SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

USERS PROGRAM

Presenter

Colleen McGahan, Biostatistical Lead, Cancer Surveillance & Outcomes, BC Cancer Agency

Colleen has been a Biostatistician for 20 years and started using SAS when she moved from the UK to Vancouver, Canada over 15 years ago. Colleen currently leads a team of Biostatisticians at the British Columbia Cancer Agency in Vancouver providing a statistical consulting service to researchers and policy makers across the agency. Colleen enjoys her involvement in the SAS user community and has been the President of the Vancouver SAS User Group for over 10 years.

Use of a Population Average Model to Investigate the Success of a Customer Retention Strategy

About the Presentation

- Assumes knowledge of logistic regression
- Aimed at those seeking to learn more about mixed effects or hierarchical modelling
- Focus is on a binary outcome
- 2 methods:
  - Population Average (PA) using generalized estimating equations (GEE)
  - Subject or cluster-specific (CS) using random effects model

Objective of Presentation

- To give a brief description of each method.
- Provide guidance and understanding as to when each model is appropriate to use.
- Examine the success of a new customer retention strategy within the breast screening environment (the business case).
- To lead you through the decision process so you can understand why the PA model is the most appropriate model in this scenario.

Objective of Presentation

- Conclude with the implementation of the GEE model using the GENMOD procedure

Use of Population Average Model to Investigate the Success of a Customer Retention Strategy

The Business Case

Natural News

The world's top news source on natural health

ARTICLES BLOGS LABS SCIENCE REFERENCE REPORTS VIDEOS RADIO INFOGRAPHICS MUSIC

CNN commits credibility suicide: Runs HATE

Breast cancer widespread: subjected to and radiati

Monday, March 30, 2015 by: Dav

Tags: mammogram, breast cancer, over

THE GLOBE AND MAIL

SUBSCRIBE NOW

LOG IN

thestar.com

Breast cancer screening guide

Routine mammograms, self-exams and MRIs not needed f

CBC News Posted: Nov 21, 2011 12:12 PM ET | Last Updated: Nov 21

HEALTH

American Ca

By DENISE GRADY OCT. 20, 2015

The Ame women to

News · World

New U.S. breast cancer screening guidelines cause confusion among medical groups

The American Cancer Society released new guidelines for mammograms calling for women to start screening at age 45 versus 40 if they are considered low-risk.

