ANALYTICS CRM
EXCHANGE

Testing Approaches in Marketing

Design your test to ensure clarity of results

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Introduction

© 2012 Merkle Inc. All Rights Reserved | Confidential

Agenda

- Intro to Testing Methodologies
  - Approach Options
  - Focus on DOE
    - Full Factorial
    - Fractional Factorial
    - D – Optimal
  - Summary
    - Pros and Cons of Each method
    - Determining the appropriate method
- Testing Pitfalls
- Case Studies

© 2012 Merkle Inc. All Rights Reserved | Confidential

Introduction to Testing Methodologies

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Testing Methodologies: Champion/Challenger

ANALYTICS CRM EXCHANGE

Champion / Challenger

- Begin with a control strategy and test a strategy that differs in many different factors.
- Compare the two strategies against one another.

Control Cell

- Control Audience Criteria
- Control Offer
- Control Creative

→

Test Cell

- Test Audience Criteria
- Test Offer
- Test Creative

- Compare the change in Primary KPI from pre to post period of the control cell vs. the test cell and test for significance.
- The difference in change, if any, can be attributed to the test.

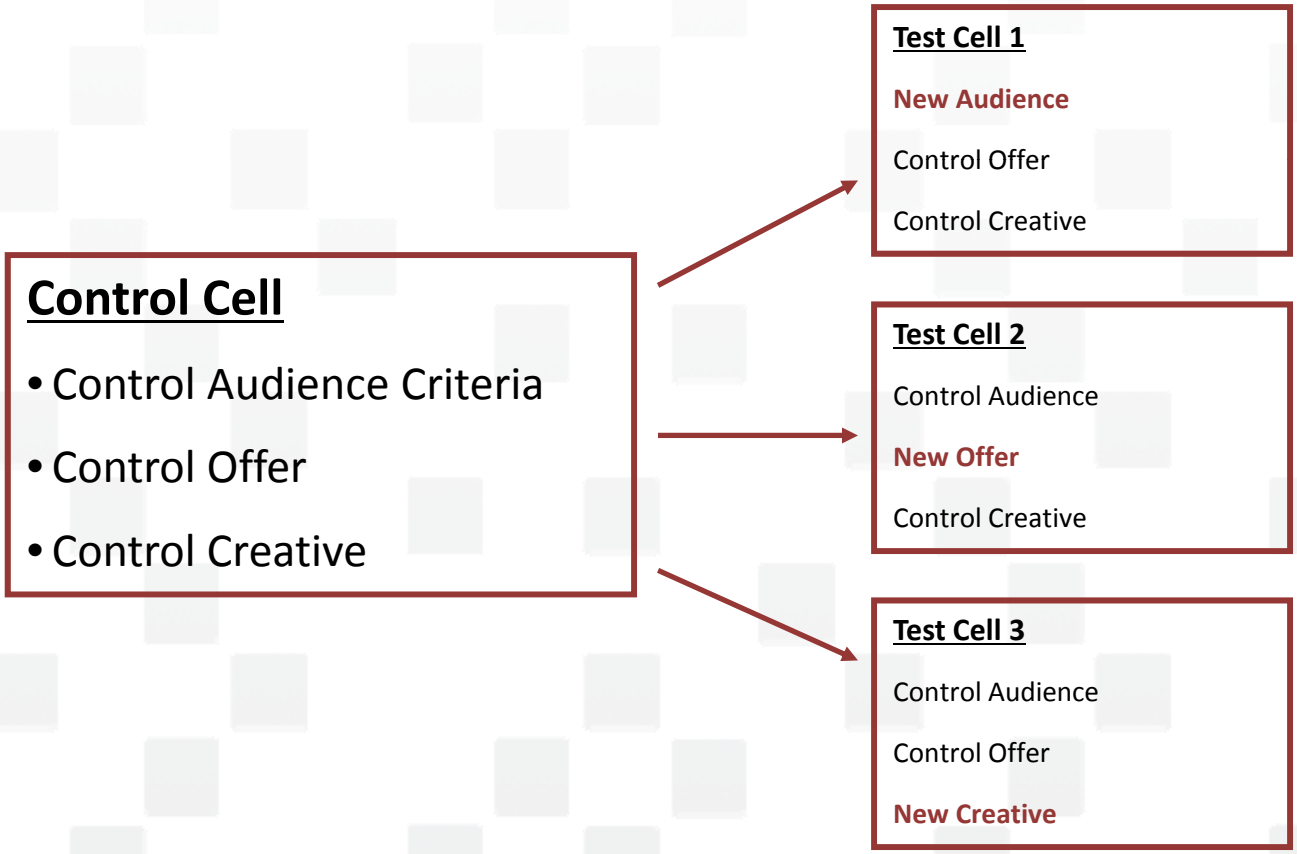
MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Testing Methodologies - OFAT

One Factor At A Time - "OFAT"

- Begin with a control strategy and multiple test cells which differ from the control in one factor only.

