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
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.
Some responders within the Treatment Group are “incremental” while other responders would have responded without the offer. IRM tries to predict “which are which”.

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

<table>
<thead>
<tr>
<th>For each Prospect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment: 1 = treated or 0 = control</td>
</tr>
<tr>
<td><strong>Response:</strong> 1 = responded or 0 = not responded</td>
</tr>
<tr>
<td>Predictors for <strong>Response:</strong> Xi</td>
</tr>
<tr>
<td>Optional: <strong>Outcome</strong> (Revenue or Profit): Usually dollar amount</td>
</tr>
<tr>
<td>Outcome equals “missing” for non-responders</td>
</tr>
<tr>
<td>Optional: Predictors for <strong>Outcome:</strong> Zi</td>
</tr>
<tr>
<td>Xi and Zi may overlap.</td>
</tr>
</tbody>
</table>
History and References

<table>
<thead>
<tr>
<th>Reference</th>
<th>Details</th>
</tr>
</thead>
</table>
A **link function** is a function that transforms the expected response $\mu$ to a linear combination of predictors $\sum \beta_j X_j$. That is, $\text{LINK}(\mu) = \sum \beta_j X_j = \beta \ast X$

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

**Logit Link:**
$$\log \left( \frac{\mu}{1 - \mu} \right) = \beta \ast X$$
**Equivalently:**
$$\mu = P(Y=1) = e^{\beta \ast X} \bigg/ (1 + e^{\beta \ast X})$$

**Probit Link:**
$$\Phi^{-1}(\mu) = \beta \ast X$$
where $\Phi^{-1}$ is the inverse of the standard normal cumulative distribution.

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

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.
A. Treated and control groups are appended.

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

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

D. Fit to Response \( Y \). The Combined Model is logistic:
\[
\hat{Y}_{\text{Resp}} = P(Y=1) = \frac{\exp(\alpha + \beta X + \gamma T + \phi(X*T))}{1 + \exp(\alpha + \beta X + \gamma T + \phi(X*T))}
\]

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

... **Now for each prospect compute 2 probabilities:**

1. \( \hat{Y}_{\text{Resp}_T} = P(Y=1 \mid T=1) = \frac{\exp(1 + 2X + 3 + 4X)}{1 + \exp(1 + 2X + 3 + 4X)} \)
   \[= \frac{\exp(4 + 6X)}{1 + \exp(4 + 6X)} \]

2. \( \hat{Y}_{\text{Resp}_C} = P(Y=1 \mid T=0) = \frac{\exp(1 + 2X)}{1 + \exp(1 + 2X)} \)

The difference score (incremental rate) is: \( \hat{D}S = \hat{Y}_{\text{Resp}_T} - \hat{Y}_{\text{Resp}_C} \)
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: \( \hat{Y}_{\text{Outcome}} = \eta Z + \delta T + \psi (Z*T) \) ... Fit for Outcome ≠ . (Responders)

Set \( T = 1 \) to obtain \( \hat{Y}_{\text{Outcome}_T} \) and \( T = 0 \) for \( \hat{Y}_{\text{Outcome}_C} \)

Finally, there is EXPECTED incremental Outcome for a prospect:

\[
\hat{R}_{\text{Incr}_T} = \hat{Y}_{\text{Resp}_T} \times \hat{Y}_{\text{Outcome}_T} - \hat{Y}_{\text{Resp}_C} \times \hat{Y}_{\text{Outcome}_C} - \text{Fixed Cost}
\]

(Estimated outcome for a treated responder) = \( \hat{Y}_{\text{Resp}_T} \) (Likelihood of Response if Treated)

(Estimated outcome for a control responder) = \( \hat{Y}_{\text{Resp}_C} \) (Likelihood of Response if Control)
Why not set Outcome to 0 for non-Responders?

\[ \hat{Y}_{\text{Outcome}} = \eta^*Z + \delta^*T + \psi^*(Z*T) \]... 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 ...
Suppose there is **one** predictor $X$.

**For treated “$T$”:** A model is fit to response $Y=1$ vs $Y=0$. Let $\hat{Y}_{\text{Resp}_T} = P(Y=1 \mid T)$

If no Outcome target, then logistic is used:

$$\hat{Y}_{\text{Resp}_T} = \frac{e^{\alpha_T + \beta_T X}}{1 + e^{\alpha_T + \beta_T X}}$$

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

**Likewise for the control group “$C$”:** $\hat{Y}_{\text{Resp}_C} = P(Y=1 \mid C)$

The difference score (incremental rate) is:

$$\hat{DS} = \hat{Y}_{\text{Resp}_T} - \hat{Y}_{\text{Resp}_C}$$

Of course, the predictors for Treated Model and Control Model can be completely different.
If there is Outcome Target: Linear Regression predicts the Outcome.
Model is fit when Outcome ≠ . (Responders)

If “T”, then \( \hat{Y}_{\text{Outcome}_T} = \eta_T Z + b_T M \)
This added term \( b_T M \) 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
\( \phi \) is std normal density and \( \Phi \) is std normal cum. distribution
The coefficient \( b_T \) is fit by the regression

Theory of using Inverse Mills ratio is discussed in Appendix A.
[\( \Rightarrow \) Using Logistic instead of Probit in \( M \) would give very similar results.]

