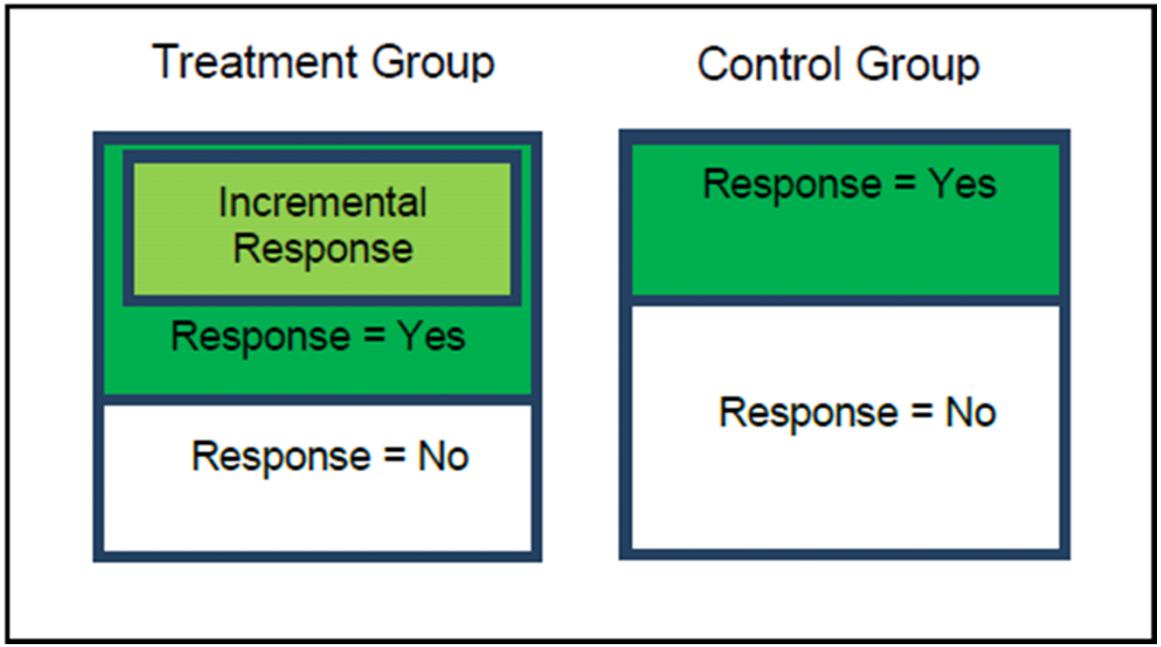
What is an Incremental Response Model? (aka Net Lift Model)

- An "Incremental Response" by a prospect is a response that would not have occurred without the "treatment" (i.e. the marketing campaign).
- Incremental Response Model (IRM) predicts the rate at which prospects will be "incremental" due to the treatment. For example, an incremental response model rate of 0.02 means that 2% of such prospects will make an incremental response due to the treatment. Note: An IRM rate can be negative.
- … Different from a Propensity Model (PM). PM assigns a probability to a prospect giving the likelihood that the prospect will respond (during a fixed future time period). Until recent years, PM's were used in lieu of IRM's.
- … A IRM is fit to a campaign with randomly selected treated and control groups. … With expectation that future campaign will have same treatment and similar prospects. … Must be able to track responses by control group.
- … IRM is relatively recent. An early paper on this topic dates to 2002.

A schematic

Some responders within the Treatment Group are “incremental” while other responders would have responded without the offer.

IRM tries to predict “which are which”.



From Lee, et al. “Incremental Response Modeling using SAS(R) Enterprise Miner (2013)”, SGF Paper 096-2013

Outcome: Incremental Revenue or Profit

After considering incremental Responses … the next step is to estimate incremental revenue or profit from a prospect. SAS uses Outcome to refer to the revenue or profit.

- This extra measurement is needed when the Outcome from a Response will vary according to the response.

  Generally the Outcome will vary

  … buy a car (profit varies greatly),

  … open a bank account (initial deposit will vary),

  … make a donation (amount varies).

Input Data for Incremental Response Modeling

For each Prospect:

Treatment: 1 = treated or 0 = control

Response: 1 = responded or 0 = not responded

Predictors for Response: Xi

Optional: Outcome (Revenue or Profit): Usually dollar amount

Outcome equals "missing" for non-responders

Optional: Predictors for Outcome: Zi

Xi and Zi may overlap.

History and References

- Lo, V. (2002), "The True Lift Model: A Novel Data Mining Approach to Response Modeling in Database Marketing." *ACM SIGKDD Explorations Newsletter* 4:78–86.
- Zhong, Jun (2009), VP Targeting and Analytics, Card Services Customer Marketing, Wells Fargo in the presentation: "Predictive Modeling & Today's Growing Data Challenges" at Predictive Analytics World in San Francisco, CA 2009.
- Kim Larsen (2009) introduces Net Lift Modeling to SAS at M2009, 12th Annual Data Mining Conference, Las Vegas.
- Larsen, Kim. (2010), "Net Lift Models: Optimizing the Impact of Your Marketing Efforts." *SAS Course Notes. Cary, NC: SAS Institute Inc.* … SAS Training Class on Net Lift Model
- Lund, Bruce (2012) Direct Marketing Profit Model, *MWSUG 2012*, paper CI-04
- Lee T., Zhang R., Meng X., and Ryan L. (2013), "Incremental Response Modeling Using SAS® Enterprise Miner", *Proceedings of the SAS Global Forum 2013 Conference*, Paper 096-2013. (See other references given in this paper.)