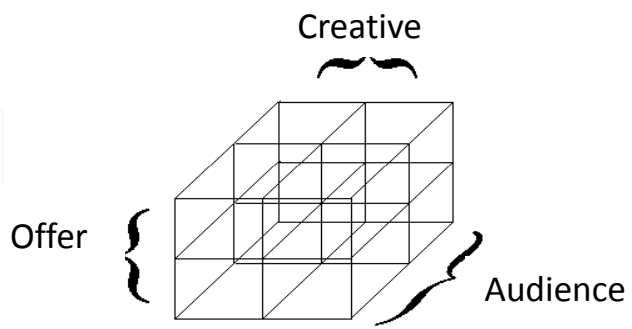


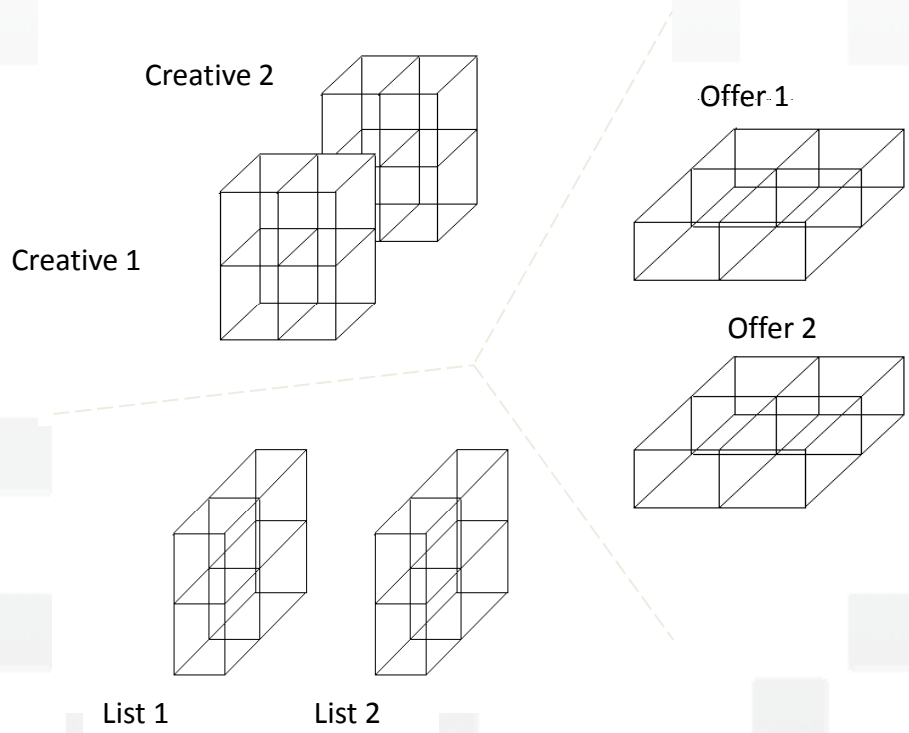
© 2012 Merkle Inc. All Rights Reserved | Confidential

Testing Methodologies – D.O.E.

D.O.E.

- Test multiple factors jointly in a structured manner. Individual test cells are combined to create larger cells which differ only based on one factor. Each matrix can be split multiple times to analyze several factors.





© 2012 Merkle Inc. All Rights Reserved | Confidential

DOE Methodologies

ANALYTICS CRM
EXCHANGE

MERKLE

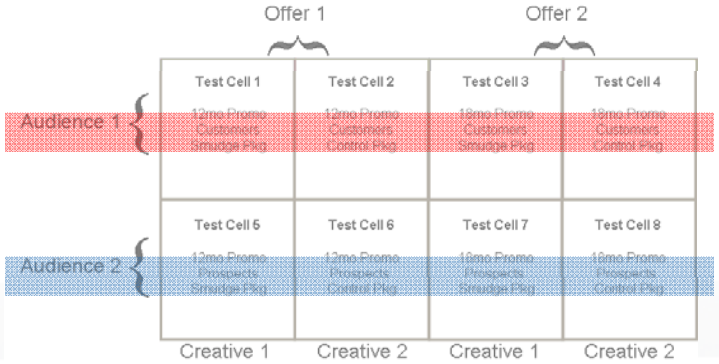
© 2012 Merkle Inc. All Rights Reserved | Confidential

