

Logistic Models for Net Lift & Propensity Scores - Applied to Marketing v06

by Bruce Lund

Statistical Modeling Consultant and Trainer, Novi, MI

blund_data@mi.rr.com and blund.data@gmail.com

Send an email for a copy of slides



This Presentation:

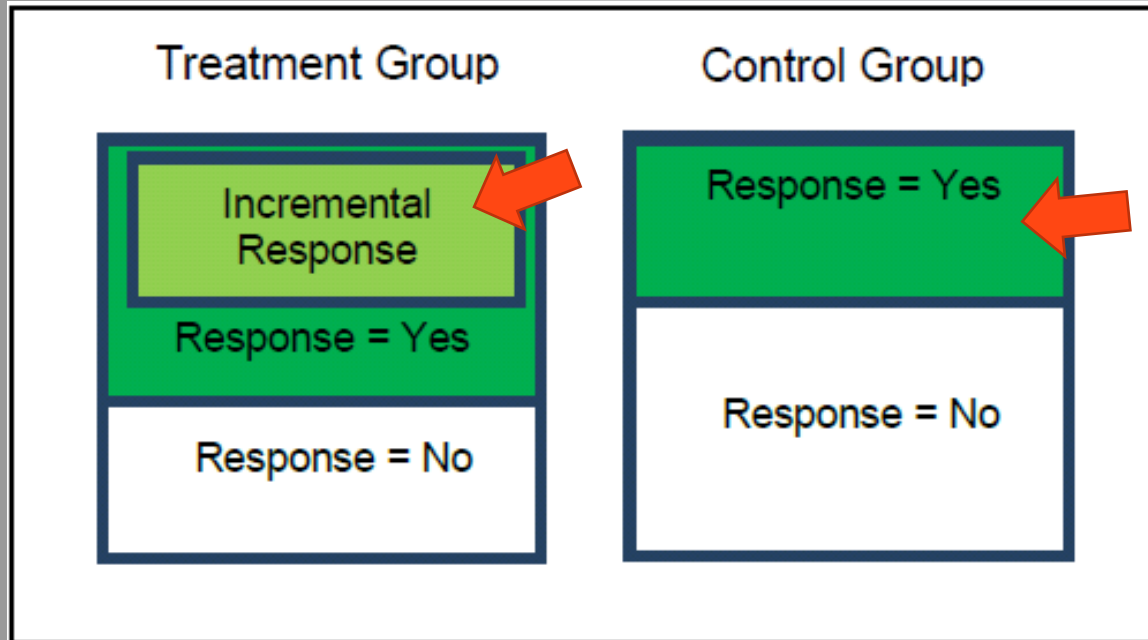
Talks about two Topics where Logistic Models play an important role ...

1. Net Lift Modeling
2. Propensity Scoring applied to Causal Inference

Net Lift (=Incremental Response) Modeling. Let's Begin

An Example from Automotive Direct Marketing

Automotive Marketing: Two groups of customers (treated and control ... randomly split) are taken from the Customer Database. Treated group receives an offer of \$750 cash-back to buy a new-vehicle within next 90 days. This is a private offer – extended by mail or email.



Some households in control group will buy a vehicle in the next 90 days because of various needs. These are the control "responders". Maybe, 0.5% would buy (no cash-back).

In the treatment group (with \$750), maybe the buy-rate is 1.5%. But of these, some would have bought anyway (needed a vehicle) but others were motivated to buy because of the \$750 (wouldn't have bought, at least now), without the cash. These are the "incremental responders".

The Goal of Net Lift Models is to identify the "incremental responders". Then the Marketing Department can target these prospects

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

Focus is Marketing Campaigns, but other applications (*)

- An “**Incremental Response (IR)**” by a prospect is a response that would not have occurred without the marketing campaign “treatment”.
- **IR Model** (or Net Lift) predicts **rate** at which prospects are “incremental” due to treatment.
 - E.g.: An incremental response model rate of 0.02 means the model *predicts* that 2% of these prospects make an **incremental** response.
 - Note: An IR rate for some prospects can be negative ... offer may “back-fire”
- We cannot say John or Mary Doe's response was incremental, unless we interviewed them, but they might not know! – reasons might be subconscious.
- An **IR Model** is different from a **Propensity Model (PM)**. PM gives probability that a prospect will respond. Until recent years at my automotive company, PM's were all that were available.
 - Prospects with upper mid-range scores were targeted on the hope that they would be incremental. Top range was excluded “They would buy anyway”.

Fitting an Incremental Response Model comes at a cost




A pilot marketing campaign must be conducted which has sufficient scale to allow good incremental response modeling.

The organization must:

1. Create treated and control groups for the pilot where ...
 - Selection of prospects **is wide enough** to encompass future campaigns selections
 - Organization must be committed to offering **same** treatment in future campaigns
2. Communicate a treatment having a momentary value to this treatment group
3. Perform post campaign processing to determine which prospects responded
4. Develop an Incremental Response Model

Note: Cost of the pilot could be substantial

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.
- Larsen, Kim (2009) introduced Net Lift Modeling at M2009, 12th Annual Data Mining Conference, Las Vegas.
- Lee T., Zhang R., Meng X., and Ryan L. (2013), "Incremental Response Modeling Using SAS® Enterprise Miner", *SAS Global Forum 2013* 
- Lavery, R. (2016) An Animated Guide: Incremental Response Modeling in Enterprise Miner, *WUSS 2016*.
Link: https://www.lexjansen.com/wuss/2016/135_Final_Paper_PDF.pdf
- Lund, Bruce (2016), Incremental Response Modeling with SAS® EM (14.1), *MSUG Feb 2016*
Link: http://www.misug.org/uploads/8/1/9/1/8191072/blund_incremental_response.pdf
- Michel, R., Schnakenburg, I., von Martens, T. (2019), *Targeting Uplift*, Springer, Cham, Switzerland. 
 - See the extensive references in Michel, et. al.
- SAS Institute Inc. 2018. SAS® Enterprise Miner™ 15.3: Reference Help. Cary, NC: SAS Institute Inc 

SAS added Incremental Response Node to [Enterprise Miner](#) in 2013 (See paper by Lee, et al. above) 

Some Incremental Response Model Methods

Difference Model (This is included in SAS EM, Incremental Response Node)

This will be our first example to illustrate the important concepts

Combined Model (Included in SAS EM, Incremental Response Node) ... in Appendix

Other methods are discussed in the book Michel, et al (2019) and one is discussed in Appendix

For ALL modeling methods this information is REQUIRED:

A large sample of Prospects is needed. For each Prospect there will be:

Treatment T: 1 = treated or 0 = control ... random assignment

Response Y: 1 = responded or 0 = not responded ... determined after the campaign

Assignment to TRAIN or VALIDATE ... random assignment ... to fit and then validate IR Model

Predictors X's for the Prospects which are used in fitting the IR Model

NEXT

The Difference Model

A lot of SAS code to come

... follow along and then get the slides from me

A Marketing Campaign Dataset for Illustration of Difference Model

```
%let seed = 12345;
DATA Campaign;
do i = 1 to 250000;
x1 = ranuni(&seed); x2 = ranuni(&seed); x3 = ranuni(&seed);
if mod(i,4) in(0 2) then T = 1; else T = 0;
if T = 0 then do;
if ranuni(&seed) < .04*x1 + 0.004*x3 then Y = 1;
else Y = 0;
end;
else do;
if ranuni(&seed) < .07 * (1.3*x1 + .20*x2 - .05) then Y = 1;
else Y = 0;
end;
output;
end;
run;
```

Regard this code as a black box

```
PROC FORMAT;
value T_C
1 = "TREAT" 0 = "CONTROL";
/* Divide Campaign into TRAIN and VALIDATE */
DATA Train Validate; SET Campaign;
if ranuni(123) < .5 then OUTPUT Train;
else OUTPUT Validate;
run;
PROC FREQ Data = Train;
Table T * Y /nocol nopercnt; format T T_C;
PROC FREQ Data = Validate;
Table T * Y /nocol nopercnt; format T T_C;
run;
```

There are 3 predictors: X1 X2 X3

Select PROSPECTS from DATABASE.