Background: Link Functions

A **link function** is a function that transforms the expected response μ to a linear combination of predictors $\sum \beta_j * X_j$. That is, $\text{LINK}(\mu) = \sum \beta_j * X_j = \beta * X$

Two important link function in the case of a binary response are the logit link function, and the probit link function. Let μ be the expected response.

Logit Link: $\log(\mu / (1 - \mu)) = \beta * X$

Equivalently: $\mu = P(Y=1) = e^{\beta * X} / (1 + e^{\beta * X})$

Probit Link: $\Phi^{-1}(\mu) = \beta * X$

where Φ^{-1} is the inverse of the standard normal cumulative distribution.

Equivalently: $\mu = P(Y=1) = \frac{1}{\sqrt{2*\pi}} \int_{-\infty}^{\beta * X} \exp(-0.5 * z^2)\, dz$

Both Logit and Probit links have a role in the SAS EM Incremental Response Node … as will be discussed.

The Incremental Response Node in SAS EM

Data Node: Your data (here called A_SIM).

Data Partition: Divides Your data into Training and Validation. Stratify by "Treatment" and "Target" to avoid the possibility of a biased sample. This Node can be omitted.

Incremental Response Node: This Node offers the choice of two models:

Combined Model
Difference Model

Both models include:

(1) **Response** Model for binary target
(2) **Outcome** Model for interval target (optional) … e.g. **Revenue, Profit**

If both (1) and (2), then there is a final computation to give expected incremental Outcome (revenue, profit, etc.) per prospect.

Combined Model – Illustrated Step by Step

A. Treated and control groups are appended. A

B. Indicator variable T … shows if prospect is treated (T=1) or control (T=0). B

C. Suppose there is **one** predictor X. Then interaction with T is X*T C

D. Fit to Response Y. The Combined Model is logistic: D

$$\hat{Y}_{Resp} = P(Y=1) = \exp(\alpha + \beta*X + \gamma*T + \phi*(X*T)) / (1 + \exp(\alpha + \beta*X + \gamma*T + \phi*(X*T)))$$

E. After fitting, assume $\alpha=1$, $\beta=2$, $\gamma=3$, and $\phi=4$.

… Now for each prospect compute 2 probabilities: 1

1. $\hat{Y}_{Resp\_T} = P(Y=1 \mid \mathbf{T=1}) = \exp(1 + 2*X + 3 + 4*X) / (1 + \exp(1 + 2*X + 3 + 4*X))$

$= \exp(4 + 6*X) / (1 + \exp(4 + 6*X))$

2. $\hat{Y}_{Resp\_C} = P(Y=1 \mid \mathbf{T=0}) = \exp(1 + 2*X) / (1 + \exp(1 + 2*X))$ 2

DS

The difference score (incremental rate) is: $\widehat{\mathbf{DS}} = \hat{\mathbf{Y}}_{\mathbf{Resp\_T}} - \hat{\mathbf{Y}}_{\mathbf{Resp\_C}}$

Combined Model - Illustrated

If there is Outcome Target: Linear Regression Model predicts the Outcome:

Suppose there is a single predictor Z. Then Z*T is the interaction with treatment

Model is: $\widehat{Y}_{Outcome} = \eta * Z + \delta * T + \psi * (Z * T)$ … Fit for **Outcome ≠ .** (Responders)

Set T = 1 to obtain $\widehat{Y}_{Outcome\_T}$ **T = 1** and T = 0 for $\widehat{Y}_{Outcome\_C}$ **T = 0**

Finally, there is **EXPECTED** incremental Outcome for a prospect:

$$\widehat{R}_{Incr\_T} = \widehat{Y}_{Resp\_T} * \widehat{Y}_{Outcome\_T} - \widehat{Y}_{Resp\_C} * \widehat{Y}_{Outcome\_C} - \text{Fixed Cost}$$

Incr.

(Estimated outcome for a treated responder)
X (Likelihood of Response if Treated)

(Estimated outcome for a control responder)
X (Likelihood of Response if Control)

Why not set Outcome to 0 for non-Responders?

$\widehat{Y}_{Outcome} = \eta*Z + \delta*T + \psi*(Z*T)$… Model is fit when Outcome ≠ . (Responders)

Can we set Outcome = 0 when Outcome = . and run regression on all prospects?

This would not accomplish the purpose of the Outcome Model which is:

To predict the **expected outcome** for Treated (Control) for a new Prospect List

Given that a prospect is a responder, we want the responder's Outcome to be estimated from real Outcomes. Then the **expected outcome** for a new prospect is the **Estimated Outcome X Probability of Response**.

(Using "zeros" in the Outcome Model would drastically reduce **expected outcome**.)

But what about "Selection Bias"? … Can a regression model that is fit only on Responders be used to estimate outcome for everyone in a new list? Maybe there is some "defect" among the non-responders that would affect their Outcome if in fact they responded?

There is no correction in Combined Model for selection bias. But see next slides …

Difference Model (= difference of two Models)

Suppose there is **one** predictor X.

For treated "T": A model is fit to response Y=1 vs Y=0. Let $\widehat{Y}_{Resp\_T}$ = P(Y=1 | T)

If no Outcome target, then logistic is used:

$$\widehat{Y}_{Resp\_T} = e^{\alpha_T + \beta_T * X} / (1 + e^{\alpha_T + \beta_T * X})$$

T

If there is Outcome target, then probit model is used (why? See next slides)

Likewise for the control group "C": $\widehat{Y}_{Resp\_C}$ = P(Y=1 | C)

C

The difference score (incremental rate) is: $\mathbf{\widehat{DS} = \widehat{Y}_{Resp\_T} - \widehat{Y}_{Resp\_C}}$

DS

Of course, the predictors for Treated Model and Control Model can be completely different.