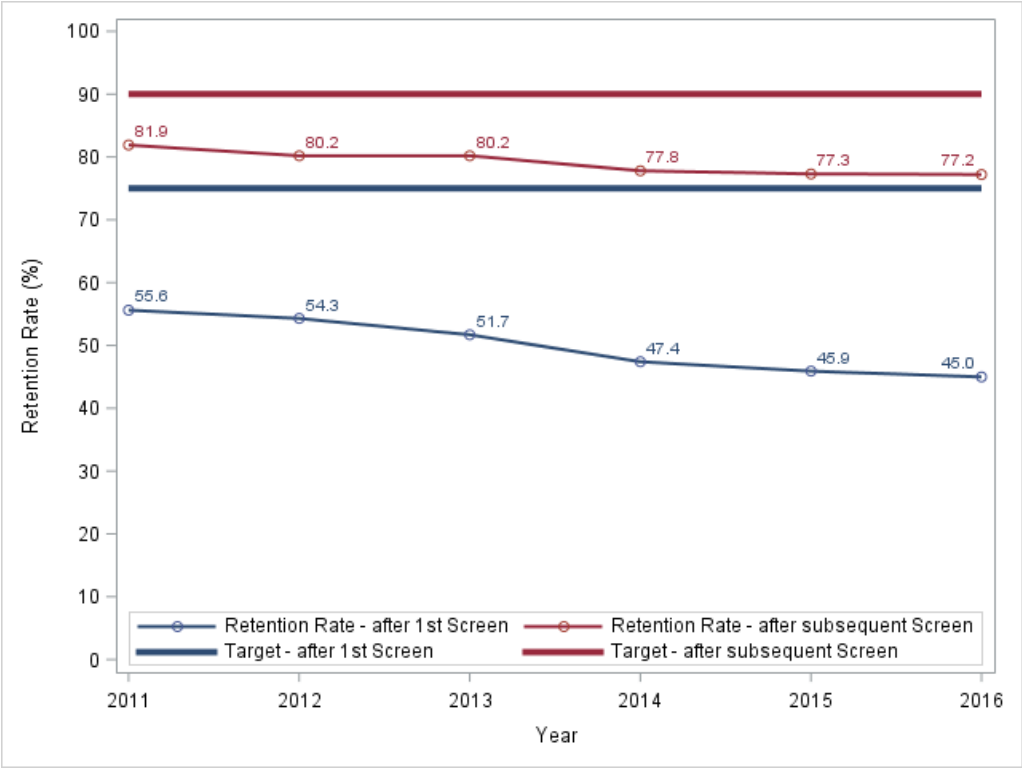
Background

Patients need to participate initially

AND

return for ongoing screening (retention)

The Problem – Retention Rates Declining



Screening Mammography Program (SMP)

British Columbia (BC) Canada

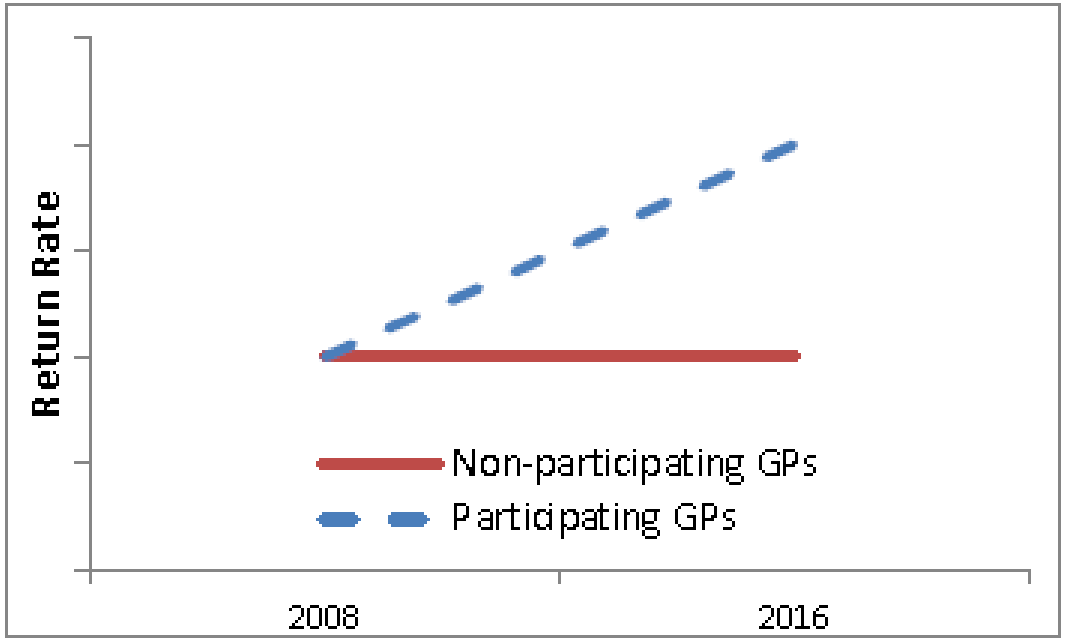
Mitigation Efforts – Investigative Study

- SMP invited all GPs practicing in one HSDA as of 1st Jan 2016 to participate in study
- Participating GP: sent invite letter (intervention)
- Non-participating GP: standard invite letter (control)
- Overdue by 24-96 months

Mitigation Efforts – Investigative Study

- Inherent biases in using GPs who chose not to participate as the control.
- Use retention rates of the same GPs but from an earlier year and use the earlier year as a baseline.
- Return rates of participating and non-participating GPs in 2016 were compared to their return rates in 2008.

Hypothesis


Return Rate
Non-participating GPs
Participating GPs
2008
2016

Characteristics of the Data Structure

- 2 levels of information – GPs and women within GP. Since women are nested within GP, each GP can be thought of as a ‘cluster’.
- One observation per woman.
- Each observation is unique to a specific cluster. Since there is only one observation per woman, a woman can only be under one GP.
- The data is unbalanced. A different number of women are within a GP.

Characteristics of the Data Structure

- The number of observations within a cluster varies quite considerably. The median number of women under one GP is 30 and ranges from 4 to 176.
- There are a large number of clusters. A total of 340 GPs; 205 participated, 135 did not.
- It is reasonable to assume data are missing completely at random (MCAR) rather than missing at random (MAR).
- In each cluster the variable of interest takes only one value.