- Randomly assign Treatment T=1 to 50%
 - Communicate the offer to Treated
 - Record the Y's for all PROSPECTS
 - Randomly split PROSPECTS into TRAIN & VALIDATE
- Fit model to the TRAIN dataset
Validate on the VALIDATION

T (treatment)	Y (Response)		
	0	1	Total
CONTROL=0	61060	1398	62458
	97.76	2.24	
TREAT=1	59490	3064	62554
	95.1	4.90	
Total	120550	4462	125012

T (treatment)	Y (Response)		
	0	1	Total
CONTROL=0	61208	1334	62542
	97.87	2.13	
TREAT=1	59390	3056	62446
	95.11	4.89	
Total	120598	4390	124988



Difference Model: Fit on Train, Score Validation

Two Logistic Regressions are used to fit the Difference Model
On **TRAIN**

Fit X's (x1 x2 x3) to the **TRAIN-TREATED** and, use "OUTMODEL"

```
PROC LOGISTIC DATA=TRAIN(where=(T=1)) desc
  OUTMODEL = MODEL1; /* Lets us Score on New Data */
  MODEL Y = x1 x2 x3;
```

Fit X's (x1 x2 x3) to the **TRAIN-CONTROL** and use "OUTMODEL"

```
PROC LOGISTIC DATA=TRAIN(where=(T=0)) desc
  OUTMODEL = MODEL0;
  MODEL Y = x1 x2 x3;
```

Score **VALIDATE** prospects twice ... using MODEL1 and MODEL0

MERGE the 2 scored datasets to form **VALIDATE_SCORED**

Each **Prospect** has two probabilities P_TREATED and P_CONTROL

Subtract two probabilities to **compute Difference Score DS:**

$$DS = P_TREATED - P_CONTROL$$

DS estimates the Incremental Response Rate (IRR)

Can have different X's for MODEL1 and MODEL0

```
PROC LOGISTIC DATA = Train
  (where=(T=1))
  desc OUTMODEL = MODEL1;
  MODEL Y = x1 x2 x3;
run;
```

```
PROC LOGISTIC DATA = Train
  (where=(T=0))
  desc OUTMODEL = MODEL0;
  MODEL Y = x1 x2 x3;
run;
```

```
PROC LOGISTIC INMODEL = MODEL1;
SCORE Data = VALIDATE
Out = VALIDATE_Scored1
(rename=(P_1 = P_TREATED));
run;
PROC LOGISTIC INMODEL = MODEL0;
SCORE Data = VALIDATE
(Out = VALIDATE_Scored0
(rename=(P_1 = P_CONTROL)));
run;
```

```
DATA Validate_Scored; MERGE
  VALIDATE_Scored1 VALIDATE_Scored0;
/* No BY needed */
DS = P_TREATED - P_CONTROL;
run;
```

Importance of DS for planning Future Campaigns

We need to see if DS *actually predicts* the incremental response rate.

This is **important** because we want to use DS to **predict** the profits for FUTURE campaigns. We use DS to decide which prospects to include in the FUTURE campaign.

- An **accurate** DS **enables** the type of calculation shown below:
 - If DS = 0.02 for a group of prospects, then these prospects would produce incremental responses at a rate of 2%.
 - Assume that every response has a fixed profit amount of \$100 (this is a **simplistic assumption**... see a later slide for more discussion)
 - If a response gives us \$100 profit, then the *incremental* profit from the DS=0.02 group would give us profit equal to (Size of Group) * 2% * \$100
 - If treated response rate is 14% and incentive cost is \$10 and if there is fixed cost of \$1 for each treated prospects then for this DS = 0.02 group:

$$\text{NET INCREMENTAL PROFIT} = (\text{Size of Group}) * [2\% * \$100 - 14\% * \$10 - \$1]$$
- Future campaign: Offer treatment to DS ranks that give NET INCREMENTAL PROFIT > 0

Validating the Difference Score DS

Here is the last DATA Step.

```
DATA Validate_Scored; MERGE
VALIDATE_Scored1 VALIDATE_Scored0;
/* No BY needed */
DS = P_TREATED - P_CONTROL;
run;
```

Does DS accurately estimate the Incremental Response Rate?

Rank the TREATED prospects in Validate_Scored by DS.

Here is code:

```
PROC RANK DATA = Validate_Scored(where=(T=1))
OUT = RANKED GROUPS = 10 DESCENDING;
VAR DS; RANKS RANK_DS;
run;
PROC MEANS DATA = RANKED NOPRINT;
CLASS RANK_DS; VAR DS Y;
OUTPUT OUT = MEANOUT(DROP=_TYPE_)
MEAN = Mean_DS Mean_Y;
PROC PRINT DATA = MEANOUT;
run;
```

RANK_DS	_FREQ_	Mean_DS	Mean_Y
ALL	62446	0.027	0.049
0	6244	0.063	0.088
1	6245	0.047	0.081
2	6245	0.038	0.068
3	6244	0.031	0.066
4	6245	0.025	0.053
5	6245	0.019	0.045
6	6244	0.015	0.034
7	6245	0.012	0.026
8	6245	0.010	0.021
9	6244	0.007	0.007

%Y=1

- Mean_Y (response rate of Y=1) has large drop from RANK=0 to RANK=9 ... 0.088 to 0.007
- Mean_Y decreases steadily down the ranks
- ➔ So far, GOOD ... but have not yet validated DS

What is %Y=1 if these prospects were untreated??

Validating the Difference Score DS

```

PROC RANK DATA = Validate_Scored(where=(T=0))
OUT = RANKED GROUPS = 10 DESCENDING;
VAR DS; RANKS RANK_DS;
run;
PROC MEANS DATA = RANKED NOPRINT;
CLASS RANK_DS; VAR DS Y;
OUTPUT OUT = MEANOUT
(rename = (RANK_DS = Control_RANK_DS))
MEAN = Control_Mean_DS Control_Mean_Y;
PROC PRINT DATA = MEANOUT;
VAR
Control_RANK_DS _FREQ_ Control_Mean_DS Control_Mean_Y;
run;

```

/* CHART on RIGHT combines TREATED and CONTROL reports → */

- A flip-flop at ranks 2,3
- Fairly good agreement between DS and IRR ... but perhaps collapse to quintiles for a better agreement and work with 5 ranks

The hope is that IRR has decreasing trend and DS is close. Support FUTURE campaigns

Incremental_Rate = Treated_Resp_Rate - Control_Resp_Rate
 Compare Incremental_Rate to Treated_Mean_DS
 (Treated Mean DS ≈ Control Mean DS ... see Appendix)

Treated_RANK_DS	Treated_Mean_DS	Incremental_Rate	Treated_Resp_Rate	Control_Resp_Rate
.	0.027	0.028	0.049	0.021
0	0.063	0.053	0.088	0.035
1	0.047	0.046	0.081	0.035
2	0.038	0.035	0.068	0.033
3	0.031	0.040	0.066	0.026
4	0.025	0.031	0.053	0.022
5	0.019	0.026	0.045	0.019
6	0.015	0.016	0.034	0.018
7	0.012	0.013	0.026	0.013
8	0.010	0.014	0.021	0.007
9	0.007	0.001	0.007	0.005

Comments on Incrementality Models

Comments on Incrementality Models

NEXT

SAS EM Incremental Response Node:

- Offers predictor preparation tools using "Net Information Value" (NIV)
Customized for Incrementality Modeling ... Developed by Kim Larsen (2009).
I have not seen studies of the benefits of NIV ... I think it needs published research
Good explanation of NIV is by Lavery (2016). See Lund (2016) for an example and commentary
- Difference Model is criticized in Michel, et al (2019) p. 51 ... see comments in Appendix
- Perhaps, better method, also in SAS EM, is the Combined Model ... see Appendix for example
- A third method: Probability Decomposition Model (PDM) ... In automotive marketing, we had some success with a PDM with a cash-back offer ... see Appendix for discussion.