Difference Model and Outcomes

If there is Outcome Target: Linear Regression predicts the Outcome.

Model is fit when Outcome ≠ . (Responders)

If **"T"**, then $\widehat{Y}_{Outcome\_T} = \eta_T * Z + \mathbf{b_T * M}$

This **added** term $\mathbf{b_T * M}$ **corrects** for selection bias.

The variable **M** is called the inverse Mills ratio and it arises from Probit Model

M = $\phi(p) / \Phi(p)$ where p is the response probability from Probit Model

ϕ is std normal density and Φ is std normal cum. distribution

The coefficient $\mathbf{b_T}$ is fit by the regression

Theory of using Inverse Mills ratio is discussed in Appendix A.

[➔ Using Logistic instead of Probit in **M** would give very similar results.]

Likewise for **"C"** to give $\widehat{Y}_{Outcome\_C} = \eta_C * Z + \mathbf{b_C * M}$

Finally, there is the **expected** incremental Outcome for a prospect:

$$\widehat{R}_{Incr\_T} = \widehat{Y}_{Resp\_T} * \widehat{Y}_{Outcome\_T} - \widehat{Y}_{Resp\_C} * \widehat{Y}_{Outcome\_C} - \text{Fixed Cost}$$

Combined Model and inverse Mills ratio ?

SAS EM documentation:

"The default outcome model (Difference Model) is a two-stage model that uses the inverse Mills ratio.

The Combined Model uses separate regressions for the binary and interval target."

This statement says (I think) that inverse Mills ratio is never used for the "Combined" Outcome Model.

My Conjecture: Although there may still be a selection bias, the Selection Bias correction using inverse Mills ratio does not apply to the Combined Output model since this approach does not account for a "treatment" variable.

The original reference is: Heckman, J. (1979). "Sample Selection Bias as a Specification Error." Econometrica 47:153-161.

Weight of Evidence (WOE) for X (for treated)

SAS EM (IRM Node) can "pre-screen" predictors using Net Information Value (NIV). The next slides explain NIV

	Y				
X	0	1	$P_T(X=i \mid Y=0)$	$P_T(X=i \mid Y=1)$	WOE
1	800	200	16.0%	23.5%	0.38566
2	1200	170	24.0%	20.0%	-0.18232
3	900	160	18.0%	18.8%	0.04474
4	1000	160	20.0%	18.8%	-0.06062
5	1100	160	22.0%	18.8%	-0.15593
SUM	5000	850			

800/5000 = 16% and 200/850 = 23.5%

WOE = LOG(16% / 23.5%) = 0.38566

First, here is the definition of **Weight of Evidence** of a predictor X

Net Weight of Evidence (NWOE) for X

Treated					
	Y				
X	0	1	$P_T(X=i \mid Y=0)$	$P_{T(}X=i \mid Y=1)$	WOE
1	800	200	16.0%	23.5%	0.38566
2	1200	170	24.0%	20.0%	-0.18232
3	900	160	18.0%	18.8%	0.04474
4	1000	160	20.0%	18.8%	-0.06062
5	1100	160	22.0%	18.8%	-0.15593

Control					
	Y				
X	0	1	$P_C(X=i \mid Y=0)$	$P_{C(}X=i \mid Y=1)$	WOE
1	1000	80	13.5%	24.2%	0.58441
2	1300	70	17.6%	21.2%	0.18852
3	1500	60	20.3%	18.2%	-0.10873
4	1700	60	23.0%	18.2%	-0.23390
5	1900	60	25.7%	18.2%	-0.34512

This concept was presented by K Larsen (2009)

T

T – C =

C

X	NWOE
1	-0.1988
2	-0.3708
3	0.1535
4	0.1733
5	0.1892

If Treated WOE is “big +” and Control WOE is “big -” … OR Conversely, then Abs(NWOE) is large

Weighting NWOE to compute Net Information Value (NIV)

$$P_T(X=i \mid Y=1)*P_C(X=i \mid Y=0) - P_T(X=i \mid Y=0)*P_C(X=i \mid Y=1)$$

	Treated		
X	$P_T(X=i \mid Y=0)$	$P_T(X=i \mid Y=1)$	WOE
1	16.0%	23.5%	0.38566
2	24.0%	20.0%	-0.18232
3	18.0%	18.8%	0.04474
4	20.0%	18.8%	-0.06062
5	22.0%	18.8%	-0.15593

	Control		
X	$P_C(X=i \mid Y=0)$	$P_C(X=i \mid Y=1)$	WOE
1	13.5%	24.2%	0.58441
2	17.6%	21.2%	0.18852
3	20.3%	18.2%	-0.10873
4	23.0%	18.2%	-0.23390
5	25.7%	18.2%	-0.34512

X	NWOE (a)	Weight (b)	= (a) * (b)
1	-0.1988	-0.0070	0.0014
2	-0.3708	-0.0158	0.0058
3	0.1535	0.0054	0.0008
4	0.1733	0.0069	0.0012
5	0.1892	0.0083	0.0016
	NIV =	Sum * 1000=	**108.4**