Likewise for “C” to give \( \hat{Y}_{\text{Outcome}_C} = \eta_C Z + b_C M \)

Finally, there is the expected incremental Outcome for a prospect:
\( \hat{R}_{\text{Incr}_T} = \hat{Y}_{\text{Resp}_T} * \hat{Y}_{\text{Outcome}_T} - \hat{Y}_{\text{Resp}_C} * \hat{Y}_{\text{Outcome}_C} - \text{Fixed Cost} \)
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.

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

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0</td>
<td>1</td>
<td>( P_T(X=i \mid Y=0) )</td>
<td>( P_T(X=i \mid Y=1) )</td>
</tr>
<tr>
<td>1</td>
<td>800</td>
<td>200</td>
<td>16.0%</td>
<td>23.5%</td>
</tr>
<tr>
<td>2</td>
<td>1200</td>
<td>170</td>
<td>24.0%</td>
<td>20.0%</td>
</tr>
<tr>
<td>3</td>
<td>900</td>
<td>160</td>
<td>18.0%</td>
<td>18.8%</td>
</tr>
<tr>
<td>4</td>
<td>1000</td>
<td>160</td>
<td>20.0%</td>
<td>18.8%</td>
</tr>
<tr>
<td>5</td>
<td>1100</td>
<td>160</td>
<td>22.0%</td>
<td>18.8%</td>
</tr>
<tr>
<td>SUM</td>
<td>5000</td>
<td>850</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

\[ \text{WOE} = \log(16\% / 23.5\%) = 0.38566 \]
Net Weight of Evidence (NWOE) for X

<table>
<thead>
<tr>
<th>Treated</th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0</td>
<td>1</td>
<td>P_T(X=i</td>
</tr>
<tr>
<td>1</td>
<td>800</td>
<td>200</td>
<td>16.0%</td>
</tr>
<tr>
<td>2</td>
<td>1200</td>
<td>170</td>
<td>24.0%</td>
</tr>
<tr>
<td>3</td>
<td>900</td>
<td>160</td>
<td>18.0%</td>
</tr>
<tr>
<td>4</td>
<td>1000</td>
<td>160</td>
<td>20.0%</td>
</tr>
<tr>
<td>5</td>
<td>1100</td>
<td>160</td>
<td>22.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control</th>
<th>Y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0</td>
<td>1</td>
<td>P_C(X=i</td>
</tr>
<tr>
<td>1</td>
<td>1000</td>
<td>80</td>
<td>13.5%</td>
</tr>
<tr>
<td>2</td>
<td>1300</td>
<td>70</td>
<td>17.6%</td>
</tr>
<tr>
<td>3</td>
<td>1500</td>
<td>60</td>
<td>20.3%</td>
</tr>
<tr>
<td>4</td>
<td>1700</td>
<td>60</td>
<td>23.0%</td>
</tr>
<tr>
<td>5</td>
<td>1900</td>
<td>60</td>
<td>25.7%</td>
</tr>
</tbody>
</table>

This concept was presented by K Larsen (2009)

\[
T - C = \text{NWOE}
\]

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)

### Treated

<table>
<thead>
<tr>
<th>X</th>
<th>( P_T(X=i \mid Y=0) )</th>
<th>( P_T(X=i \mid Y=1) )</th>
<th>WOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.0%</td>
<td>23.5%</td>
<td>0.38566</td>
</tr>
<tr>
<td>2</td>
<td>24.0%</td>
<td>20.0%</td>
<td>-0.18232</td>
</tr>
<tr>
<td>3</td>
<td>18.0%</td>
<td>18.8%</td>
<td>0.04474</td>
</tr>
<tr>
<td>4</td>
<td>20.0%</td>
<td>18.8%</td>
<td>-0.06062</td>
</tr>
<tr>
<td>5</td>
<td>22.0%</td>
<td>18.8%</td>
<td>-0.15593</td>
</tr>
</tbody>
</table>