To Fit a useful Incrementality Model is challenging ... difficult to fit a model that predicts Incremental Responses for future campaigns ... when model is based on transactional data (purchase and service history, demographic overlays).

... what is needed is a predictor X "If only I were given an offer, then I'd buy"

Incremental Response Rate v. Incremental Profit?

In many applications the response is binary ... 0 or 1

Here, a ranking by Net Lift Model score is sufficient to decide who to target with a treatment.

But if the response is quantitative and if profit varies with the response ... then must **rethink**

e.g. New-Vehicle Sale ... each vehicle has a different profit for the Automotive Company

luxury vehicle higher profit than small sedan ... better to sell an incremental luxury vehicle

A Net Lift Model score might **incorrectly** rank prospects for targeting if profit is ignored.


This adds complexity to NET LIFT modeling ... see next slide.

Variable Profit from a Response ... How to Rank Prospects

Self-study - get the slide and work through the calculations

Example: The "MIDDLE" Net Lift Model rank has the highest Net Profit per Prospect (\$1.60) but only the second highest Incremental Response Rate (0.02)

Develop this TABLE from PILOT data. Update TABLE when new Campaigns are run.



NET LIFT MODEL Rank	Treated Model Response Rate	Profit from a Treated Response	Control Model Response Rate	Profit from a Control Response	Incremental Profit (before Costs) per Prospect	Treatment Cost per Prospect \$20	Fixed Cost per Prospect \$1	Net Profit Per Prospect	Memo: Incremental Response Rate
	(A)	(B)	(C)	(D)	(E = A*B - C*D)	(F = A*\$20)	(G)	(H = E - F - G)	(I = A - C)
HIGH	0.10	\$100	0.07	\$100	\$3.00	\$2.00	\$1.00	(\$0.00)	0.03
MIDDLE	0.20	\$150	0.18	\$130	\$6.60	\$4.00	\$1.00	\$1.60 ←	0.02
LOW	0.05	\$50	0.04	\$50	\$0.50	\$1.00	\$1.00	(\$1.50)	0.01

See Appendix for prospect level data for the table

SAS EM offers "linear regression" for quantitative response (assumes the response distribution is suitable for a regression model)

A discussion of estimating the Treatment Effect in an
Observational Study
using
Propensity Scores
... computed via a Logistic Model

An Observational Study is Described ...

“Observational study”: subjects are assigned to treatments by a non-randomized process.

Assume 2 treatments:

$T=1$ treated group

$T=0$ control group

Associated with subjects is an outcome Y that follows the treatment T

- Y might be binary ... sick/well, buy/rent
- Y might have continuous numeric values ... income 1 year after treatment

Goal: Estimate the average effect of T on Y

Called: Average Treatment Effect ... abbreviated by ATE

Automotive Marketing Example

Auto Co. offered \$10 cash-back if vehicle owner had a “deluxe” oil change at franchised dealer during a marketing campaign. Widely advertised.

Owner “self selects” by having oil change & claiming \$10 via web.

Treatment Groups:

T = 1 if received \$10 cash back

T = 0 random sample from Database with an oil change in prior period

Y: Binary Outcome:

Owner **BUYS** new vehicle from Company within 1 year after the oil change or **NOT**

X: Facts about vehicle owner's purchase and service history with Company

ATE = Increase in % buying

... Can't simply compare buy-rates between treatments: T=1 v. T=0

... T=1 were generally recent buyers & loyal servicers and T=0 were not.

We now explore methods that can measure ATE for an example like this one.

Propensity Scores

NEXT

The Propensity Score (PS) is the probability of receiving treatment given the observed X's.

PS does not involve outcome Y

Propensity Score is $PS = P(T=1 | X)$... Logistic Regression can be used to fit PS

PROC PSMATCH provides an option "**PSMODEL**" to use Logistic Regression to fit PS.

Example given later.

PSMODEL has limited functionality but PROC PSMATCH provides a method to "import" Propensity Scores which were fitted externally using other procedures.

Propensity Scores are used to help in the estimation ATE for an observational study:

There are three ways the PS can be used:

1. Matching
2. Stratification
3. Inverse Probability Weighting

These are discussed on the next slide.

PROC PSMATCH offers three ways to use PS to help estimate ATE

(1) MATCHING

PROC PSMATCH has a "MATCH METHOD" statement. Here is a high level explanation:

- Given a case in TREAT, then a case in CONTROL is found where $PS_{\text{treat}} \sim PS_{\text{control}}$
- This implies that $X_{\text{treat}} \sim X_{\text{control}}$... i.e. very similar covariate values.
- Therefore, TREATED case is "very much like" CONTROL case ... **apples to apples**
- For this case: Treatment_Effect = $Y_{\text{treat}} - Y_{\text{control}}$... like treating and not treating same case
- To remove random chance ... average over all treated cases ... $ATE = \text{AVG}(Y_{\text{treat}} - Y_{\text{control}})$

Other options for the MATCH METHOD statement in PSMATCH. See Appendix.

(2) STRATIFICATION ... TREAT and CONTROL are put into Strata (Groups) with similar PS

Less "exact" than MATCH but is feasible for large TREAT and CONTROL groups ... MATCH is limited by size of TREAT and CONTROL groups.

Regarding \$10 oil-change rebate, we used an ad hoc stratification approach ... this work was before PROC PSMATCH was introduced (2016) ... stratification is not further discussed today

(3) INVERSE PROBABILITY WEIGHTING (IPW): Weights are computed from propensity scores.

The weights adjust TREAT and CONTROL to make these groups comparable. The weights can then be incorporated into a subsequent analysis that estimates the ATE.

Requirements for an unbiased estimate of ATE

Regardless of METHOD (matching, stratification, weighting) ... for unbiased estimate ATE ... PS and X's must satisfy:

1. **Strong Ignorability Assumption (SIA)**: ... there is no omitted X that is strongly related to both treatment and outcome.
2. **Balance**: PS and X's (for PS) must have similar distributions (histograms) on T=1 and T=0 **after** applying either matching, stratification or weighting.

To evaluate **Balance**: PSMATCH has an option called **ASSESS**.

ASSESS provides statistical measures and graphics to measure **Balance**. **ASSESS** is not discussed today due to time constraints.

Concerning PS: **More important** to have **Balance** than to have good fit.

SIA has a formal statistical definition ... see Appendix for definition.

No statistical tests for **SIA** ... Modeler must apply Subject Matter Expertise.

But see <https://www.youtube.com/watch?v=tUbVkMGlqmM> and PROC CAUSALGRAPH for an approach.

SIA Violation: Example of Omitted but Strongly Related X

- The treatment T is attending type of college: S = "State U" vs I = "Ivy U"
- Outcome Y: Student's Annual Income 5 years after graduation
- Suppose X1 is GENDER and X2 is SAT Score from High School.
- Comparison of Y can be made across T's for equal values of X1-X2 ... apples to apples ?
- No, ATE estimate will be biased!
 - X3 = Parent's Income is not included. X3 is strongly related to:
 - Treatment (high tuition requires high parent income)
 - Outcome (high parent income implies "connections" leading to high paying jobs / businesses opportunities for the graduate).
- Outcome Y cannot be viewed as the result of a randomized treatment after controlling only for X1-X2 (wrong apples)

In the Appendix the mathematical definition of SIA is given and then it is shown mathematically how this example violates SIA.