	Y=0	Y=1
T	A	C
C	B	D
= B*C - A*D		

If Treatment is strong for X_i, then C >> A and weight B*C - A*D >> 0

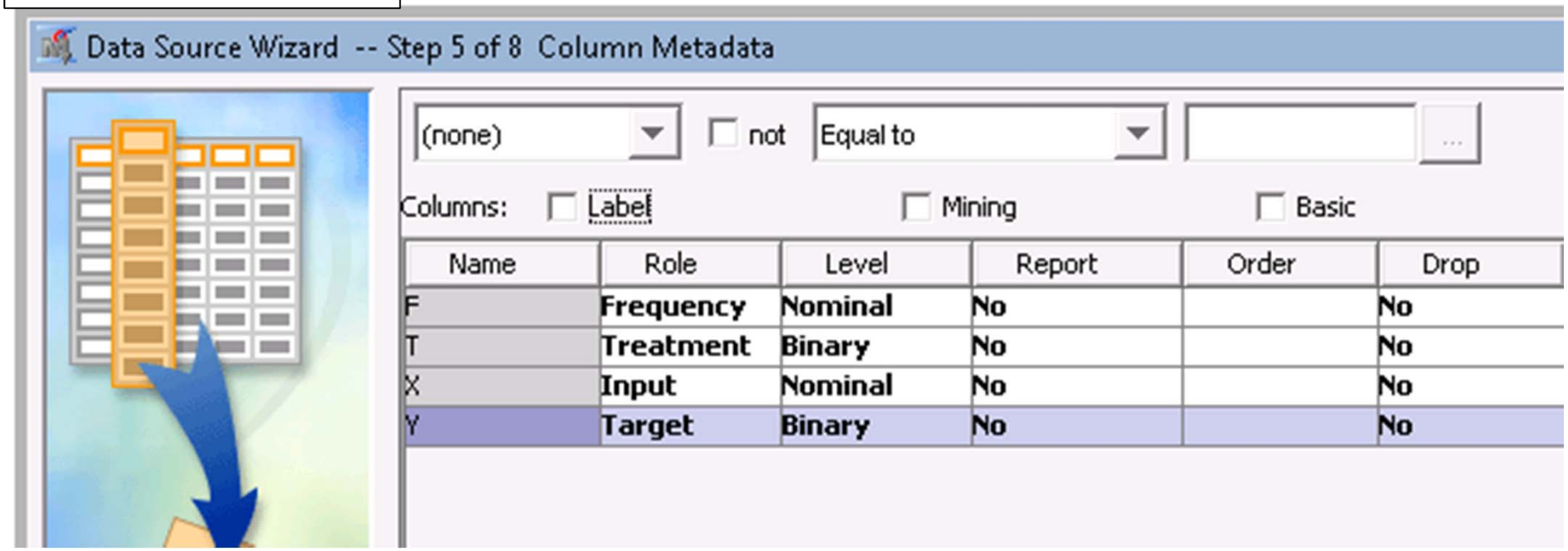
Example #1: Data Source & Select Roles

```
Data NetL.Test_NIV;
input X Y T F;
datalines;
1 0 0 1000
1 1 0 80
1 0 1 800
1 1 1 200
<16 lines omitted>
```

- X has 5 levels, Y is response, T is treatment
- F is a freq variable

… warning SAS EM IRM does not recognize a freq variable!

- There is no Outcome variable

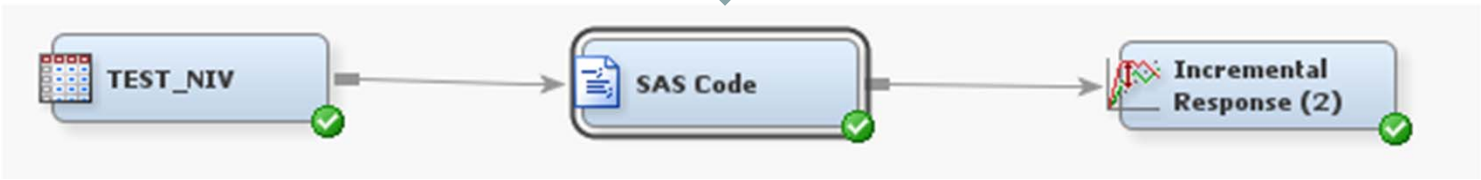


Name	Role	Level	Report	Order	Drop
F	Frequency	Nominal	No		No
T	Treatment	Binary	No		No
X	Input	Nominal	No		No
Y	Target	Binary	No		No

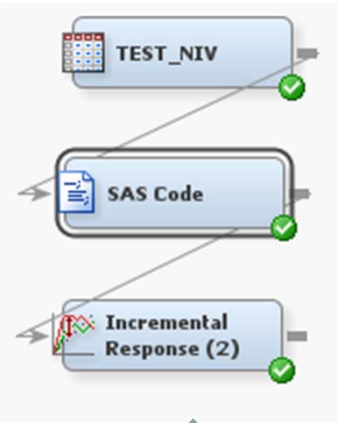
Example #1: Diagram Including a SAS Code Node

Create “R” as another INPUT and unwind the Freq Variable F

```
data &EM_EXPORT_TRAIN;
set &EM_IMPORT_DATA;
drop i;
do i = 1 to F;
   R = ranuni(12345);
   output;
   end;
run;
```



Example #1: Set-Up for “Combined” Response Model



General	
Node ID	IRM2
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Treatment Level Selection	
Prescreen Variables	Yes
Prescreening Criterion	Default
Rank Percentage Cutoff	90
Response Model Selection	
Combined Model	Yes
Selection Method	Forward Selection
Selection Criterion	None
Significance Level for Entry	0.95
Stay Significance Level	0.05
Suppress Intercept	No
Two-Way Interaction	No
Outcome Model Selection	

= NIV

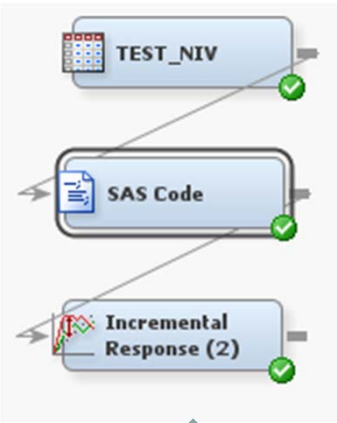
In this example **DEFAULT is NIV**

If Validation Data, then “penalized NIV” becomes the DEFAULT.