Variables

Outcome '*returned*':

- Return to screening within 6 months of the invite letter [0=No, 1=Yes]
- Measured on each woman (level-1)

Intervention of interest '*participate*':

- GP sent invite letters [0=No, 1=Yes]
- Cluster-varying variable (level-2)

Variables

Level-1 or women-invariant variables:

- *lapse_gp:* Amount of time elapsed since a woman's last screen [24 to 36 months, 37-96 months]
- *total_screens:* Number of previous mammography screens a woman has had [1, 2, 3, 4, 5, 6, 7, 8+]
- *age_gp:* Woman's age at her last screen grouped [49-53, 54-57, 58-70]
- *period:* Period at which point the woman had not returned for screening [2008 or 2016]

Use of Population Average Model to Investigate the Success of a Customer Retention Strategy

Considerations for Model Selection

Consider 2 methods

- Focus: Binary outcome
- Method 1: Population average' (PA) using Generalized Estimating Equations (GEE) by Liang and Zeger,
- Method 2: Subject or cluster-specific (CS) method using random effects model

Comparison of the 2 methods

- Both address the problem of correlated observations but incorporate it into the model differently.
- Regression coefficients or odds ratios obtained from the two approaches are numerically different.
- Interpretation is therefore different.
- GEE models the average response effect **across all clusters** (i.e. marginal expectations of the outcome)
- For a CS model, the response effect is specific **for a given cluster**

GEE

- Uses a quasi-likelihood approach
- Separately models:
  - mean response across all clusters (the primary interest)
  - within-cluster association (a nuisance taken into account for valid inference)
- A cluster effect is not explicitly included in the model
- Observations within the same cluster are allowed to have different variances and nonzero covariances

GEE

- Within-cluster correlation is specified through a working correlation
- Assumes the observations in different clusters are independent
- It can handle unbalanced data
- Assumes data are missing completely at random (MCAR)
- The behavior of GEEs is asymptotic to the number of clusters so the more clusters you have, the better.

GEE

- No formal cut-off for what quantifies as ‘too few’ clusters.
- Some research puts the threshold of concern around 40 clusters
- If clusters have different numbers of observations then as many as 100 clusters can still be problematic
- Computational complexity increases with the size of the largest cluster.
- Standard statistical software does not easily implement more than 2 clustering levels.

GEE

- Involves fewer assumptions than the CS model
- Does not require knowledge of the ways in which individuals are correlated within a GP nor how GPs vary in order to provide consistent regression parameter estimates.
- Use if we want to estimate the average effect over the entire sample rather than estimate the effect for a particular cluster.

GEE

- It cannot make inferences about:
  - the reasons for the variation at the cluster level
  - cluster-specific effects

GEE

Relating this to our business case, the question a GEE could answer is:

How does the probability of a woman returning to screening mammography change if she received a GP invite letter compared to receiving the SMP standard letter?

Cluster-specific (CS) model

- Estimates the variance among cluster means and among observations within a cluster.
- It assumes the within-cluster correlation has a specific distribution and attempts to model it.
- It treats the cluster effect as a random variable
- Each cluster has a different intercept estimated from this random effect.

Cluster-specific (CS) model

- The change in absolute risk for different levels of covariates from one cluster to another will depend on the baseline rate for the cluster.
- Accuracy of estimates is affected by:
  - The number of groups
  - Number of level-one units
  - Amount of dependency between the observations

Cluster-specific (CS) model

- As with GEE it is better to have more clusters than more observations within a cluster.
- As few as 10 clusters, with 5 observations within each cluster is on the verge of unacceptable.
- More than 50 clusters is ideal.
- More clusters may be needed for convergence if the model is more complex or data is unbalanced.

Cluster-specific (CS) model

- Assumes data are missing at random (MAR) i.e. dependent on the covariates and observed responses.
- Can be susceptible to biases from sensitive and difficult-to-verify assumptions about the random effects distribution.
- Can be computationally intensive, especially as the number of random effects to be estimated increases beyond one or two.

Cluster-specific (CS) model

- There is no concept of ‘average’ across a covariate level
- It can:
  - estimate the cluster effect adjusting for covariates of interest.
  - evaluate and compare the performance of the clusters with respect to the outcome.
  - compare the difference between two clusters.