### Control

<table>
<thead>
<tr>
<th>X</th>
<th>( P_C(X=i \mid Y=0) )</th>
<th>( P_C(X=i \mid Y=1) )</th>
<th>WOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.5%</td>
<td>24.2%</td>
<td>0.58441</td>
</tr>
<tr>
<td>2</td>
<td>17.6%</td>
<td>21.2%</td>
<td>0.18852</td>
</tr>
<tr>
<td>3</td>
<td>20.3%</td>
<td>18.2%</td>
<td>-0.10873</td>
</tr>
<tr>
<td>4</td>
<td>23.0%</td>
<td>18.2%</td>
<td>-0.23390</td>
</tr>
<tr>
<td>5</td>
<td>25.7%</td>
<td>18.2%</td>
<td>-0.34512</td>
</tr>
</tbody>
</table>

\[
P_T(X=i \mid Y=0) \cdot P_C(X=i \mid Y=0) - P_T(X=i \mid Y=1) \cdot P_C(X=i \mid Y=1)
\]

NIV = Sum \* 1000 = 108.4

<table>
<thead>
<tr>
<th>X</th>
<th>NWOE (a)</th>
<th>Weight (b)</th>
<th>= (a) * (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.1988</td>
<td>-0.0070</td>
<td>0.0014</td>
</tr>
<tr>
<td>2</td>
<td>-0.3708</td>
<td>-0.0158</td>
<td>0.0058</td>
</tr>
<tr>
<td>3</td>
<td>0.1535</td>
<td>0.0054</td>
<td>0.0008</td>
</tr>
<tr>
<td>4</td>
<td>0.1733</td>
<td>0.0069</td>
<td>0.0012</td>
</tr>
<tr>
<td>5</td>
<td>0.1892</td>
<td>0.0083</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

If Treatment is strong for \( X_i \), then \( C \gg A \) and weight \( B \* C - A \* D \gg 0 \)
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
Example #1: Diagram Including a SAS Code Node

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

```sas
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

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)

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.
Example #1: NIV Ranks for Inputs X and R

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Net Information Value</th>
<th>Rank Percentile</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>108.4</td>
<td>50</td>
<td>Yes</td>
</tr>
<tr>
<td>R</td>
<td>30.9</td>
<td>100</td>
<td>No</td>
</tr>
</tbody>
</table>

- 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)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.4294</td>
<td>0.0339</td>
<td>5128.86</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>0.4734</td>
<td>0.0641</td>
<td>54.48</td>
</tr>
<tr>
<td>X</td>
<td>2</td>
<td>-0.00857</td>
<td>0.0665</td>
<td>0.02</td>
</tr>
<tr>
<td>X</td>
<td>3</td>
<td>-0.0437</td>
<td>0.0697</td>
<td>0.39</td>
</tr>
<tr>
<td>X</td>
<td>4</td>
<td>-0.1589</td>
<td>0.0695</td>
<td>5.23</td>
</tr>
<tr>
<td>EM_IRM_TREATMENT</td>
<td>0</td>
<td>-0.6637</td>
<td>0.0339</td>
<td>382.83</td>
</tr>
<tr>
<td>X*EM_IRM_TREATMENT</td>
<td>1</td>
<td>0.0940</td>
<td>0.0641</td>
<td>2.15</td>
</tr>
<tr>
<td>X*EM_IRM_TREATMENT</td>
<td>2</td>
<td>0.1801</td>
<td>0.0665</td>
<td>7.34</td>
</tr>
<tr>
<td>X*EM_IRM_TREATMENT</td>
<td>3</td>
<td>-0.0821</td>
<td>0.0697</td>
<td>1.39</td>
</tr>
<tr>
<td>X*EM_IRM_TREATMENT</td>
<td>4</td>
<td>-0.0920</td>
<td>0.0695</td>
<td>1.75</td>
</tr>
</tbody>
</table>
Example #2: Computing $\hat{Y}_{\text{Outcome}_T}$ and $\hat{Y}_{\text{Outcome}_C}$

A Data set constructed via random functions:
- 3 Inputs $X_1$ $X_2$ $X_3$
- $Y$ = binary target (buying)
- $\text{REV}$ = interval target (revenue)
- $\text{TC}_n$ = binary “treatment”

Train: 60%
Validation: 40%

In Data Partition: Designate $Y$ and $\text{TC}_n$ as stratification variables
Example #2: Using Difference Model for $\hat{Y}_{\text{Outcome\_T}}$ and $\hat{Y}_{\text{Outcome\_C}}$

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

X1, X2 pass the Prescreen (not X3). So X1, X2 are eligible for both Response and Outcome Models.
Example #2: Difference Model – Results for Treated Group

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.5330</td>
<td>0.0157</td>
<td>26040.58</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x1</td>
<td>1</td>
<td>0.7911</td>
<td>0.0188</td>
<td>1774.25</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>0.1905</td>
<td>0.0177</td>
<td>116.32</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**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 \( \text{Lift} \times \text{S} = 5\% \times 2\% = 0.1\% \) (1 per 1,000)

**Lift** could be negative

**Lift * S** = Incremental Rate from prospect (e.g. 5% * 2% = 0.1%)
PDM ... involves two probabilities

\[ \text{IR} = \text{Lift} \times \text{S} \]

**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**

\[ \text{Lift} = \frac{(P_T - P_C)}{P_C} \]

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

```r
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

```sas
%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:
```

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: \( Y_i = X_i \beta + v_i \) where \( X_i \) are Inputs, \( \beta \) are unknown parameters, and \( v_i \) 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: \( \text{REV}_i = Z_i \eta + \mu_i \) where \( Z_i \) are Inputs, \( \eta \) are unknown parameters, and \( \mu_i \) is an error term.