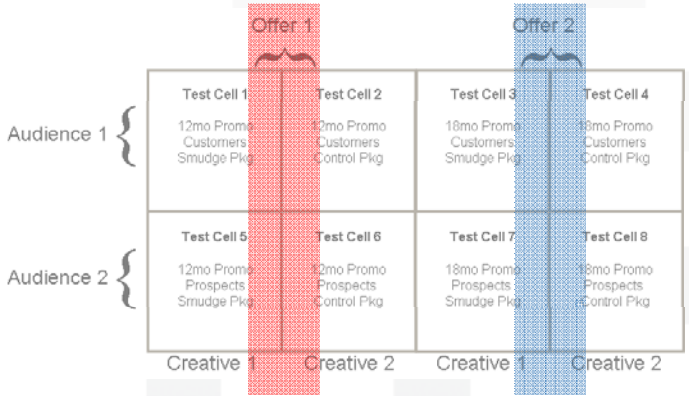
Full Factorial D.O.E.

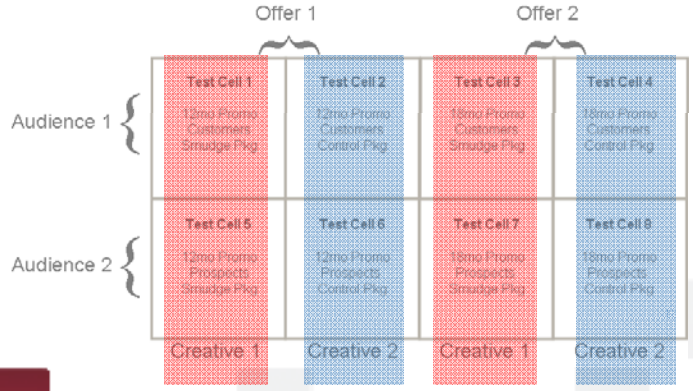
A full factorial design allows for measuring all of the factors as well as every interaction.
Example of 2^3 design – three factors (Offer, Audience, Creative) each with two levels

	Offer 1		Offer 2	
Audience 1	**Test Cell 1** 12mo Promo Customers Smudge Pkg.	**Test Cell 2** 12mo Promo Customers Control Pkg.	**Test Cell 3** 18mo Promo Customers Smudge Pkg.	**Test Cell 4** 18mo Promo Customers Control Pkg.
Audience 2	**Test Cell 5** 12mo Promo Prospects Smudge Pkg.	**Test Cell 6** 12mo Promo Prospects Control Pkg.	**Test Cell 7** 18mo Promo Prospects Smudge Pkg.	**Test Cell 8** 18mo Promo Prospects Control Pkg.
	Creative 1	Creative 2	Creative 1	Creative 2

© 2012 Merkle Inc. All Rights Reserved | Confidential







Shared Volume Across Test Cells

- A typical test might require 200,000 per cell
- With D.O.E. the test volume is shared across cells and so a much smaller quantity per cell is required to test factor level main effects and interactions.
- In this example instead of using 1.2MM to conduct 3 tests only 400,000 is needed.
- Properly calculating the correct mail volume is critical to ensuring there is enough for significance testing without mailing and spending more than is needed.

🟦 ← TEST vs. CONTROL → 🟥

© 2012 Merkle Inc. All Rights Reserved | Confidential

What do you factor when calculating sample size?

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Determining Mail Volumes

ANALYTICS CRM EXCHANGE

The goal is to mail no more than is required to correctly identify true differences that are both statistically and practically significant.

- *Statistical* Significance: Difference unlikely to have occurred by chance
- *Practical* Significance: Difference of sufficient size to result in a strategy change

To Determine Mail Size Need To Know:

1. Expected performance of lowest performing factor
2. Desired significance (typically 90% or 95%)
3. Difference that is practically significant (5%, 10%, etc.)

Expected Performance	Significance Level	Practical Significance	Practical Performance	Required Mail Volume
0.40%	90%	10%	0.44%	751,704
0.40%	90%	15%	0.46%	342,010
0.30%	90%	10%	0.33%	1,003,328
0.30%	90%	15%	0.35%	456,506

* Test Power = 85%

The required mail volume is then spread evenly across the treatment cells

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Full Factorial and Fractional Factorial Designs

- A reduced test design is used when a very large number of factors wish to be tested. This involves mailing a specific subset of a full factorial design such that the non-mailed cells can be inferred through modeling. Testing of interaction effects is limited.

Example - A full factorial 2^3 design – three factors at two levels each.