Cluster-specific (CS) model

Relating this to our business case, the questions a CS model could answer are:

How does the probability of a woman returning to screening mammography change, if she received an invite letter from a specific GP compared to receipt of the standard SMP letter instead?

How does the change in return rates differ if a woman received an invite letter from one GP compared to a woman receiving the invite from another GP?

Use of Population Average Model to Investigate the Success of a Customer Retention Strategy

Model Selection for Business Case

Model Selection for Business Case

- SMP wanted to evaluate if customer retention would improve using the GP invite letter compared to the standard SMP invite letter
- It was not of interest to compare retention within specific GPs or between GPs, but rather examine the overall effect.
- This is the most important aspect that leads to the decision a PA model is more appropriate than a CS model.

Working Correlation Structure Selection

- Estimates from the model are meant to be robust even if the incorrect correlation structure is specified.
- However, it is good scientific practice to carefully select the correlation structure of your data
- Can have substantive loss of efficiency in the estimation of the regression parameters if the working correlation is misspecified with small sample sizes or strong correlations

Working Correlation Structure Selection

- Independent (IND): assumes zero correlation within the same cluster
- Autoregressive (AR): correlation between consecutive measurements on a cluster to decrease with increased separation in time or space
- Unstructured (UN): assumes unique correlation
- Exchangeable (EXCH): assumes the same correlation

Model Implementation

GEE model using an exchangeable correlation structure

Model Implementation

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Implementation

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Implementation

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Implementation

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Implementation

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3 WALD;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Implementation

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Implementation

woman_ID	gp_ID
1	1
2	1
3	1
4	2
5	2
6	2

woman_ID	gp_ID
1	1
2	1
3	1
1	2
2	2
3	2

Unique observation ID:
REPEATED SUBJECT=gp_id

Unique observation ID within cluster:
REPEATED SUBJECT= woman_id(gp_id)
REPEATED SUBJECT= woman_id*gp_id

Model Output

The GENMOD Procedure

Model Information	
Data Set	WORK.STUDY5
Distribution	Binomial
Link Function	Logit
Dependent Variable	returned

Number of Observations Read	11660
Number of Observations Used	11660
Number of Events	2173
Number of Trials	11660

Model Output

Class Level Information		
Class	Levels	Values
gp_id	340	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 ...
total_screens	8	1 2 3 4 5 6 7 8
age_group	3	49-53 54-57 58-70
lapse_group	2	25-36 mo 37-96 mo
period	2	2008 2016
participate	2	0 1

Model Output

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
total_screens	7	153.32	<.0001
age_group	2	25.81	<.0001
lapse_group	1	105.63	<.0001
period	1	1.49	0.2216
participate	1	38.22	<.0001
period*participate	1	10.95	0.0009

Model Output

GEE Model Information	
Correlation Structure	Exchangeable
Subject Effect	gp_id (340 levels)
Number of Clusters	340
Correlation Matrix Dimension	176
Maximum Cluster Size	176
Minimum Cluster Size	4

Exchangeable Working Correlation	
Correlation	0.0071330977

REPEATED SUBJECT=gp_id / TYPE=UN CORRW

Model Output

GEE Fit Criteria	
QIC	10016.4047
QICu	10011.6860

Model Output

Analysis Of GEE Parameter Estimates								
Empirical Standard Error Estimates								
Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept			-1.3774	0.1278	-1.6278	-1.1270	-10.78	<.0001
total_screens	1		-2.0872	0.1183	-2.3191	-1.8553	-17.64	<.0001
total_screens	2		-1.0223	0.0901	-1.1989	-0.8457	-11.35	<.0001
total_screens	3		-0.7709	0.0876	-0.9426	-0.5991	-8.80	<.0001
total_screens	4		-0.6059	0.0946	-0.7912	-0.4205	-6.41	<.0001
total_screens	5		-0.3762	0.1029	-0.5780	-0.1745	-3.66	0.0003
total_screens	6		-0.2877	0.1008	-0.4853	-0.0901	-2.85	0.0043