See Documentation

SLE = 0.95 to force in variables

Example #1: Prescreen Inputs using Net Information Value (NIV)



General	
Node ID	IRM2
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Treatment Level Selection	
Prescreen Variables	Yes
Prescreening Criterion	Default
Rank Percentage Cutoff	90
Response Model Selection	
Combined Model	Yes
Selection Method	Forward Selection
Selection Criterion	None
Significance Level for Entry	0.95
Stay Significance Level	0.05
Suppress Intercept	No
Two-Way Interaction	No
Outcome Model Selection	

Variables are ranked by NIV. Rank for X is 50 and R is ranked 100.

Cutoff at "90" will cause R (lowest NIV) to be removed

"Rank 90" removes "R"

Example #1: NIV Ranks for Inputs X and R

Variable Name	Net Information Value	Rank Percentile	Selection
X	108.4	50	Yes
R	30.9	100	No

- NIV of R is computed for 20 equal width bins (the default when possible).
- Since cutoff was 90% and R is in the 100% rank, then R is excluded.
- The Default = NIV unless a Validation Sample has been selected (e.g. by a Partition Node). Then Default is Penalized NIV. See SAS documentation.

No paper exists (as far as I know) that studies NIV vs. "Model Fit" perhaps as measured by Log Likelihood. … Good topic for a SAS User Group paper

Example #1: IRM Results for Response Model

- 4 levels of X, EM_IRM_TREATMENT, and X * Treat Interactions
- R was dropped during "Pre-Screening"

Analysis of Maximum Likelihood Estimates (omitted P-Value column)

Parameter		DF	Estimate	Standard Error	Wald Chi-Square
Intercept		1	-2.4294	0.0339	5128.86
X	1	1	0.4734	0.0641	54.48
X	2	1	-0.00857	0.0665	0.02
X	3	1	-0.0437	0.0697	0.39
X	4	1	-0.1589	0.0695	5.23
EM_IRM_TREATMENT	0	1	-0.6637	0.0339	382.83
X*EM_IRM_TREATMENT	1 0	1	0.0940	0.0641	2.15
X*EM_IRM_TREATMENT	2 0	1	0.1801	0.0665	7.34
X*EM_IRM_TREATMENT	3 0	1	-0.0821	0.0697	1.39
X*EM_IRM_TREATMENT	4 0	1	-0.0920	0.0695	1.75

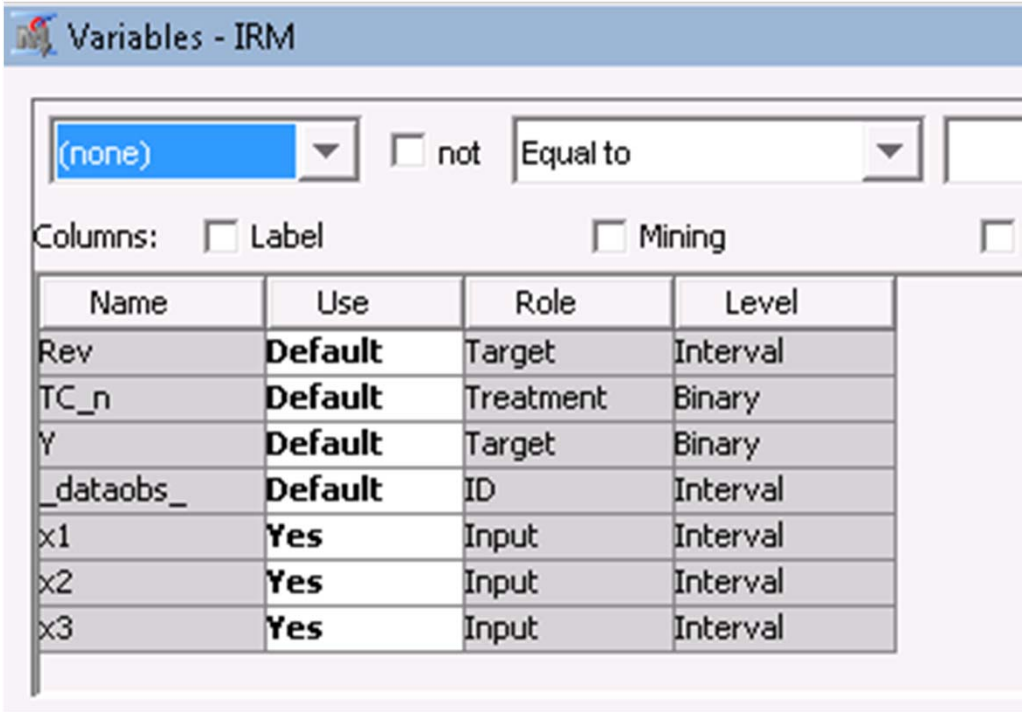
Example #2: Computing $\widehat{Y}_{Outcome\_T}$ and $\widehat{Y}_{Outcome\_C}$

A Data set constructed via random functions:

- 3 Inputs X1 X2 X3
- Y = binary target (buying)
- REV = interval target (revenue)
- TC_n = binary "treatment"

Train: 60%

Validation: 40%



Name	Use	Role	Level
Rev	**Default**	Target	Interval
TC_n	**Default**	Treatment	Binary
Y	**Default**	Target	Binary
dataobs	**Default**	ID	Interval
x1	**Yes**	Input	Interval
x2	**Yes**	Input	Interval
x3	**Yes**	Input	Interval

In Data Partition: Designate Y and TC_n as stratification variables

Example #2: Using Difference Model for $\widehat{Y}_{Outcome\_T}$ and $\widehat{Y}_{Outcome\_C}$

Train	
Variables	...
Treatment Level Selection	...
Prescreen Variables	Yes
Prescreening Criterion	Default
Rank Percentage Cutoff	95
Response Model Selection	
Combined Model	No
Selection Method	Forward Selection
Selection Criterion	None
Significance Level for Entry	0.5
Stay Significance Level	0.05
Suppress Intercept	No
Two-Way Interaction	No

X1, X2 pass the Prescreen (not X3). So X1, X2 are eligible for both Response and Outcome Models.