NEXT

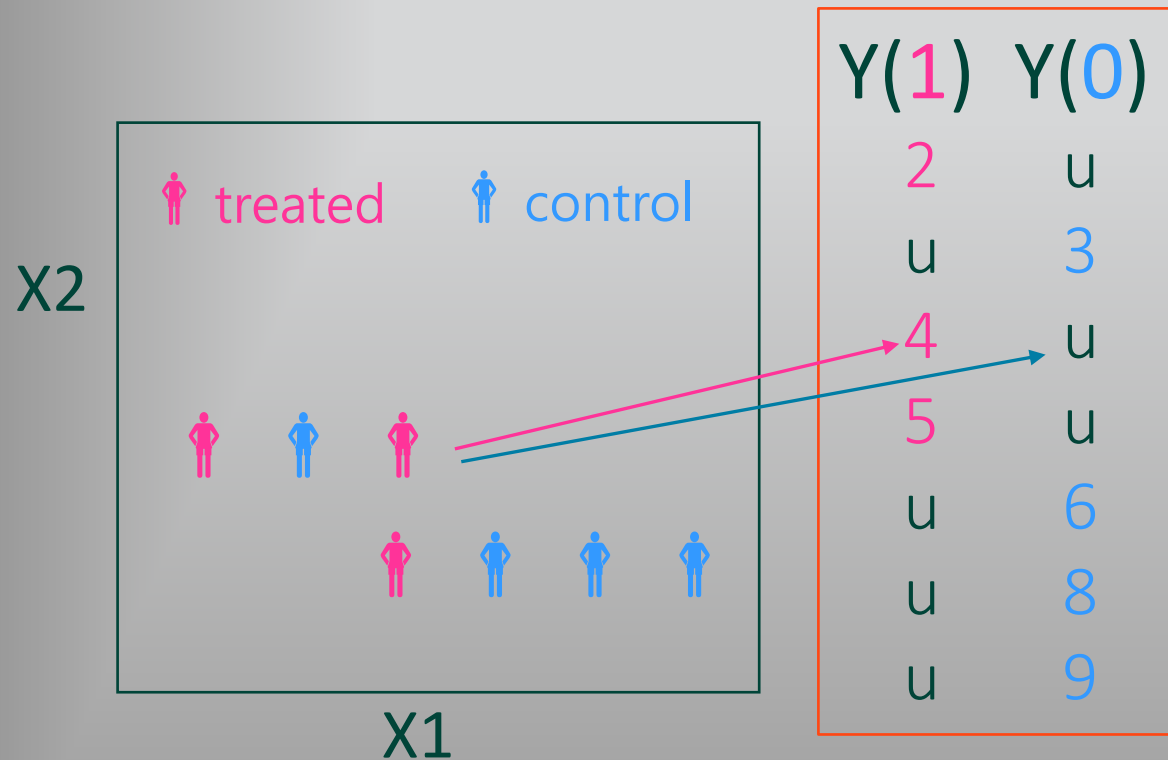
Inverse Probability Weights and ATE

Potential Outcomes

NEXT

If a prospect is treated, then it has a real treated outcome, but has a *potential* control outcome
 $Y(1)$ be the real **treated** outcome and $Y(0)$ be the potential **control** outcome

If a prospect is untreated, then it has a real control outcome, but has a *potential* treated outcome
 $Y(1)$ be the potential **treated** outcome and $Y(0)$ be the real **control** outcome



Focus on $Y(1)$... we see some "u's". They are real values ... just never to be observed.

$Y(1)$ represents the conceptual population of all the prospects being **treated**

$E[Y(1)]$ is expected or mean value of the population of **all treated**

Likewise, $E[Y(0)]$ is mean of **all control**

So, based on these hypotheticals:

$$ATE = E[Y(1)] - E[Y(0)]$$

How can $E[Y(1)]$, $E[Y(0)]$ be estimated from a sample of Y , T , and X 's ??

This formula estimates ATE ... plug in sample ... no missing values

Get a sample size n of (Y, T, X) . Fit PS model, and then Plug-in values from the sample ... no missing.

$$(1/n) * \left\{ \sum_{i=1}^n Y_i * T_i / ps(x_i) - \sum_{i=1}^n Y_i * (1 - T_i) / (1 - ps(x_i)) \right\} \text{ estimates ATE}$$

How to interpret this formula? Consider the first summation:

Consider a term $Y * T$ where $T=1$... if $T=0$, then this term $Y * T$ would be zero

- Suppose for this Y : $ps(x) = 1/3$ and so $1/ps(x) = 3$.
- Then $Y / ps(x) = 3 * Y$ accounts for 3 Y 's from population of $Y(1)$... "filling in unobserved".

A similar discussion for the second sum.

But I realize this discussion doesn't prove the formula.

Proof of the ATE formula depends on:

- (1) Strong Ignorability Assumption (SIA) is true
- (2) Propensity score is perfect ... fitted propensity score at X exactly equals $P(T=1 | X)$

There is a proof in Appendix which utilizes basic probability rules.

Proof looks somewhat "magical".

PROC PSMATCH computes PS and WEIGHTS

PROC PSMATCH fits a propensity model for T.

```

PROC PSMATCH DATA=WORK;
CLASS T X;
PSMODEL T(TREATED="1")=X;
PSWEIGHT WEIGHT=ATEWGT; /* SAS keywords */
OUTPUT OUT=OUTwork WEIGHT= ATEwgt;
/* ATEwgt is our choice of SAS variable name */

```

Another
Choice is:
ATTWGT

WORK			
Obs	Y	T	X
1	1	0	1
2	2	1	1
3	3	0	1
4	1	1	1
5	1	0	3
6	1	1	3
7	4	0	4
8	2	0	4
9	3	1	4
10	2	0	4
11	3	0	4
12	2	1	4

ADDED to OUTwork	
PS	ATEwgt (IPW)
0.5	2
0.5	2
0.5	2
0.5	2
0.5	2
0.5	2
0.333	1.5
0.333	1.5
0.333	3
0.333	1.5
0.333	1.5
0.333	3

PSMODEL produces propensity scores: **_PS_**
PS is used to create Weights called ATEwgt.

$$\text{ATEwgt} = \frac{1}{\text{_PS_} \dots \text{treated}} \\
 \frac{1}{(1 - \text{_PS_}) \dots \text{control}}$$

Terminology: ATEwgt is called "Inverse Probability Weight" and abbreviated IPW.

Now Estimate ATE ... using DATA Step and PROC MEANS

$(1/n) * \{ \sum_{i=1}^n Y_i * T_i / ps(x_i) - \sum_{i=1}^n Y_i * (1 - T_i) / (1 - ps(x_i)) \}$ estimates ATE

DATA OUTwork2; **SET** OUTwork;

YTwgt = Y*T*ATEwgt;

← Creates all the terms for the 2 sums in the ATE formula

YCwgt = Y*(1-T)*ATEwgt;

PROC MEANS DATA=OUTwork2 **MEAN**;

VAR YTwgt YCwgt; ← Average of the terms

run;

Variable	Mean
YTwgt	1.917
YCwgt	2.208

$$\text{ATE} = 1.917 - 2.208 = -0.292$$

NEXT

PROC CAUSALTRT computes PS, IPW, and ATE

```

PROC CAUSALTRT DATA= work METHOD= IPW; ←
CLASS X;
/* PROPENSITY MODEL ... doesn't require T to be included in CLASS statement */
PSMODEL T(descending)= X;
/* MODEL Y identifies Y as OUTCOME ... other METHOD's specify MODEL Y = <X>*/
MODEL Y;
run;

```

POM =
potential
outcome mean

Analysis of Causal Effect							
Parameter	Treatment Level	Estimate	Robust Std Err	Wald 95% Confidence Limits		Z	Pr > Z
POM	1	1.917	0.275	1.377	2.456	6.96	<.0001
POM	0	2.208	0.364	1.496	2.921	6.07	<.0001
ATE		-0.292	0.382	-1.041	0.458	-0.76	0.446

NEXT

PROC CAUSALTRT ... many features not discussed today

This has been a very brief introduction to PROC CAUSALTRT.