Model Implementation: Model-Based Std Errors

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
RUN;
```

Model Output: Model-Based Std Errors

Analysis Of GEE Parameter Estimates								
Model-Based Standard Error Estimates								
Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept			-1.3774	0.1081	-1.5893	-1.1655	-12.74	<.0001
total_screens	1		-2.0872	0.1128	-2.3083	-1.8661	-18.50	<.0001
total_screens	2		-1.0223	0.0918	-1.2023	-0.8424	-11.14	<.0001
total_screens	3		-0.7709	0.0913	-0.9498	-0.5919	-8.44	<.0001
total_screens	4		-0.6059	0.0939	-0.7899	-0.4219	-6.45	<.0001
total_screens	5		-0.3762	0.0981	-0.5685	-0.1840	-3.84	0.0001

			Empirical Standard Error Estimates				Model-Based Standard Error Estimates			
Parameter			Estimate	Std Error	p-value		Estimate	Std Error	p-value	% Diff
Intercept			-1.3774	0.1278	<.0001		-1.3774	0.1081	<.0001	-15
total_screens	1		-2.0872	0.1183	<.0001		-2.0872	0.1128	<.0001	-5
total_screens	2		-1.0223	0.0901	<.0001		-1.0223	0.0918	<.0001	2
total_screens	3		-0.7709	0.0876	<.0001		-0.7709	0.0913	<.0001	4
total_screens	4		-0.6059	0.0946	<.0001		-0.6059	0.0939	<.0001	-1
total_screens	5		-0.3762	0.1029	0.0003		-0.3762	0.0981	0.0001	-5
total_screens	6		-0.2877	0.1008	0.0043		-0.2877	0.1041	0.0057	3
total_screens	7		-0.2892	0.1177	0.0140		-0.2892	0.1154	0.0122	-2
total_screens	8		0	0	.		0	0	.	

			Empirical Standard Error Estimates				Model-Based Standard Error Estimates			
Parameter			Estimate	Std Error	p-value		Estimate	Std Error	p-value	% Diff
age_group	49-53		0.3395	0.0684	<.0001		0.3395	0.0686	<.0001	0
age_group	54-57		0.1934	0.0605	0.0014		0.1934	0.0623	0.0019	3
age_group	58-70		0	0	.		0	0	.	
lapse_group	25-36 mo		0.8136	0.0538	<.0001		0.8136	0.0516	<.0001	-4
lapse_group	37-96 mo		0	0	.		0	0	.	
period	2016		-0.1839	0.1371	0.1798		-0.1839	0.0972	0.0585	-29
period	2008		0	0	.		0	0	.	
participate	1		0.1942	0.1195	0.1041		0.1942	0.0994	0.0508	-17
participate	0		0	0	.		0	0	.	

			Empirical Standard Error Estimates				Model-Based Standard Error Estimates			
Parameter			Estimate	Std Error	p-value		Estimate	Std Error	p-value	% Diff
period*participate	2016	1	0.5528	0.1543	0.0003		0.5528	0.1166	<.0001	-24
period*participate	2016	0	0	0	.		0	0	.	
period*participate	2008	1	0	0	.		0	0	.	
period*participate	2008	0	0	0	.		0	0	.	

```
NOTE: Class levels for some variables were not printed due to excessive size.
NOTE: PROC GENMOD is modeling the probability that returned='1'.
NOTE: Algorithm converged.
WARNING: The number of response pairs for estimating correlation is less than or equal to the
         number of regression parameters. A simpler correlation model might be more
         appropriate.
NOTE: The working correlation has been ridged with a maximum value of 18.447931549 to avoid
      singularity.
NOTE: The working correlation has been ridged with a maximum value of 13.396272134 to avoid
      singularity.
NOTE: The working correlation has been ridged with a maximum value of 4.0911454974 to avoid
      singularity.
ERROR: Error in computing the variance function.
ERROR: Error in parameter estimate covariance computation.
ERROR: Error in estimation routine.
NOTE: The scale parameter was held fixed.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE GENMOD used (Total process time):
      real time           1.95 seconds
      cpu time            1.70 seconds
```

Model Implementation: Comparisons/Odds Ratios

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
    LSMEANS period*participate / CL DIFF=ALL ODDSRATIO;
RUN;
```