Outcome Model Selection	
Selection Method	Forward Selection
Selection Criterion	None
Significance Level for Entry	0.95
Stay Significance Level	0.05
Suppress Intercept	No
Two-Way Interaction	No
Revenue Calculation	
Use Constant Revenue	No
Revenue Per Response	10.0
Use Constant Cost	No
Cost	0.0
Report	
Number of Bins	10

Not used

For the Difference Model, because there is an Outcome Target, the Response Model will use the Probit Link … see this Later

Example #2: Difference Model – Results for Treated Group

Treated: **Response** Model (Target is Y … **Link is PROBIT**)

Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-2.5330	0.0157	26040.58	<.0001
x1	1	0.7911	0.0188	1774.25	<.0001
x2	1	0.1905	0.0177	116.32	<.0001

Treated: **Outcome** Model (Target is REV)

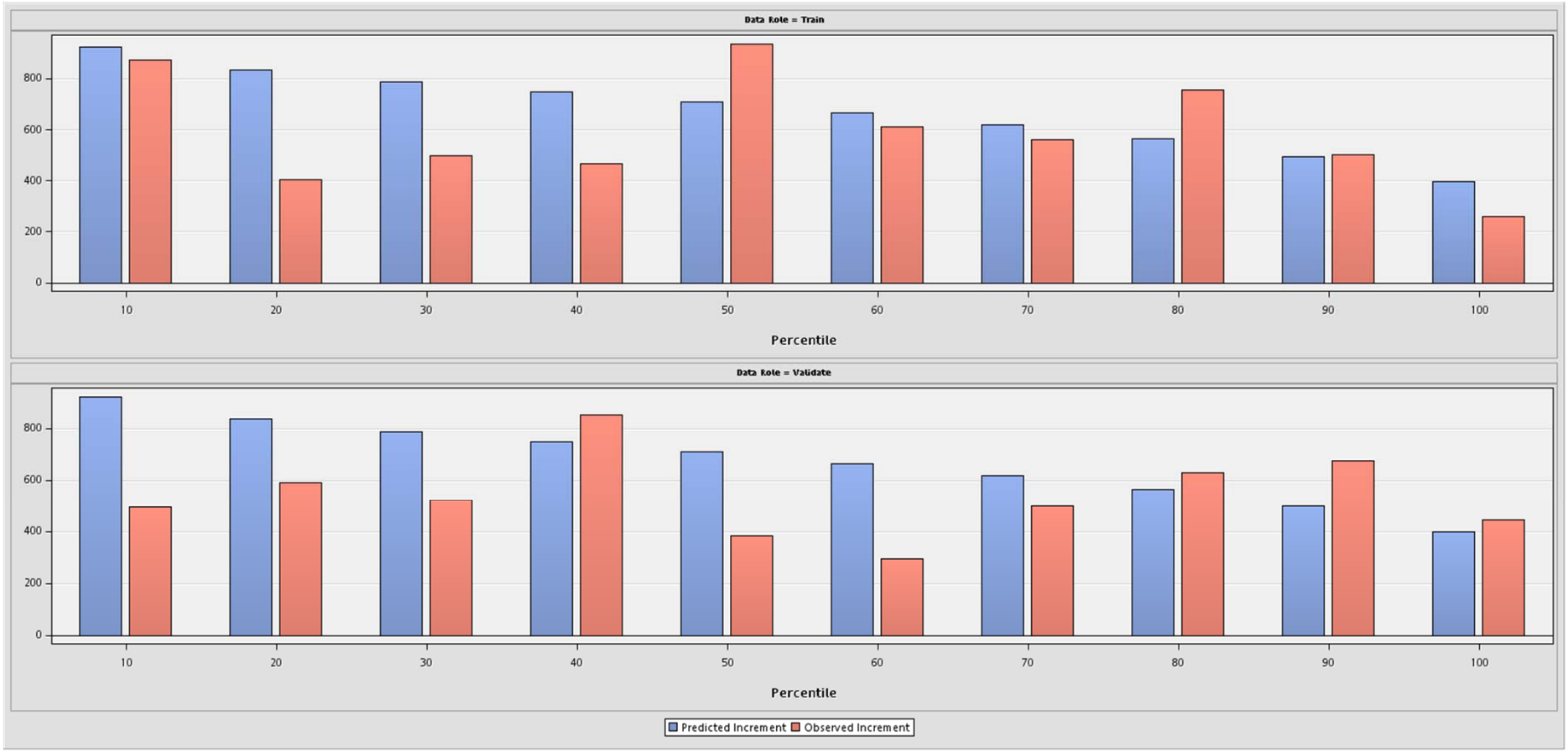
Parameter	DF	Estimate	Error	t Value	Pr > |t|
Intercept	1	-6770.9	14249.3	-0.48	0.6347
x1	1	-694.6	598.2	-1.16	0.2457
x2	1	272.6	122.7	2.22	0.0263
_IMR_1	1	13836.3	11424.7	1.21	0.2259