There are many other features:

A total of six METHOD's (PSMODEL & MODEL combinations) are offered by CASUALTRT.

Distributions offered for the MODEL statement:

Binary, Normal, Poisson, Gamma

Here are some readings ...

- First read: Yuan, Y., Yung, Y.-F., and Stokes, M. (2017), SAS Global Forum on PROC PSMATCH
- Then read: Lamm, M., and Yung, Y.-F. (2017), SAS Global Forum on PROC CAUSALTRT
- More advanced: Marie Davidian, notes NCSU (2007) <https://www4.stat.ncsu.edu/~davidian/double.pdf>

Stop Here if Running Out of Time

Why not use TTEST to estimate ATE instead of my DATA Step?

```
PROC TTEST DATA= OUTwork;
VAR Y; CLASS T;
WEIGHT ATEwgt;
run;
```

PROC TTEST does not compute ATE according formula given here



$$(1/n) * \sum_{i=1}^n y_i * t_i / e(x_i) - (1/n) * \sum_{i=1}^n y_i * (1-t_i) / (1 - e(x_i))$$

PROC TTEST replaces "1/n" with "Ratios"

replace with

$$1 / \sum_{i=1}^n t_i / e(x_i)$$



$$1 / \sum_{i=1}^n (1-t_i) / (1-e(x_i))$$

PROC TTEST gives an alternative formula for ATE.
Let's have an example ... See next slide

NEXT

PROC TTEST: different usage of inverse probability weights

```

DATA WORK2;
DO I = 1 to 500;
  NR = rannor(1); /* related to T */
  XBETA = NR + 0.5*rannor(2);
  PROB = exp(XBETA)/(1 + exp(XBETA));
  T = (PROB < 0.5);
  /* Y is not related to NR or NU */
  Y = rannor(3);
  NU = ranuni(4); /* not related to T */
OUTPUT;
END;
run;

PROC PSMATCH DATA= WORK2;
  CLASS T;
  PSMODEL T(TREATED="0")= NR NU;
  PSWEIGHT WEIGHT = ATEWGT;
  OUTPUT OUT= OUTwork2 WEIGHT= ATEwgt;
run;

PROC TTEST DATA= OUTwork2;
  VAR Y; CLASS T;
  WEIGHT ATEwgt;
run;

```

DATA WORK2;

DO I = 1 to 500;

NR = rannor(1); /* related to T */

XBETA = NR + 0.5*rannor(2);

PROB = exp(XBETA)/(1 + exp(XBETA));

T = (PROB < 0.5);

/* Y is not related to NR or NU */

Y = rannor(3);

NU = ranuni(4); /* not related to T */

OUTPUT;

END;

run;

PROC PSMATCH DATA= WORK2;

CLASS T;

PSMODEL T(TREATED="0")= NR NU;

PSWEIGHT WEIGHT = ATEWGT;

OUTPUT OUT= OUTwork2 **WEIGHT**= ATEwgt;

run;

PROC TTEST DATA= OUTwork2;

VAR Y; **CLASS** T;

WEIGHT ATEwgt;

run;

NEXT

We need a more complex example

PROC TTEST replaces "1/n" with "Ratios"

$$(1/n) * \sum_{i=1}^n y_i * t_i / e(x_i) - (1/n) * \sum_{i=1}^n y_i * (1-t_i) / (1 - e(x_i))$$

$$1 / \sum_{i=1}^n t_i / e(x_i)$$

$$1 / \sum_{i=1}^n (1-t_i) / (1-e(x_i))$$

1. Is the formula with "ratios" better or worse?
2. Does PROC CASUALTRT support this formula?

T	Method	N	Mean
0		224	-0.0317
1		276	0.0175
Diff (1-2)	Pooled		-0.0492

CAUSALTRT with METHOD = IPWR ... agrees with TTEST

NEXT

IPWR = inverse probability weighting method with ratio adjustment

PROC CAUSALTRT DATA= WORK2 **METHOD=IPWR;** ← New METHOD

PSMODEL T= NR NU;

MODEL Y;

run;

Analysis of Causal Effect (IPWR)		
Parameter	Treatment Level	Estimate
POM	0	-0.03174
POM	1	0.01749
ATE		-0.04923

Analysis of Causal Effect (IPW)		
Parameter	Treatment Level	Estimate
POM	0	-0.02634
POM	1	0.01927
ATE		-0.04561

← Same value as PROC TTEST

IPW →

Both ATE methods (IPW, IPWR) are consistent and approximately unbiased in medium size samples.

Method IPWR generally has **smaller variance** than IPW in repeated sampling from same population.

If no METHOD is given in the Code above, then PROC CAUSALTRT will use IPWR by default

PROC CAUSALTRT ... PSMODEL has limited functionality

PSMODEL has no variable SELECTION options ... simply fits all the X's in PSMODEL statement

But a preliminary PROC LOGISTIC (or HPLOGISTIC) can be run

- Use any feature of PROC LOGISTIC to fit model
- Save **xbeta** from this PROC LOGISTIC model
- Use **xbeta** in PROC CAUSALTRT

Must assess BALANCE of PS, NR, NU before running CAUSALTRT

```

DATA WORK2;
DO I = 1 to 500;
  NR = rannor(1);
  xb = NR + 0.5*rannor(2);
  PROB = exp(xb)/(1 + exp(xb));
  T = (PROB < 0.5);
  Y = rannor(3);
  NU = ranuni(4);
OUTPUT;
END;
run;
PROC LOGISTIC DATA = WORK2;
MODEL T = NR NU / SELECTION=FORWARD SLE=.1;
OUTPUT OUT = WORK3 XBETA = xbeta_Logistic;
run;
PROC CAUSALTRT DATA = WORK3 METHOD= IPWR;
PSMODEL T= xbeta_Logistic;
MODEL Y;
run;
  
```

Appendices

- Appendix 1a-1d: Net Lift using the Combined Model
- Appendix 2a-2b: Net Lift using Probability Decomposition Model
- Appendix 3a-3b: Conditional Independence and SIA and example of SIA violation
- Appendix 4a-4c: Proof of ATE formula with inverse probability weights
- Appendix 5a-5b: Treated, Control Validation Groups ... DS and covariate distributions
- Appendix 6: Discussion of weakness of the Difference Model
- Appendix 7: An example of PROC PSMATCH use of Matching
- Appendix 8: Predicting Profits for a Future Campaign

blund_data@mi.rr.com and blund.data@gmail.com

Send an email for a copy of slides

Appendix 1a: Combined Model (called Victor Lo's Model)

Same CAMPAIGN dataset (as in Difference Model)

But now a tweak to the VALIDATE dataset ...
needed in later steps.

```

DATA Train Validate(drop=T); SET Campaign;
if ranuni(123) < .5 then OUTPUT Train;
else do;
    T_Actual = T; /* Rename treatment variable */
    T1 = 1; /* used later to fake out PROC LOGISTIC */
    T0 = 0; /* used later to fake out PROC LOGISTIC */
    OUTPUT Validate;
end;
run;
PROC FREQ Data = Train;
Table T * Y
/ nocol nopercnt; format T T_C.;
TITLE1 "TRAIN data";
PROC FREQ Data = Validate;
Table T_Actual * Y
/ nocol nopercnt; format T_actual T_C.;
TITLE1 "VALIDATE data";
run; TITLE; run;