Model Implementation: Comparisons/Odds Ratios

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
    LSMEANS period*participate / CL DIFF ODDSRATIO;
RUN;
```

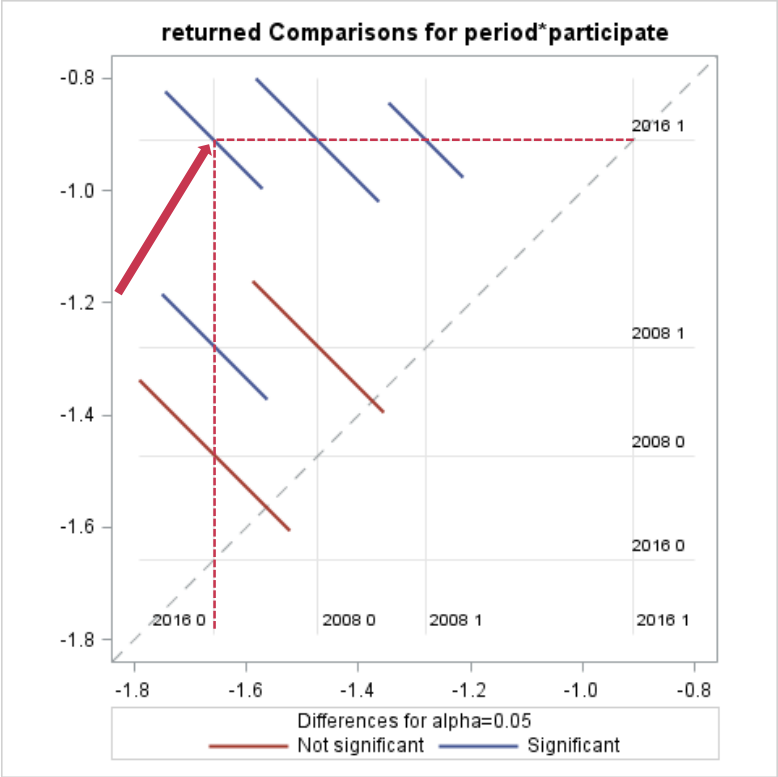
Model Implementation: Comparisons/Odds Ratios

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
    LSMEANS period*participate / CL DIFF ODDSRATIO;
RUN;
```

Model Implementation: Comparisons/Odds Ratios

Differences of period*participate Least Squares Means													
period	ITT_participate	_period	ITT_participate	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
2016	1	2016	0	0.7470	0.08849	8.44	<.0001	0.05	0.5735	0.9204	2.111	1.775	2.510
2016	1	2008	1	0.3689	0.06809	5.42	<.0001	0.05	0.2355	0.5024	1.446	1.266	1.653
2016	1	2008	0	0.5631	0.1118	5.04	<.0001	0.05	0.3441	0.7822	1.756	1.411	2.186
2016	0	2008	1	-0.3780	0.09607	-3.94	<.0001	0.05	-0.5663	-0.1898	0.685	0.568	0.827
2016	0	2008	0	-0.1839	0.1371	-1.34	0.1798	0.05	-0.4525	0.08480	0.832	0.636	1.088
2008	1	2008	0	0.1942	0.1195	1.63	0.1041	0.05	-0.03998	0.4284	1.214	0.961	1.535