Inverse Mills Ratio – Uses Probit Score from Response

Example #2: Difference Model - Decile Ranks for Outcome

Blue "Predicted" and Orange "Observed". Top is TRAIN, Bottom is VALIDATION

… This Model gives Poor Predictive Power



Probability Decomposition Model (PDM)

Incremental Response … IR = Lift * S

S: An in-market model (PM) with no treatment … avg. might be 2%

Lift is applied to **S** to give rate (**Lift * S)** of incremental purchase

[More explanation of **Lift** is needed … see next slide]

If **Lift** is 5% and **S** is 2%, then **Lift***S = 5% * 2% = 0.1% (1 per 1,000)

Lift could be negative

Lift * S = Incremental Rate from prospect (e.g. 5% * 2% = 0.1%

PDM … involves two probabilities

IR = Lift * S

S: All-purpose in-market model (PM) …. **Probability #1**

Lift : Built on a campaign. This model is fitted ONLY on the responders (treated and control) in campaign. Y = 1 if responder is treated, else Y = 0. [If Treated and Control are not same size, then "weights" are needed]

Among Responders from appended T & C, a Logistic Model is fit to predict:

P_T - probability that the Responder was treated. …. **Probability #2**

Let $P_C = 1 - P_T$ …. The incremental rate due to making an offer can be computed by **Lift**

$$\text{Lift} = (P_T - P_C) / P_C$$

This model does not address variable Revenue (Profit)
This is a modification of the model of Jun Zhong (2009). See Lund (2012)

Data Sets Used in Examples: Test_NIV

```
Data NetL.Test_NIV;
input X Y T F;
datalines;
1 0 0 1000
1 1 0 80
1 0 1 800
1 1 1 200
2 0 0 1300
2 1 0 70
2 0 1 1200
2 1 1 170
3 0 0 1500
3 1 0 60
3 0 1 900
3 1 1 160
4 0 0 1700
4 1 0 60
4 0 1 1000
4 1 1 160
5 0 0 1900
5 1 0 60
5 0 1 1100
5 1 1 160
;
run;
```

Data Sets Used in Examples: A_Sim

```
%let seed = 12345;
%let seed2 = 13579;
Data A_Sim2;
    do i = 1 to 1000000;
        x1 = ranuni(&seed); x2 = ranuni(&seed); x3 = ranuni(&seed);
        if mod(i,4) in(0 2) then TC = "T"; else TC = "C";  /* 50% of obs are control (C) */
        TC_n = (TC = "T");
/* Approx 2% of "C" are buyers.  x1 is strongly associated with buying */
        if TC = "C" then do;
            if ranuni(&seed) < .04 * x1 then Y = 1;
            else Y = 0;
            end;
/* Approx 2.3% of "T" are buyers. x2 increases buy-rate while -0.05 scales buy-rate */
        else do;
            if ranuni(&seed) < .04 * (x1 + .25*x2 - .05) then Y = 1;
            else Y = 0;
            end;
        target2 = 0; /* initialize to 0. Treated group redemptions */
        if TC = "T" & Y = 1 /* treated group redemptions */
        then
        do;
/* x3 increases redemptions, x1 reduces. Redemption rate is approx. 48% */
            if ranuni(&seed) >= .48 + .20*x3 - .10*x1 then target2 = 1; * else target2 = 0;
            end;
        if Y = 1 then Rev = floor(ranuni (&seed2)*10000 + 5000 - 750*target2 + TC_n*ranuni (&seed2)*1800 + x1*100);
        else Rev = .; /* Setting Rev to missing if no Response */
        output;
        end;
run;
```

Appendix: What is _IMR_1 ?

The linear regression of Revenue fitted only to Buyers may produce biased estimates when the Revenue model is applied to a new data set. This is because the Non-Buyers "self-select" out of the sample instead of being removed by a random process.

"Heckman adjustment" assumes there is a "selection equation" that determines if a prospect "i" will be a buyer: $Y_i = X_i \beta + \nu_i$ where X_i are Inputs, β are unknown parameters, and ν_i is an error term.

Prospect "i" will be a buyer if Y_i reaches a threshold which is: $Y_i > 0$.

For customers where $Y_i > 0$ the equation for Revenue is: $REV_i = Z_i \eta + \mu_i$

where Z_i are Inputs, η are unknown parameters, and μ_i is an error term.

But now Rev depends on Y … as shown on the next slide.

Parameter	DF	Estimate	Error	t Value
_IMR_1	1	13836.3	11424.7	1.21

Appendix: What is _IMR_1 ?

The expected value of REV is not simply $Z^*\eta$. When taking the dependency of Y into account, then

$$E[REV \mid Z \text{ and } Y > 0] = Z^*\eta + E[\mu \mid \nu > -X\beta]$$

A formula for the "bias term" $E[\mu \mid \nu > -X\beta]$ is needed.