	Factors			Interactions			
Test Cell	Offer A	Audience B	Creative C	A*B	A*C	B*C	A*B*C
1	-	-	-	+	+	+	-
2	-	-	+	+	-	-	+
3	-	+	-	-	+	-	+
4	-	+	+	-	-	+	-
5	+	-	-	-	-	+	+
6	+	-	+	-	+	-	-
7	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+

© 2012 Merkle Inc. All Rights Reserved | Confidential

Fractional Factorial Design

- In a fractional design all levels are tested against each other, but not all treatments are tested.

Example (cont'd): A ½ fraction of a 2^3 design or a 2^{3-1} design.

	Factors			Interactions			
Test Cell	Offer A	Audience B	Creative C	A*B	A*C	B*C	A*B*C
2	-	-	+	+	-	-	+
3	-	+	-	-	+	-	+
5	+	-	-	-	-	+	+
8	+	+	+	+	+	+	+

Note that columns A & B*C are identical, as are B & A*C and C & A*B. This means that this test will not be able to estimate any interaction terms.

© 2012 Merkle Inc. All Rights Reserved | Confidential

D-Optimal Designs

- D-optimal designs are one form of design provided by a computer algorithm. These types of computer-aided designs are particularly useful when classical designs do not apply.

- Unlike standard classical designs such as factorials and fractional factorials, D-optimal design matrices are usually not orthogonal and effect estimates are correlated.

- Given the total number of treatment runs for an experiment and a specified model, the computer algorithm chooses the optimal set of design runs from a *candidate set* of possible design treatment runs.
  - This candidate set of treatment runs usually consists of all possible combinations of various factor levels that one wishes to use in the experiment.

- Proc OPTEX in SAS allows a user to input the test parameters and constraints (including total number of test cells) and outputs possible designs, while highlighting which factors and interactions will be confounded

© 2012 Merkle Inc. All Rights Reserved | Confidential

Reasons for Choosing D-Optimal Designs

The reasons for using D-optimal designs instead of standard classical designs generally fall into two categories:

1. Standard or fractional factorial designs require too many runs for the amount of resources or time allowed for the experiment.
   - Example – if testing catalog versions, execution of a high number of versions may be impossible due to print vendor constraints

2. The design space is constrained (the process space contains factor settings that are not feasible or are impossible to run).
   - Example – if testing credit card terms, testing a high application fee and a high membership fee in combination violates the CARD Act

© 2012 Merkle Inc. All Rights Reserved | Confidential

D-Optimal Design Example

ANALYTICS CRV EXCHANGE

Candidate Set for Variables X1, X2, X3

X1	X2	X3
-1	-1	-1
-1	-1	+1
-1	+1	-1
-1	+1	+1
-0.5	-1	-1
-0.5	-1	+1
-0.5	+1	-1
-0.5	+1	+1
0	-1	-1
0	-1	+1
0	+1	-1
0	+1	+1
0.5	-1	-1
0.5	-1	+1
0.5	+1	-1
0.5	+1	+1
+1	-1	-1
+1	-1	+1
+1	+1	-1
+1	+1	+1

Final D-optimal Design

X1	X2	X3
-1	-1	-1
-1	-1	+1
-1	+1	-1
-1	+1	+1
0	-1	-1
0	-1	+1
0	+1	-1
0	+1	+1
+1	-1	-1
+1	-1	+1
+1	+1	-1
+1	+1	+1

As in the fractional factorial design, main effects and interactions can be inferred through modeling

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Testing Pitfalls

How to Avoid Common Mistakes

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Has anyone designed a test that failed?

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Avoid Common Testing Pitfalls

ANALYTICS CRM EXCHANGE

Planning

- Identify potential tests
  - Examine historic test results to understand what has or has not worked in the past
  - Prioritize new hypotheses for testing: which ones are most likely to have the largest impact?
- Define testing objective(s)

Pre-launch

- Determine KPIs that will be used to measure success
- Define attribution window and attribution methodology
- Determine sample sizes by leveraging expected response rates of KPIs
- Examine test and control creative to ensure objective of the test has been accomplished

Execution

- Ensure all other factors besides the test itself are identical between the test and control samples (unless using experimental design)
- Capture sufficient metadata to ensure test and control audiences can easily be identified on the back-end

Measurement

- Make sure recommendations are statistically significant and make business sense
- Watch out for confounding effects that may lead to incorrect test conclusions

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Case Study #1

Consumer Electronics Company

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Consumer Electronics Company

Background

- The objective of this test is to determine which factors contribute to a difference in open and click rates in order to use this information to improve future campaign performance
- A Design of Experiments (DOE) test was conducted using four factors as well as any interactions between those factors

Testing Strategy

- Four factors were considered with various levels for each factor
  - Email segment
  - Customer segment
  - Personalization
  - Sales Strategy
- In order to create the optimal test design to detect any factors which are significant or any two-factor interactions, 48 test cells were created using different combinations of the levels

© 2012 Merkle Inc. All Rights Reserved | Confidential

Factors & Levels of Interest

This test was designed for wide product coverage along multiple dimensions to expand the marketing universe.