Model Implementation: Comparisons Plot



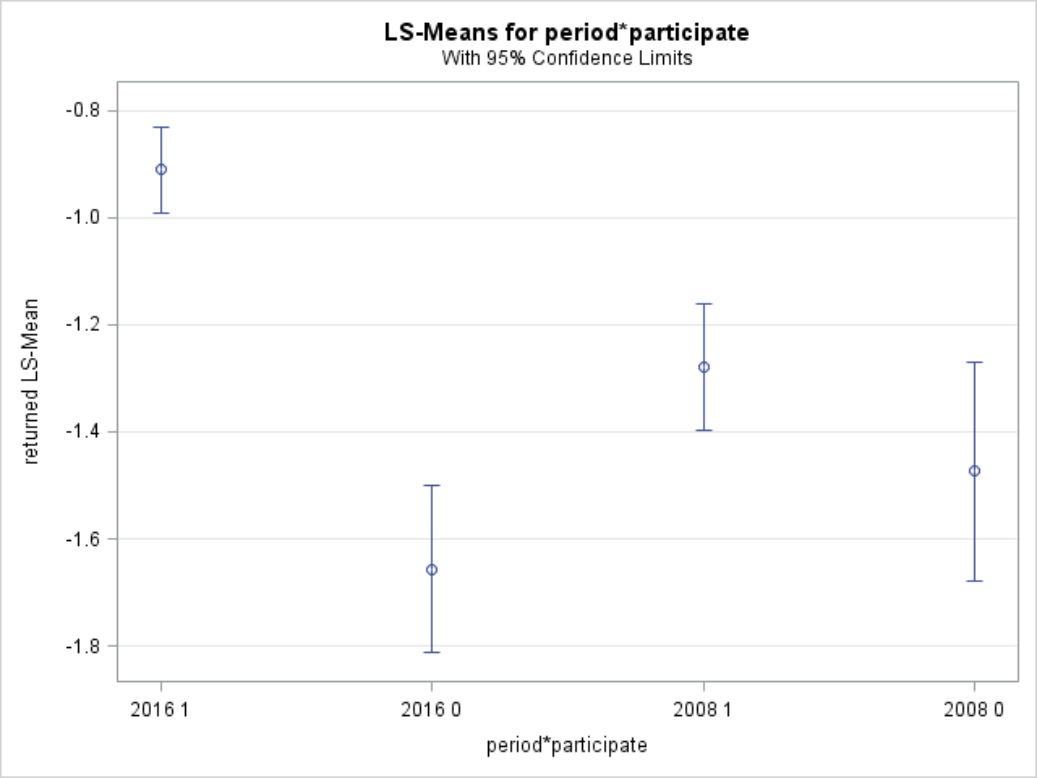
Model Implementation: Predicted Probabilities

```
ODS GRAPHICS ON;
PROC GENMOD DATA=study DESCEND;
    CLASS gp_id total_screens age_group lapse_group period(REF="2008")
          participate(REF="0")
          / PARAM=GLM;
    MODEL returned=total_screens age_gp lapse_gp period participate
                   period*participate
          / DIST=BIN LINK=LOGIT TYPE3;
    REPEATED SUBJECT=gp_id / TYPE=EXCH MODELSE;
    LSMEANS period*participate / CL DIFF ODDSRATIO;
    LSMEANS period*participate / CL ILINK;
RUN;
```

Model Implementation: Predicted Probabilities

period*participate Least Squares Means												
period	ITT_participate	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Mean	Standard Error of Mean	Lower Mean	Upper Mean
2016	1	-0.9098	0.04112	-22.12	<.0001	0.05	-0.9904	-0.8292	0.2870	0.008415	0.2708	0.3038
2016	0	-1.6568	0.07937	-20.87	<.0001	0.05	-1.8123	-1.5012	0.1602	0.01068	0.1404	0.1822
2008	1	-1.2787	0.05977	-21.39	<.0001	0.05	-1.3959	-1.1616	0.2178	0.01018	0.1985	0.2384
2008	0	-1.4729	0.1042	-14.13	<.0001	0.05	-1.6771	-1.2686	0.1865	0.01581	0.1575	0.2195

Model Implementation: Predicted Probabilities



Model Implementation: Predicted Probabilities

```
LSMEANS period*participate / CL DIFF ODDSRATIO;
LSMEANS period*participate / CL ILINK;
```

```
LSMEANS period*participate / CL DIFF ODDSRATIO ILINK;
```