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>_IMR_1</td>
<td>1</td>
<td>13836.3</td>
<td>11424.7</td>
<td>1.21</td>
</tr>
</tbody>
</table>
Appendix: What is _IMR_1?

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

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

A formula for the “bias term” $E[\mu \mid v > -X\beta]$ is needed. Heckman (1979) made assumptions leading to a formula for $E[\mu \mid v > -X\beta]$

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

where $\phi$ is the std normal density and $\Phi$ is the cum std normal distribution. Each are evaluated at $X_i \beta$. Then $\phi(X_i \beta) / \Phi(X_i \beta) = _{\text{IMR}_1}$ 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:

$$REV_i = Z_i \eta + _{\text{IMR}_1} (X_i \beta) b_M + \mu_i$$

where $\eta$ and $b_M$ are fitted by regression on REV. (See Math. 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 $i^{th}$ prospect
   where $X_i$ are predictors and $\nu_i$ is a random error term
   This equation determines whether a prospect “responds” according to:
   
   $Y_i = 1$ if $Y^*_i > 0$ 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 $\mu_i$ is a random error term

(3) It is assumed that ($\nu$, $\mu$) are jointly normal with means of 0, standard deviations of 1 and $\sigma$ (respectively) and correlation of $\rho$
The equation \( E[\mu | v > - X\beta] = \rho \sigma_\mu \phi(X\beta) / \Phi(X\beta) \) is established by the following steps:

- The definition of a conditional density gives:
  \[
  f(\mu | v > - X\beta) = \frac{f(v, \mu)}{P(v > - X\beta)} \quad \text{and where } f(v, \mu) \text{ is not defined for } v \leq - X\beta
  \]

- \( P(v > - X\beta) = \int_{-\infty}^\infty \int_{-X\beta}^{\infty} f(v, \mu) dv d\mu \)

- \( E[\mu | v > - X\beta] = \int_{-\infty}^\infty \int_{-X\beta}^{\infty} \mu f(v, \mu) dv d\mu / P(v > - X\beta) \)

- 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.** \[
\int_{-\infty}^{\infty} \int_{-X\beta}^{\infty} \mu f(v, \mu) dv d\mu = \frac{1}{2\pi \sigma_\mu \sqrt{1-\rho^2}} \int_{-X\beta}^{\infty} \int_{-\infty}^\infty \mu \exp\left\{ -\frac{\left( \frac{\mu}{\sigma_\mu} - \frac{2\mu v}{\sigma_v} \right)^2 + v^2}{2(1-\rho^2)} \right\} d\mu dv
\]
First \( \mu \) is standardized by transforming \( w = \frac{\mu}{\sigma_\mu} \). This gives:

\[
= \frac{\sigma_\mu}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w \exp\left\{-\frac{\left(w^2 - 2\rho w v + v^2\right)}{2(1-\rho^2)}\right\} dv dw
\]

Next the transformation \( z = (w - \rho_\nu) / \sqrt{1-\rho^2} \) gives:

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

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

Left-hand side integral has an integrand in “z” that is an odd-function with respect to 0 for each \( \nu \).

So, integral of \( z \exp\left\{-\frac{(z^2 + v^2)}{2}\right\} \) over \( z \) from \(-\infty\) to \(+\infty\) is zero and the left-hand side integral is zero.

The right-hand side integral is simplified by:

\[
\frac{\rho \sigma_\mu}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v \exp\left\{-\frac{(z^2 + v^2)}{2}\right\} dz dv = \frac{\rho \sigma_\mu}{2\pi} \int_{-\infty}^{\infty} v \exp\left\{-\frac{v^2}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz dv
\]

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

\[
= \frac{\rho \sigma_\mu}{2\pi} \sqrt{2\pi} \int_{-\infty}^{\infty} v \exp\left\{-\frac{v^2}{2}\right\} = \rho \sigma_\mu \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{X_\beta^2}{2}\right\} = \rho \sigma_\mu \phi(X_\beta)
\]

In summary: \( E[\mu \mid \nu > -X \ast \beta] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu f(v, \mu) dv d\mu \) / \( P(\nu > -X \ast \beta) = \rho \sigma_\mu \phi(X_\beta) / \Phi(X_\beta) \)