Email Segment	New, Active, Inactive
X	
Customer Segment	Prospect, New, Active, Inactive
X	
Personalization	Consumer Name, Generic
X	
Sales Strategy	Call to Action, Teaser, Benefits, Promo

= **96 Subject Line Combinations**

© 2012 Merkle Inc. All Rights Reserved | Confidential

Test Design

D.O.E. with a fractional factorial design was used to reduce the number of test cells required to assess all main effects and two-way interactions

Full factorial design

96 Subject Line Combinations

Entire population

Sufficient sample to analyze co-variates

Balanced distribution across all drivers

Test all two-way combinations

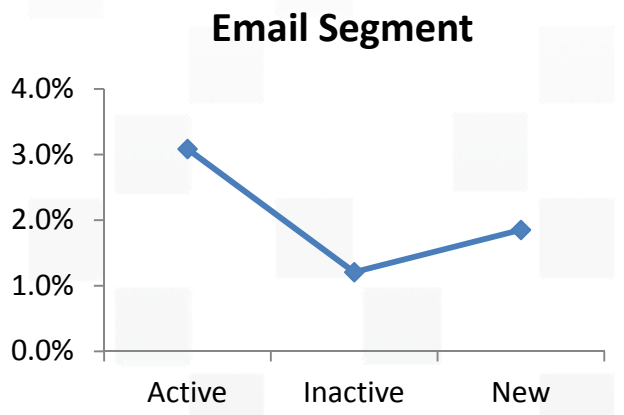
Fractional design

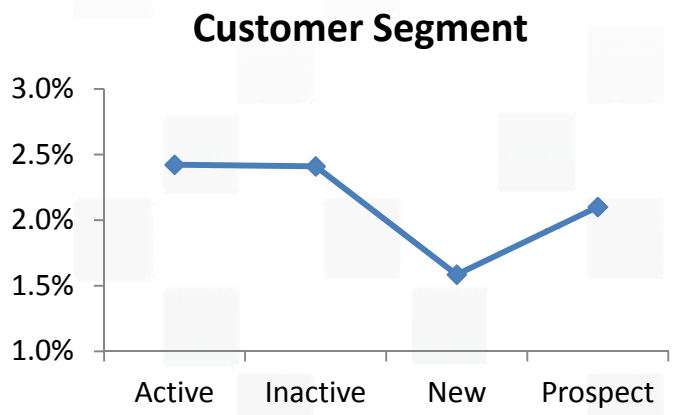
48 Test Cells

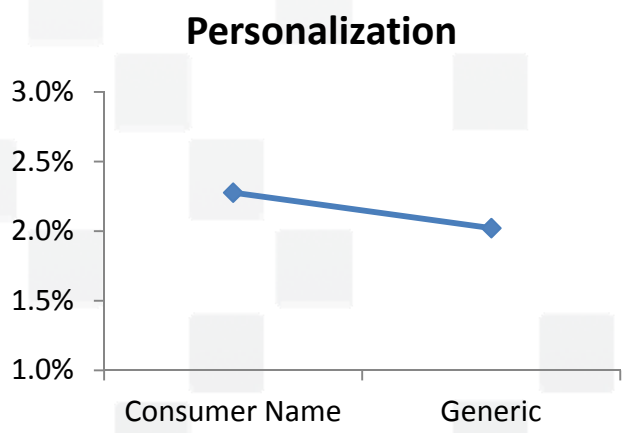
© 2012 Merkle Inc. All Rights Reserved | Confidential

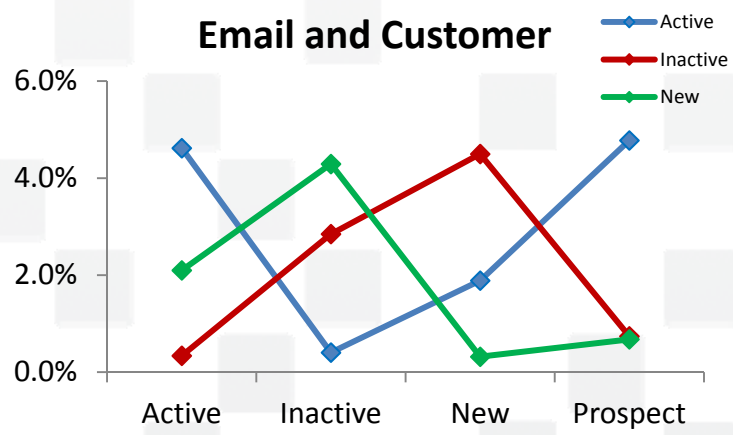
DOE Test Results

Email engagement, customer segment and personalization are significant, as well as the interaction between email engagement and customer segment









© 2012 Merkle Inc. All Rights Reserved | Confidential

Conclusions

- Results
  - Email segment, customer segment, the interaction between the two, and Personalization were found to be significant factors for click rate
  - Using the Consumer's name instead of a generic callout forecasts to result in an additional $1.4MM in email attributed revenue annually
  - **Due to challenges in test execution, it was not possible to measure the impact of variations in Sales Strategy. Client could not execute the numerous message versions they wanted to test and ultimately abandoned this portion of the test.**

- Recommendations
  - Implement using the Consumer's name on emails
  - Investigate other factors that might be significant for different segments of the email population, including other sales strategies as a part of the Phase II subject line test

© 2012 Merkle Inc. All Rights Reserved | Confidential

Case Study #2

Major Credit Card Company

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Major Credit Card Company

Background

- Building balances within the customer portfolio was extremely important, but traditionally focused on low risk, low return asset generation
- Testing up to this point had been conservative, changing only one or two pricing terms at a time from the control
- There was a desire to test different approaches to balance build and to understand the effect of different levers and some of the interactions between them

Testing Strategy

- A pricing terms test was designed to understand the effects of 5 different profit drivers and certain interactions between them using a fractional factorial design, which reduced the number of test cells needed from 950 to 50
- The test allowed for unique combinations of terms to be offered, all at a large enough sample size to be able to significantly read response rates and balance transfer amounts
- New products were rolled out as a result of this test, that allowed for achieving desired response rates, increasing balances, and maintaining similar levels of risk

© 2012 Merkle Inc. All Rights Reserved | Confidential

Factors & Levels of Interest

This test was designed for wide product coverage along multiple dimensions to expand the marketing universe.

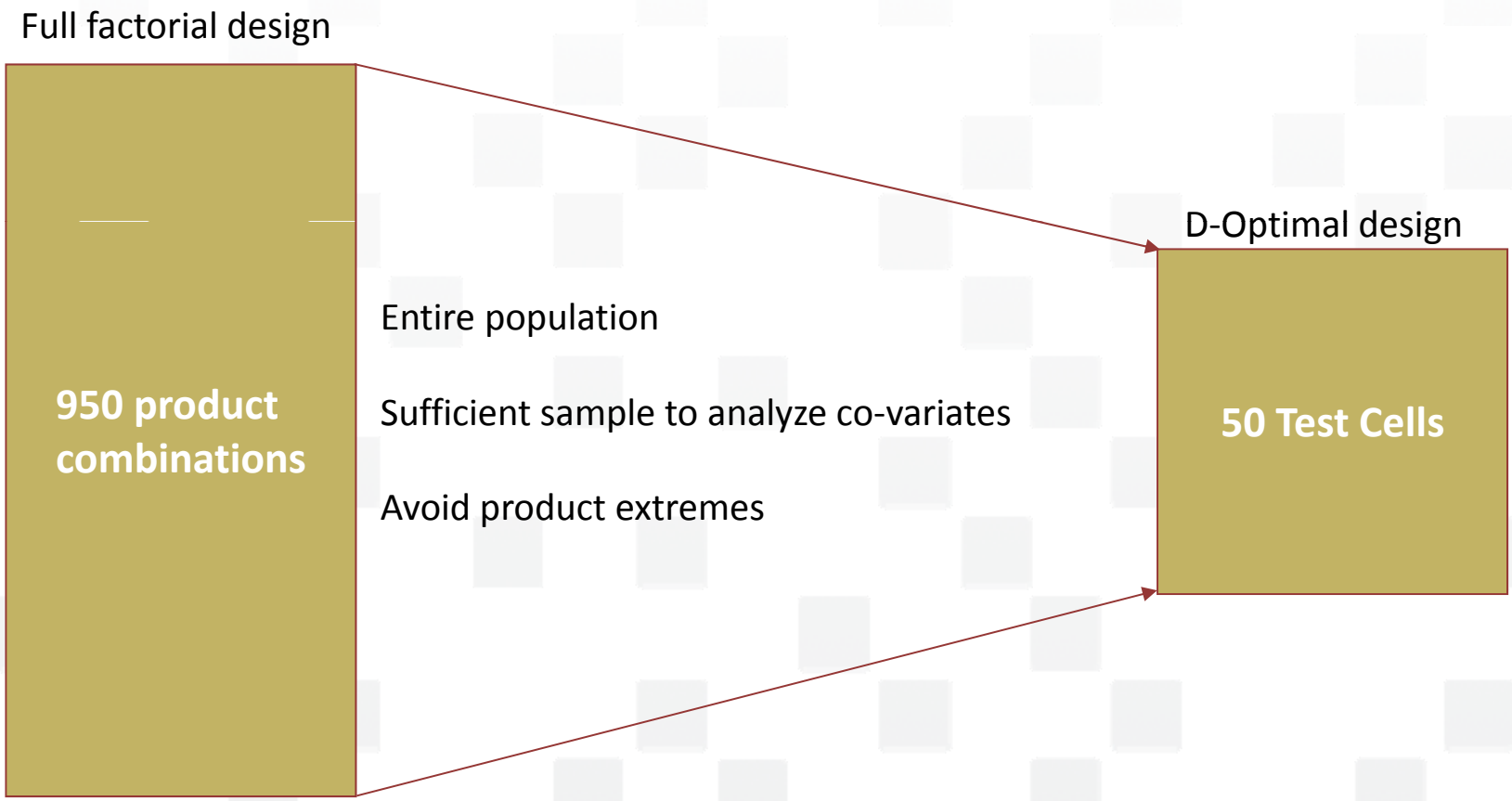
Intro Rate	0, 1.99, 2.99, 3.99 and 4.99%
X	
Duration	0 (fixed rate), 3, 6, 9 & 12 months
X	
Go to Rate	1.99% - 3.99%, increments of 1% 6.99% - 10.99%, increments of 2%
X	
BT Fees	0 – 3%, increments of 1%
X	
BT Fee Cap	$25 cap, $99 cap & uncapped