```

TRAIN data

Table of T by Y			
T	Y		Total
	0	1	
CONTROL	61060	1398	62458
TREAT	59490	3064	62554
Total	120550	4462	125012

VALIDATE data

Table of T_Actual by Y			
T_Actual	Y		Total
	0	1	
CONTROL	61208	1334	62542
TREAT	59390	3056	62446
Total	120598	4390	124988

Appendix 1b: Fit Model to TRAIN with X's, T, Interactions with T

* Fit combined model on TRAIN ... both treated and control prospects;

PROC LOGISTIC DATA = Train desc **OUTMODEL** = OUTMODEL;

MODEL Y = x1 x2 x3 T x1*T x2*T x3*T;

run;

Analysis of Maximum Likelihood Estimates			
Parameter	DF	Estimate	Pr > ChiSq
Intercept	1	-4.7920	<.0001
x1	1	1.7399	<.0001
x2	1	-0.1105	0.2386
x3	1	0.1580	0.0915
T	1	0.4371	0.0002
x1*T	1	0.3573	0.0036
x2*T	1	0.4400	0.0001
x3*T	1	-0.1348	0.2362

c-Stat	0.691
--------	-------

- This is Lo's Combined Model. It is fit to all of TRAIN ... whether or not Treated.
- Not focusing here on Model fitting techniques ... simply use the model with all 7 predictors.
- Do a MODEL COMPARISON test of MODEL with only x1 x2 x3 vs. FULL MODEL in order to test if Treatment has any effect ... clearly, Treatment has an effect in this example.

Appendix 1c: The Lift Chart for new Prospects who are Treated

```

PROC LOGISTIC INMODEL = OUTMODEL;
SCORE DATA = Validate(WHERE=(T_Actual=1)
RENAME=(T0=T)) OUT=T0(RENAME=(P_1 = P_1_T0));
run;
PROC LOGISTIC INMODEL = OUTMODEL;
SCORE DATA = Validate(WHERE=(T_Actual=1)
RENAME=(T1=T)) OUT=T1(RENAME=(P_1 = P_1_T1));
run;
DATA VALIDATE_DS_1; MERGE T1 T0;
DS = P_1_T1 - P_1_T0;
run;
PROC RANK DATA = VALIDATE_DS_1
OUT=VALIDATE_DS_1 GROUPS=10 DESCENDING;
VAR DS;
RANKS VALIDATE_DS_1;
run;
PROC MEANS DATA = VALIDATE_DS_1 NOPRINT;
CLASS VALIDATE_DS_1; VAR DS Y T_Actual;
OUTPUT OUT = MEANOUT_1
MEAN = DS Y T_Actual;
run;
PROC PRINT DATA = MEANOUT_1;
TITLE "VALIDATION TREATED";
run; TITLE; run;

```

VALIDATE-TREATED (where=(T_Actual=1) is scored **twice**

1. With T=0 ... compute control score
2. With T=1 ... compute treated score

Then **DS** is difference of treated and control scores

- Create deciles using **DS**.
- Look at **Y** (equals %Y=1) across the deciles.
- Good decreasing **Y** rate

In future marketing campaigns, select and treat the best DS deciles
... these are most Incremental due to T.

DS_1	_FREQ_	DS	Y
All	62446	0.027	0.049
0	6244	0.063	0.088
1	6245	0.047	0.081
2	6245	0.038	0.068
3	6244	0.031	0.066
4	6245	0.025	0.052
5	6245	0.019	0.045
6	6244	0.015	0.034
7	6245	0.012	0.026
8	6245	0.010	0.021
9	6244	0.007	0.007

Appendix 1d: The Lift Chart for new Prospects who are Treated

Can repeat the process for **VALIDATE-CONTROL**

Expectation is that there is no relationship between DS and Y down the ranks.

- The Y's are flat for ranks 0, 1, 2
- Then there is a noticeable decrease for ranks 3 to 9
- All Y's are much lower than for **VALIDATE-TREATED**
- This is, of course, a simulation. The relationship between DS and Y may have been inadvertently programmed into the simulation.

DS_0	_FREQ_	DS	Y
All	62542	0.027	0.021
0	6254	0.063	0.035
1	6254	0.047	0.035
2	6254	0.038	0.034
3	6255	0.031	0.026
4	6254	0.025	0.022
5	6254	0.019	0.019
6	6255	0.015	0.018
7	6254	0.012	0.013
8	6254	0.010	0.007
9	6254	0.007	0.005

Appendix 2a: Net Lift using Probability Decomposition Model

PROPENSITY MODEL S: = Probability to Respond ... fitted on CONTROL ... $S = P(Y=1 | T=0)$

BUYER MODEL PT: Model is fitted to Responders (the Buyers) from both TREATED and CONTROL

PT = Probability that the Responder was treated ... **PT = Prob(T=1 | Y=1)**

Let **PC** = $1 - \text{PT} = \text{Prob}(T=0 | Y=1)$

In fitting the model to predict PT and PC it is required that treated and control records are weighted to account for differences in the sampling fractions to select the treated and control groups.

For example, suppose the eligible population for a campaign was 300,000, treated group sampled at 200,000 and control at 100,000. Then weight "W", with W=1 for treated and W=2 for control is used in fitting PT.

With weighting, the mean of P_T would equal the percentage of "treatments" among the buyers from a population where the treatment had been randomly assigned to half of the population.

LIFT is defined by **Lift** = $(\text{PT} - \text{PC}) / \text{PC}$

Now combining Propensity and Lift gives the Incremental Response Rate for a prospect.

Incremental response rate = IRR = S * Lift

When $\text{PT} > 0.50$ the quantity $S * (\text{PT} - \text{PC}) / \text{PC}$ gives the prospect's expected increase in buy-rate over the base control rate.

Appendix 2b: Net Lift using Probability Decomposition Model

Example of IR Rate calculation:

Suppose for a Prospect:

If S is 2% and LIFT is 5%

Then $IRR = S * LIFT = 2\% * 5\% = 0.1\%$

(equivalently, IR Rate is 1 per 1,000)

LIFT could be negative but such a result would arise only if the Treatment was viewed negatively by the prospects.

This model is called a Probability Decomposition Model because IRR is a product which involves two probabilities S and PT.

The Model presented here is a modification of a model first developed by Jun Zhong (2009). See Zhong (2009) and Lund (2012) in History and References Slide

Appendix 3a: Conditional Independence and SIA

Equivalent Conditions for Conditional Independence:

Conditions (1), (2), and (3) are equivalent.

$$(1) \quad P(A, B \mid C) = P(A \mid C) * P(B \mid C)$$

$$(2) \quad P(A \mid B, C) = P(A \mid C)$$

$$(3) \quad P(B \mid A, C) = P(B \mid C)$$

In each case A and B are conditionally independent given C ... this is denoted by $A \perp\!\!\!\perp B \mid C$

The Strong Ignorability Assumption (SIA) is given by the three conditions below:

(a) Conditional independence of $Y(1)$ and T given X . This is denoted $Y(1) \perp\!\!\!\perp T \mid X$

(b) Conditional independence of $Y(0)$ and T given X . This is denoted $Y(0) \perp\!\!\!\perp T \mid X$

(c) $0 < P(T=1 \mid X=x) < 1$ for all x . This says that if $X=x$ appears for a treated subject, then $X=x$ also appears for some subjects in control, and vice versa.