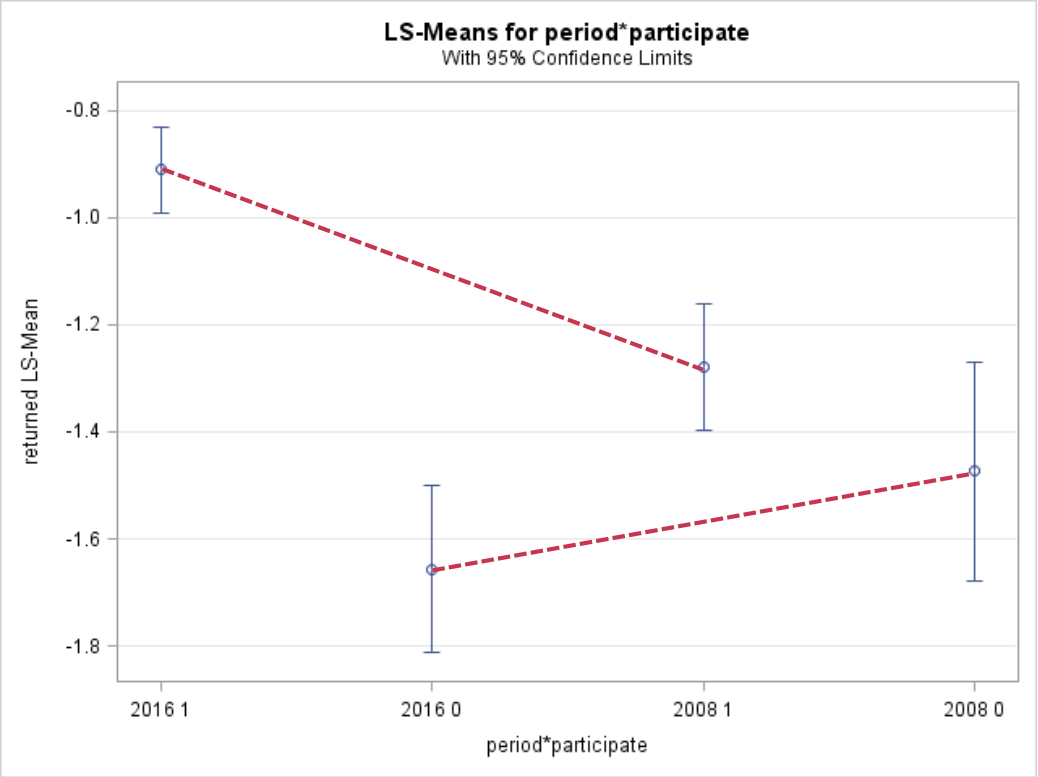
Use of Population Average Model to Investigate the Success of a Customer Retention Strategy

Interpreting the Results

Interpreting the Results

			Empirical Standard Error Estimates		
Parameter			Estimate	Std Error	p-value
period*participate	2016	1	0.5528	0.1543	0.0003
period*participate	2016	0	0	0	.
period*participate	2008	1	0	0	.
period*participate	2008	0	0	0	.

Interpreting the Results



Interpreting the Results

Differences of period*participate Least Squares Means													
period	ITT_participate	_period	ITT_participate	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
2016	1	2016	0	0.7470	0.08849	8.44	<.0001	0.05	0.5735	0.9204	2.111	1.775	2.510
2016	1	2008	1	0.3689	0.06809	5.42	<.0001	0.05	0.2355	0.5024	1.446	1.266	1.653
2016	1	2008	0	0.5631	0.1118	5.04	<.0001	0.05	0.3441	0.7822	1.756	1.411	2.186
2016	0	2008	1	-0.3780	0.09607	-3.94	<.0001	0.05	-0.5663	-0.1898	0.685	0.568	0.827
2016	0	2008	0	-0.1839	0.1371	-1.34	0.1798	0.05	-0.4525	0.08480	0.832	0.636	1.088
2008	1	2008	0	0.1942	0.1195	1.63	0.1041	0.05	-0.03998	0.4284	1.214	0.961	1.535

Interpreting the Results

period*participate Least Squares Means												
period	ITT_participate	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Mean	Standard Error of Mean	Lower Mean	Upper Mean
2016	1	-0.9098	0.04112	-22.12	<.0001	0.05	-0.9904	-0.8292	0.2870	0.008415	0.2708	0.3038
2016	0	-1.6568	0.07937	-20.87	<.0001	0.05	-1.8123	-1.5012	0.1602	0.01068	0.1404	0.1822
2008	1	-1.2787	0.05977	-21.39	<.0001	0.05	-1.3959	-1.1616	0.2178	0.01018	0.1985	0.2384
2008	0	-1.4729	0.1042	-14.13	<.0001	0.05	-1.6771	-1.2686	0.1865	0.01581	0.1575	0.2195

Conclusion

The strategy of utilizing the GP to send invite letters increased customer retention

Reflection

- Concept applies to all industries
- Improper analysis of correlated data can lead to erroneous statistical inference
- CS model parameter interpretation is cluster specific
- GEE is population average
- Understand your study, the data, and the assumptions and limitations of the methods
- Use a common sense approach

Don’t Forget to Provide Feedback!

1. Go to the Agenda icon in the conference app.
2. Find this session title and select it.
3. On the sessions page, scroll down to Surveys and select the name of the survey.
4. Complete the survey and click Finish.

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL