Experimentation in Data Science

Course Notes

Experimentation in Data Science Course Notes was developed by Catherine Truxillo, Ph.D. Editing and production support was provided by the Curriculum Development and Support Department.

Experimentation in Data Science Course Notes

Book code E70533, course code LWXDS/XDS, prepared date 03Dec2015. LWXDS_001

ISBN 978-1-62960-161-8

Table of Contents

Course Notes Description

This course explores the essentials of experimentation in data science, why experiments are central to any data science efforts, and how to design efficient and effective experiments.

To learn more…

For information about other courses in the curriculum, contact the SAS Education Division at 1-800-333-7660, or send e-mail to training@sas.com. You can also find this information on the web at http://support.sas.com/training/ as well as in the Training Course Catalog.

For a list of other SAS books that relate to the topics covered in this course notes, USA customers can contact the SAS Publishing Department at 1-800-727-3228 or send e-mail to sasbook@sas.com. Customers outside the USA, please contact your local SAS office.

Also, see the SAS Bookstore on the web at http://support.sas.com/publishing/ for a complete list of books and a convenient order form.

Prerequisites

None at this time

Chapter 1 Experimentation in Business

1.1 Why Experiment?

Objectives

- Explain the role of experiments in answering business questions.

3

You Need to Know

Work is full of questions that you need answers to.

Some have answers that require only a lookup:

- What is the policy regarding the use of demographic variables in predictive models?
- When did you last send a marketing email to segment 17?

Some do not have readily available answers:

- Does it really matter whether you use first-class postage when sending direct mailings for a cruise line?
- Is it better to assign consultants to projects based on geography or on line of business?

4

Statistical Models Can Answer Questions

The models that data scientists use can answer many of the questions that you have.

- Do you have the data to perform an analysis and answer the question?
- Did you account for the types of variables that are in your control as well as the types of variables over which you have no control?

5

Just as it is important to understand your subject area when fitting models, it is also important – and perhaps more important – to understand your subject area in designing an experiment that will be useful in answering questions.

Questions Often Require Comparisons

Does your question imply that a comparison is needed?

- First-class versus bulk-rate postage
- Geographical versus line-of-business assignments

Did you conduct an experiment?

6

The questions that do not have answers readily available are the subject of this course.

What Is an Experiment?

According to Merriam-Webster:

1. a scientific test in which ***you perform a series of actions*** and carefully ***observe their effects*** in order to learn about something.
2. something that is done as a test
3. something that you do to see how well or how badly it works

Experiments differ from observational data in the types of conclusions you can draw from them.

7

Observational versus Experimental

Observation: “As my son gets taller, the national debt increases.”

Incorrect conclusion: “Stop feeding your son!”

Correct conclusion: “A child’s height and the national debt increased over time.”

Correlation does not imply causation.

8

Observational versus Experimental

Experiment: "As I increase the dose of a drug by 5 mg, the symptoms decrease by 12%, on average, compared to a control group."

With an experiment, you can evaluate the cause-effect relationship between the things that you change in the experiment and the variable that you are measuring.

Experiments enable you to identify causal drivers.

9

Consider This…

What is the question that you want to answer?

What is the population that you want the answer to pertain to?

What types of things do you want to compare that you can control?

How is the outcome measured (Y_{obs})?

What else impacts Y_{obs} that you cannot control?

10

Each of these questions is addressed individually in the discussion that follows.

Consider This…

What is the question that you want to answer?

1. Does postage make a difference in the response rate?

2. Did the recent changes to human resource allocation increase profit on projects?

11

You ask these types of questions, explicitly or implicitly, most every time you make a decision. Can you formalize your question? Are there questions whose answers come merely from a gut instinct or acumen? Could those answers come from data? Do you have the relevant data available?

Chase (2009) discusses the usefulness of gut instinct versus data-based decisions in the context of demand forecasting. He makes a compelling argument that not only is gut instinct worse than evidence-based decision-making, it can actually be worse than a naive decision or a decision of no action.

Formulating a research question is the first step to getting valuable, evidence-based answers that lead to good decisions.

Consider This…

What is the question that you want to answer?

What is the population that you want the answer to pertain to?

1. The “luxury traveler” segment

2. Analytical consultants in the field

12

When you make evidence-based decisions, it is important that the data be relevant to the population to which you will be applying your decisions? Are you interested in a particular demographic? A region? A customer segment?

By making the population explicitly known, it is easier to extract appropriate and relevant data.

Consider This…

What is the question that you want to answer?

What is the population that you want the answer to pertain to?

What types of things do you want to compare that you can control?

1. The class of postage on the offer envelope

2. The specific project a consultant is assigned to work on

13

Knowing what you can control also relates to your question of interest. If you want to decide what class of postage to use, have you also considered the color of the envelope? What messages are on the envelope? What types of inserts are in the envelope? These are things within your control. Whenever possible, if you are not interested in comparing them, then you should be certain that they are constant throughout the study. For example, if you are not concerned with envelope color, then make sure that all offers have the same envelope color, or that envelope color is not confounded with other variables such as postage. Do not use a postcard for cheap postage and an envelope for first-class postage; avoid confounding postage with mailing format.

Consider This…

What is the question that you want to answer?

What is the population that you want the answer to pertain to?

What types of things do you want to compare that you can control?