As a result of assumption (a) of SIA, the version (1) of conditional independence says:

$$P(Y(1)=y, T=t \mid X=x) = P(Y(1)=y \mid X=x) * P(T=t \mid X=x) \dots \text{and likewise for } Y(0) \text{ for assumption (b)}$$

As a result of assumption (a) of SIA, the version (2) of conditional independence says:

$$P(Y(1)=y \mid T=t, X=x) = P(Y(1)=y \mid X=x) \dots \text{and likewise for } Y(0) \text{ for assumption (b)}$$

Appendix 3b: Example of SIA Violation in terms of mathematical definition

In this example of a causal analysis the covariate set will omit a covariate which is strongly related to both the treatment and the outcome.

The treatment has two values (with non-random selection into the treatment):

T=0: Attending State University

T=1: Attending Ivy University (NOTE: Ivy University has high tuition)

The outcome Y will be the student's annual income 5 years after graduation.

Does Ivy University improve the student's annual income 5 years after graduation?

The two covariates to support this analysis are:

X1: Gender

X2: SAT Score from High School

It is certainly true that comparisons of the outcome across the treatments could be made for subgroups of subjects with similar values of X1 and X2.

However: X3 = Parent's Income is not included. X3 is strongly related both to treatment (high tuition requires high parent income) and Outcome (high parent income leads to high paying jobs / businesses for the graduate).

So the "treatment effect" may be nothing other than measuring the effect of parent income on the outcome, and not the effect of the university. Expressed in the framework of SIA the condition (B) would fail ... see prior slide for condition (B)

$$P(Y(1)=\text{big} \mid T=1, X=(x_1, x_2)) > P(Y(1)=\text{big} \mid X=(x_1, x_2))$$

Having T=1 (Ivy U.) implies high parent income as well as high income after 5 years ... so $P(Y(1)=\text{big} \mid T=1, X=(x_1, x_2))$ is high.

But for the right side of the inequality, there are some subjects with T=1 and some with T=0. Parent Income would be lower for T=0. So the potential treated outcome would tend to be lower ... so a lower value of $P(Y(1)=\text{big} \mid X=(x_1, x_2))$.

Appendix 4a: Proof of IPW formula for ATE

THEOREM: Let (Y_i, T_i, X_i) be a sample of size n from Population of (Outcomes, Treatments, Covariates) Assuming SIA and that $e(X)$ is the perfect propensity score, then

$$\text{ATE} = E[Y(1)] - E[Y(0)] = (1/n) * E[\sum_{i=1}^n Y_i * T_i / e(X_i)] - (1/n) * E[\sum_{i=1}^n Y_i * (1 - T_i) / (1 - e(X_i))]$$

Proof of THEOREM

Consider only $E[Y(1)] = (1/n) * E[\sum_{i=1}^n Y_i * T_i / e(X_i)]$... since the proof for $E[Y(0)]$ is similar.

Use the identity: $y * t = \{ y(1) * t + y(0) * (1 - t) \} * t = y(1) * t$ to establish:

$$\sum_{i=1}^n Y_i * T_i / e(X_i) = \sum_{i=1}^n Y_i(1) * T_i / e(X_i)$$

To simplify, look at one term from the sum, where the "i" is dropped ... $Y(1) * T / e(X)$

Note that: $Y(1) * T / e(x)$ is a function of $Y(1), T, X$...

Then $E[Y(1), T, X]$ is the expected value over the multivariate distribution $P(Y(1)=y, T=t, X=x)$

$$E[Y(1) * T / e(X)] = \sum_{y(1), t, x} y(1) * t / e(x) * P(Y(1)=y, T=t, X=x)$$

$$= \sum_{y(1), t, x} y(1) / e(x) * P(Y(1)=y, T=1, X=x) + 0 \text{ for terms where } t=0$$

Here is a basic probability rule: $P(A, B, C) = P(A | B, C) * P(B, C) = P(A | B, C) * P(B | C) * P(C)$

Apply the rule to obtain: $P(Y(1)=y, T=1, X=x) = P(Y(1)=y | T=1, X=x) * P(T=1 | X=x) * P(X=x)$

Substitute: $P(T=1 | X=x) = e(x)$ and cancel $e(x)$ from numerator and denominator as shown on next slide.

Appendix 4b: Proof of IPW formula for ATE

$$\begin{aligned}
 &= \sum_{y(1), t=1, x} y(1)/e(x) * P(Y(1)=y | T=1, X=x) * e(x) * P(X=x) \\
 &= \sum_{y(1), t=1, x} y(1) * P(Y(1)=y | T=1, X=x) * P(X=x)
 \end{aligned}$$

Recall Appendix 3a where SIA is defined.

By the **conditional independence** of SIA: $P(Y(1)=y | T=1, X=x) = P(Y(1)=y | X=x)$ giving

$$= \sum_{y(1), t=1, x} y(1) * P(Y(1)=y | X=x) * P(X=x) = \sum_{y(1), t=1, x} y(1) * P(Y(1)=y, X=x)$$

Sum over X to obtain to marginal probability for Y(1)

$$= \sum_{y(1), t=1} y(1) * P(Y(1)=y)$$

Summing over T=1 removes t from the index of summation:

$$= \sum_{y(1)} y(1) * P(Y(1)=y)$$

Note that $\sum_{y(1)} y(1) * P(Y(1)=y)$ is the definition of $E[Y(1)]$

There were n terms in the original sum $(1/n) * E[\sum_{i=1}^n Y_i * T_i / e(X_i)]$

$$(1/n) * E[\sum_{i=1}^n Y_i * T_i / e(X_i)] = (1/n) * \sum_{i=1}^n E[Y_i * T_i / e(X_i)] = (1/n) * \sum_{i=1}^n E[Y_i(1) * T_i / e(X_i)]$$

Discussion above shows each term in the right hand sum equals $E[Y(1)]$

Therefore: $(1/n) * E[\sum_{i=1}^n Y_i * T_i / e(X_i)] = (1/n) * n * E[Y(1)] = E[Y(1)]$

Appendix 4c: Proof of IPW formula for ATE

Published proofs of ATE formula utilize the mathematical expectation operation. I find these challenging to understand. One such proof is given in Angrist and Pischke (2009) *Mostly Harmless Econometrics* on page 82:

First: $E[Y^*T/e(x)] = E\{ E[Y^*T/e(x) | X=x] \}$.

Taking inside expectation of the right-side:

$$E[Y^*T/e(x) | X=x]$$

$$= E[Y | T=1, X] * e(x) / e(x)$$

$$= E[Y(1) | T=1, X]$$

$$= E[Y(1) | X].$$

$$\text{Finally, } E\{ E[Y(1) | X] \} = E[Y(1)]$$

Appendix 5a: Treated, Control Validation Groups ... DS and covariate distributions

In the Validation Sample, the treated prospects were ranked by DS.
Likewise the control prospects were ranked by DS.

- Are prospects in a treated-decile *similar* to the prospects in the same control-decile in terms of X1 X2 X3 values?

YES, the Prospects are actually very similar when measured in terms of X1 X2 X3 values.

- X1, X2, X3 are random across T=0 and T=1
- Treated validation group has essentially the same DS as corresponding control validation group
- ranking by DS of treated validation is ~same as ranking of control validation by DS
- "Same" prospects in a treated rank as in the corresponding control rank.

... Apples to apples!

(Logic above does not depend on having same X's in MODEL0 and MODEL1 ... see next slide)

Appendix 5b: Treated, Control Validation Groups ... DS and covariate distributions

Suppose MODEL1 is MODEL $Y = w1$; and MODEL0 is MODEL $Y = z1$;

Each row in VALIDATE has $Y T w1 z1$

... The assignment of values to $T=1$ and $T=0$ is random

... So, $w1$ and $z1$ have very similar distributions in TREATED and CONTROL

VALIDATE is scored by MODEL1 and MODEL0

TREATED and CONTROL in VALIDATE will have MODEL1 scores and MODEL0 scores with the same distributions.