= **950 products**

© 2012 Merkle Inc. All Rights Reserved | Confidential

Test Design

D.O.E. with a D-optimal fractional design was used to ensure adequate sample size for analysis of population co-variates.



© 2012 Merkle Inc. All Rights Reserved | Confidential

Testing and Analysis Methodology

The test design enabled regression models to predict Net Response Rate and Balance Transfer Amount for a wide variety of products

D-Optimal DOE Test

Only 50 of 950 combinations tested

- Simplifies execution and fulfillment significantly

Allows reads on:

- Main effects
- Key 2-way interactions

Product Combinations

Intro Rate	5 Levels
Duration of Intro	5 Levels
Goto Rate	7 Levels
BT Fee	4 Levels
Fee Cap	4 Levels

950 Total Combinations

Statistical Modeling to analyze the DOE

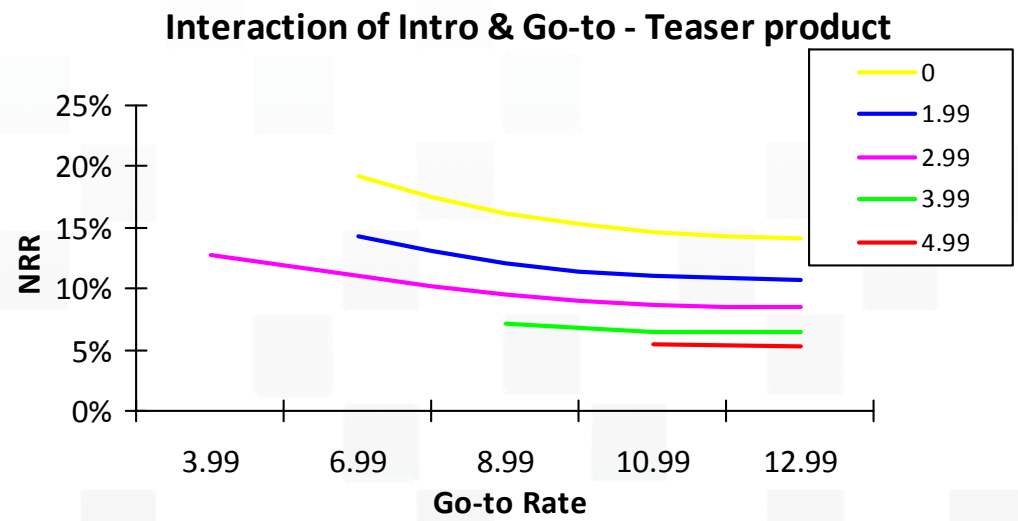
- Regression models built on account-level test data
- NRR = Logistic Function (Intro, Duration, Goto, Fee, Cap, Interactions)
- BT$ = Linear Function (Intro, Duration, Goto, Fee, Cap, Interactions)
- Models trained using all 50 test cells as input data and can predict all Product Combinations
- Incorporating account-level attributes allows better product optimization within segments