Heckman (1979) made assumptions leading to a formula for $E[\mu \mid \nu > -X\beta]$

$$E[\mu \mid \nu > -X\beta] = \rho\, \sigma_\mu\, \phi(X\beta) / \Phi(X\beta)$$

where ϕ is the std normal density and Φ is the cum std normal distribution. Each are evaluated at $X_i\,\beta$. Then $\phi(X_i\,\beta) / \Phi(X_i\,\beta)$ = _IMR_1 is the inverse Mills ratio (for treated)

For prospect "i" the quantity _IMR_1 is estimated from the Response Model. The bias-corrected regression for REV becomes:

$$REV_i = Z_i\,\eta + \_IMR\_1\,(X_i\,\beta)\, b_M + \mu_i$$

where η and b_M are fitted by regression on REV. (See Math. Appendix)

Mathematical Appendix

The equation $E[\mu \mid \nu > -X\beta] = \rho\, \sigma_\mu\, \phi(X\beta) / \Phi(X\beta)$ is established subject to the following assumptions:

(1) There is a latent variable equation $Y^*_i = X_i\, \beta + \nu_i$ for the ith prospect

where X_i are predictors and ν_i is a random error term

This equation determines whether a prospect "responds" according to:

$$Y_i = 1 \text{ if } Y^*_i > 0 \text{ else } Y_i = 0$$

(2) A second equation gives REV for the set of responders ($Y_i = 1$) through:

$REV_i = Z_i\, \eta + \mu_i$ where Z_i are predictors and μ_i is a random error term

(3) It is assumed that (ν, μ) are jointly normal with means of 0, standard deviations of 1 and σ (respectively) and correlation of ρ

Mathematical Appendix

The equation $E[\mu \mid v > -X\beta] = \rho\, \sigma_\mu\, \varphi(X\beta) / \Phi(X\beta)$ is established by the following steps:

- The definition of a conditional density gives:
- $f(\mu \mid v > -X\beta) = f(v, \mu) / P(v > -X\beta)$ and where $f(v, \mu)$ is not defined for $v \leq -X\beta$
- $P(v > -X\beta) = \int_{-\infty}^{\infty} \int_{-X\beta}^{\infty} f(v,\mu)\, dv\, d\mu$
- $E[\mu \mid v > -X * \beta] = \int_{-\infty}^{\infty} \int_{-X\beta}^{\infty} \mu f(v,\mu)\, dv\, d\mu \;\; / \;\; P(v > -X * \beta)$
- These double integrals are evaluated using $f(v, \mu)$ with a bivariate normal density to yield the results as shown below in (A) and (B) … recall that it is assumed: $(v, \mu) \sim N(0, 0, \sigma_v, \sigma_\mu, \rho)$ and $\sigma_v = 1$

A. $P(v > -X\beta) = \int_{-\infty}^{\infty} \int_{-X\beta}^{\infty} f(v,\mu)\, dv\, d\mu = \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} f(v,\mu)\, d\mu\, dv = \int_{-X\beta}^{\infty} f_v(v)\, dv = 1 - \Phi(-X\beta) = \Phi(X\beta)$

where $f_v(v)$ is the marginal distribution for v which is normal with mean 0 and variance 1 and where the change of the order of integration is justified by standard theory.

B. $$\int_{-\infty}^{\infty} \int_{-X\beta}^{\infty} \mu\, f(v,\mu)\, dv\, d\mu = \frac{1}{2\pi\sigma_\mu\sqrt{1-\rho^2}} \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} \mu \exp\left\{\frac{-\left(\left(\frac{\mu}{\sigma_\mu}\right)^2 - \frac{2\rho\mu v}{\sigma_\mu} + v^2\right)}{2(1-\rho^2)}\right\} d\mu\, dv$$

Mathematical Appendix

First μ is standardized by transforming w = μ/σ_μ. This gives:

$$= \frac{\sigma_\mu}{2\pi\sqrt{1-\rho^2}} \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} w \exp\{\frac{-(w^2 - 2\rho w v + v^2)}{2(1-\rho^2)}\} dw dv$$

Next the transformation z = (w – ρv) / sqrt(1-ρ^2) gives:

$$= \frac{\sigma_\mu}{2\pi} \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} \left(z\sqrt{1-\rho^2} + \rho v\right) \exp\{\frac{-(z^2+v^2)}{2}\} dz dv$$

$$= \frac{\sigma_\mu}{2\pi} \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} \left(z\sqrt{1-\rho^2}\right) \exp\{\frac{-(z^2+v^2)}{2}\} dz dv + \frac{\rho\sigma_\mu}{2\pi} \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} v \exp\{\frac{-(z^2+v^2)}{2}\} dz dv$$

Left-hand side integral has an integrand in "z" that is an odd-function with respect to 0 for each v. So, integral of z $\exp\left\{\frac{-(z^2+v^2)}{2}\right\}$ over z from -∞ to +∞ is zero and the left-hand side integral is zero.

The right-hand side integral is simplified by:

$$\frac{\rho\sigma_\mu}{2\pi} \int_{-X\beta}^{\infty} \int_{-\infty}^{\infty} v \exp\{\frac{-(z^2+v^2)}{2}\} dz dv = \frac{\rho\sigma_\mu}{2\pi} \int_{-X\beta}^{\infty} v \exp\{\frac{-v^2}{2}\} \int_{-\infty}^{\infty} \exp\{\frac{-z^2}{2}\} dz dv$$

The inside integral in z is integrated to obtain $\sqrt{2\pi}$. This leaves:

$$= \frac{\rho\sigma_\mu}{2\pi} \sqrt{2\pi} \int_{-X\beta}^{\infty} v \exp\{\frac{-v^2}{2}\} = \rho\, \sigma_\mu \frac{1}{\sqrt{2\pi}} \exp\{\frac{-X\beta^2}{2}\} = \rho\, \sigma_\mu\, \varphi(X\beta)$$

In summary: E[μ | v > - X * β] = $\int_{-\infty}^{\infty} \int_{-X\beta}^{\infty} \mu f(v,\mu) dv d\mu$ / P(v > - X * β) = ρ σ_μ φ(Xβ) / Φ(Xβ)

See Heckman, J. (1979). "Sample Selection Bias as a Specification Error." Econometrica 47:153–161.