Therefore, $DS = \text{difference of scores}$ will have the same distribution in TREAT and CONTROL.

So, the ranks of TREAT by DS and the ranks of CONTROL by DS very similar DS values and very similar $w1$ and $z1$ values ... apples to apples.

So, for each rank, the difference in $\%Y=1$ in TREAT v. CONTROL is "due to" treatment.

Appendix 6: Discussion of weakness of the Difference Model

See the book by Michel, et, al. (2019) starting on page 51.

Here is one point that the authors make:

MODEL1 is the model of $P(Y=1 \mid T=1, X\text{'s})$. Let's think of P as the sum of:

= Lift (treatment effect) + X_T (selected: X1 X2, etc.) + random

MODEL0 is the model of $P(Y=1 \mid T=0, X\text{'s})$. Let's think of P as the sum of:

= X_C (selected: X1 X3 X4, etc.) + random

In NET LIFT modeling it could easily be true that X-effect and random are large v. Lift.

Then ranking by $DS = P(Y=1 \mid T=1, X\text{'s}) - P(Y=1 \mid T=0, X\text{'s})$ could simply be due to X-effects and random-effects ... $DS = X\text{-effect} + \text{random}$

Therefore, there is no relationship of DS to Incremental Response Rates due to T.

No course, the Validation step should then fail ... so, at least, with a Validation step, we are not led to think we have a predictive model.

Appendix 7a: Example of PROC PSMATCH using MATCHING

Back to the question of validating the Difference Score Model. One question was whether the prospects in a validation treated rank were comparable to the prospects in a validation control rank. PROC PSMATCH Matching can be applied to settle this question.

MATCH METHOD specifies criteria for matching:

- A. PSMODEL fits propensity scores with X1 X2 X3
- B. OPTIMAL(K=1): Selects all matches simultaneously and without replacement to minimize total absolute difference in propensity score across all matches. K=1: matches one treated to one control.
- C. CALIPER=0.60 ... allowable distance between pairs in order to allow a match. (Default is 0.25)

If PS for control do not overlap sufficiently with PS for treated, then there are problems. Not an issue for this example

With $\text{mod}(N,5)=0$ the PSMATCH ran a LONG time. For larger samples, memory limit will be exceeded.

```

DATA Sample; SET Validate_Scored;
/* Includes DS from Validation */
if mod(_N_,5)=0; /* = 20% sample */
run;
PROC PSMATCH DATA=Sample;
CLASS T;
PSMODEL T (Treated="1") = x1 x2 x3;
MATCH METHOD=optimal(k=1)
STAT=PS caliper=0.60;
OUTPUT OUT(obs=match)=OutMatch
matchid=_MatchID;
run;

```

NEXT

Appendix 7b: Split the Matched Pairs into TREAT and CONTROL datasets

```

PROC SORT DATA=OutMatch
OUT=OutMatchx;
BY _MatchID T;
run;

PROC PRINT DATA = OutMatchx(obs=10);
Var T X1 X2 X3 Y _MatchID _PS_ DS;

DATA
T_Score(rename=(Y = YT))
C_Score(rename=(Y = YC));
SET OutMatchx; BY _MatchID T;
KEEP DS Y _MatchID T;
if first._MatchID then output C_Score;
else output T_Score;
PROC PRINT DATA = T_Score(obs=5);
Var DS YT T _MatchID;
PROC PRINT DATA = C_Score(obs=5);
Var DS YC T _MatchID;
run;

```

Obs	T	x1	x2	x3	Y	_MatchID	_PS_	DS
1	0	0.9692	0.0143	0.9548	0	1	0.4902	0.0421
2	1	0.9324	0.0136	0.9843	0	1	0.4903	0.0386
3	0	0.9723	0.0061	0.9035	0	2	0.4903	0.0424
4	1	0.9582	0.0133	0.9338	0	2	0.4903	0.0411
5	0	0.9546	0.0260	0.9413	0	3	0.4904	0.0412
6	1	0.9514	0.0393	0.9706	0	3	0.4905	0.0412
7	0	0.9855	0.0272	0.9933	0	4	0.4901	0.0438
8	1	0.9254	0.0103	0.9410	0	4	0.4905	0.0381
9	0	0.9397	0.0116	0.9085	0	5	0.4905	0.0396
10	1	0.9039	0.0092	0.9655	0	5	0.4905	0.0362

CONTROL

TREATED

Obs	DS	YC	T	_Match ID
1	0.0421	0	0	1
2	0.0424	0	0	2
3	0.0412	0	0	3
4	0.0438	0	0	4
5	0.0396	0	0	5

Obs	DS	YT	T	_Match ID
1	0.0386	0	1	1
2	0.0411	0	1	2
3	0.0412	0	1	3
4	0.0381	0	1	4
5	0.0362	0	1	5

NEXT

Appendix 7c: Validate accuracy of DS using Matched Pairs

```

PROC RANK DATA = T_Score OUT = RANKED_T
GROUPS = 5 DESCENDING;
VAR DS;
RANKS RANK_DS;
PROC MEANS DATA = RANKED_T NOPRINT;
CLASS RANK_DS; VAR DS YT;
OUTPUT OUT = MEANOUT
MEAN = Mean_DS Mean_YT;
PROC PRINT DATA = MEANOUT;
PROC RANK DATA = C_Score OUT = RANKED_C
GROUPS = 5 DESCENDING;
VAR DS;
RANKS RANK_DS;
PROC MEANS DATA = RANKED_C NOPRINT;
CLASS RANK_DS; VAR DS YC;
OUTPUT OUT = MEANOUT
MEAN = Mean_DS Mean_YC;
PROC PRINT DATA = MEANOUT;
run;

```

RANK DS
Treated

RANK DS
Control

RANK_DS=0: **_MATCHID's** in **RANK_DS=0** for TREAT are the same as in **RANK_DS=0** for CONTROL

Therefore, Incremental Rate Y is "completely due" to T.
Good Accuracy ... DS = **0.055** and Incremental Y = **0.059**

RANK_DS	TYPE	FREQ	Mean_DS	Mean_YT
ALL	0	12479	0.027	0.049
0	1	2495	0.055	0.089
1	1	2496	0.034	0.068
2	1	2496	0.022	0.037
3	1	2496	0.014	0.036
4	1	2496	0.008	0.013

RANK_DS	TYPE	FREQ	Mean_DS	Mean_YC
ALL	0	12479	0.027	0.020
0	1	2495	0.055	0.031
1	1	2496	0.034	0.027
2	1	2496	0.022	0.021
3	1	2496	0.014	0.013
4	1	2496	0.008	0.006

Incremental Y
ALL = 0.029
0.059
0.040
0.016
0.023
0.007

Mostly decreasing Rate of Incremental Y down the Ranks.

Appendix 8: Predicting Profits for a Future Campaign

Here is raw data from the TABLE columns (A), (B), (C), (D)

Model Treated Responses (0 / 1)	Profit from a Treated Response	Model Control Responses (0 / 1)	Profit from a Control Response	
1	\$100	1	\$100	
1	\$110	1	\$105	
1	\$120	1	\$110	
1	\$130	1	\$120	
1	\$150	1	\$130	
1	\$150	1	\$140	
1	\$170	1	\$150	
1	\$180	1	\$155	
1	\$190	1	\$160	
1	\$200			
40 0's	40 \$0's	41 0's	41 \$0's	
Treated Model Response Rate	Profit from a Treated Response	Control Model Response Rate	Profit from a Control Response	Incremental Profit (before Costs) per Prospect
(A)	(B)	(C)	(D)	$E=(A*B - C*D)$
0.20	\$150	0.18	\$130	\$6.60