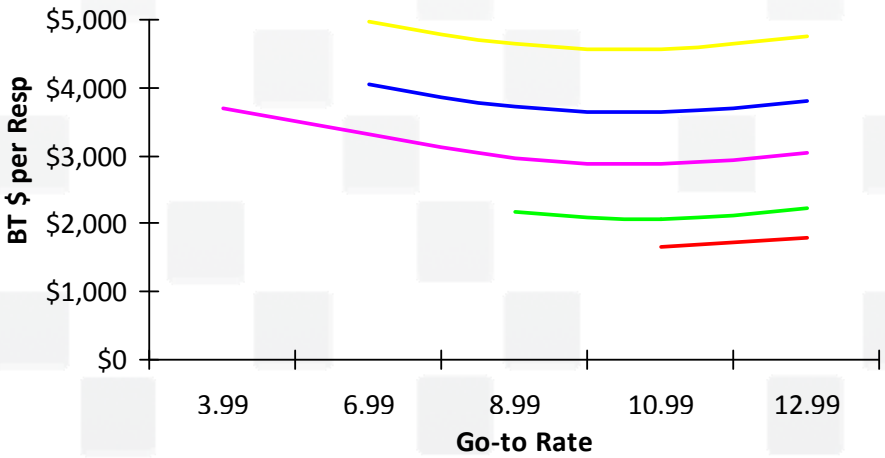
© 2012 Merkle Inc. All Rights Reserved | Confidential

Results: Product Type "Go-To Rate"

Response sensitivity to Go-to rate is significant only at low Intro and Go-To rates

- A 3.99 Go-To having less response than 6.99 Go-To is driven by the bias in test design
  - 3.99 Go-To is used only in combination with 2.99 Intro where as 6.99 Go-To is used with 0 and 1.99 Intro rates
- Go-To Rate has significant effect on NRR until 10.99, the significant effect on BT size is only until 8.99
- Effect of Go-To Rate is limited to lower intro rates (0 – 2.99)





© 2012 Merkle Inc. All Rights Reserved | Confidential

Conclusions

- Customers are very focused on introductory rate
  - A 1% decrease in intro rate can be balanced by 10% increase in Go-to without impacting response
- The benefit of doing longer teaser durations exists only for low intro rates
- Sensitivity to Fee depends on various population covariates
  - Effect of Fee on high FICO customers is almost 3x of low FICO customers
  - High Purchase APR customers are almost 2x less sensitive to Fee
- The more aggressive the product is, the more impact customer's purchase APR has on BT size
- Customers who rarely use their card are more price sensitive than customers who use their cards frequently

© 2012 Merkle Inc. All Rights Reserved | Confidential

Summary of Testing Methodologies

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Pros and Cons of Each Approach

Although complex, DOE tests provide the most robust results at the lowest cost.

	Champion/Challenger	OFAT	DOE
Most robust solution	x	x	✓
Accurate response estimate	✓	x	✓
Cost efficient	x	x	✓
Portable learnings	x	✓	✓
Simplicity	✓	✓	x
Bottom Line	Sub -optimal decisions	High cost & population utilization	**Efficient & precise learnings**

© 2012 Merkle Inc. All Rights Reserved | Confidential

When does it make sense to use Champion/Challenger vs. DOE?

Why can DOEs be risky?

ANALYTICS CRM
EXCHANGE

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential

Determining Which Approach to Use

- When trying to measure two overall strategies, without concern for the individual factors, a champion-challenger test would be a good fit.
- When trying to measure one factor with several different levels (3 different price points) an OFAT test would be a good fit.
- When measuring several factors, particularly if sample size is a restriction, a D.O.E. test would be a good fit.
- When measuring several factors and the interactions between each factor are of interest, a D.O.E test is the option for measuring the interactions effectively.
- A D.O.E with many test cells can be reduced by a fractional factorial design, but some interactions will not be measurable and therefore results may not be accurate.

© 2012 Merkle Inc. All Rights Reserved | Confidential

ANALYTICS CRM
EXCHANGE

Merkle Analytics

Shirli Zelcer
Senior Director, Merkle Analytics
443.542.4433
szelcer@merkleinc.com

MERKLE

© 2012 Merkle Inc. All Rights Reserved | Confidential