How is the outcome measured (Y_{obs})?

1. The number of responses from each postage group
2. Profit as a fraction of budget for each project

14

This might seem obvious, but in fact it is critically important to good data analysis and it is one of the key questions in the scientific method. You want to know the answer to a question, but if your measured variables are not meaningful and reliable indicators of the thing that you want to know about, then the data might be useless. For example, if you want to know about the effect of postage on responses to a cruise offer, but you measure the number of respondents who actually went on the cruise, you might be looking at only part of the picture. The response to the card is the initial contact: requesting more information about the cruise. Whether they take the cruise is likely to be influenced by a host of other variables.

Consider This…

What is the question that you want to answer?

What is the population that you want the answer to pertain to?

What types of things do you want to compare that you can control?

How is the outcome measured (Y_{obs})?

What else impacts Y_{obs} that you cannot control?

1. Gender, vacation already taken that year, children
2. Consultant’s past performance, budget size

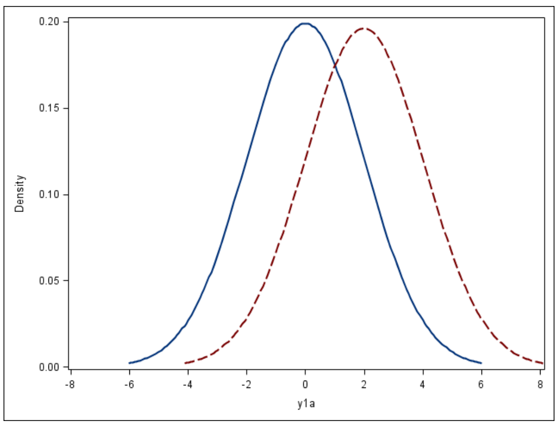
15

There are things that you can know but not control that can have an impact on the response. People with small children or people who have already taken a vacation that year might be less likely to inquire about cruises. Just because these variables are not part of your primary question of interest does not mean that you should not measure them. The “background noise” in responses, such as demographics, previous behavior, and environmental influences, has an impact on the precision of the estimates that you can obtain from your data analysis.

Who Cares about Things You Cannot Control?

You do!

Accounting for only the things in the experiment that you can control:



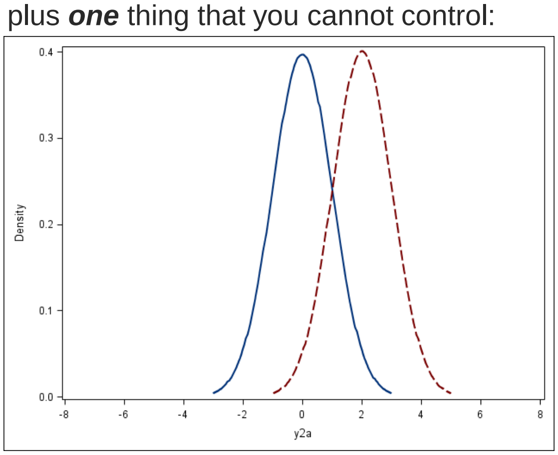
16

Suppose you want to compare the effect of postage on whether someone responds to a cruise offer. Comparing the two postage groups, as shown above, shows a difference of two units. However, this difference is not statistically significant for the observed sample.

Who Cares about Things You Cannot Control?

You do!

Accounting for the things in the experiment that you can control plus ***one*** thing that you cannot control:



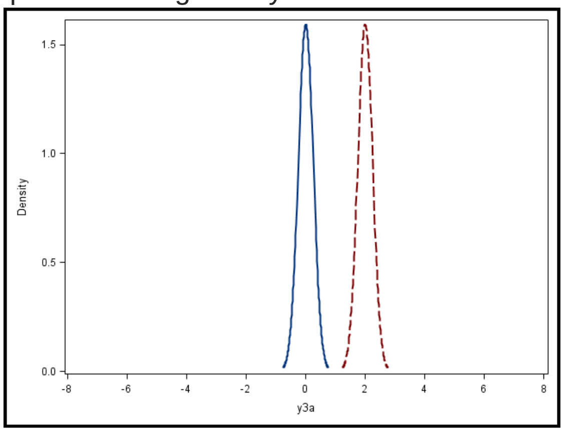
17

Controlling for a nuisance variable (for example, whether the person has children) reduces the amount of noise variation in response rate. Although the difference between the postage groups is still exactly the same (two units), the noise variance in the two groups is smaller, which makes the difference easier to detect.

Who Cares about Things You Cannot Control?

You do!

Accounting for the things in the experiment that you can control plus ***two*** things that you cannot control:



18

Accounting for another nuisance variable (for example, whether the person had already taken a vacation) reduces the noise variance even further. The difference between the postage groups is still two units, but the groups are clearly different, and easy to detect even with a fairly small sample. This is because the signal-to-noise ratio is much higher when you systematically account for nuisance variables.

Consider This…

What is the question that you want to answer?

What is the population that you want the answer to pertain to?

What types of things do you want to compare that you can control?

How is the outcome measured (Y_{obs})?

What else impacts Y_{obs} that you cannot control?

Work smarter: conduct an experiment!

19

In order to draw causal inference about whether one variable affects another, it is necessary to use an experiment. To determine whether postage affects response rate, you could randomly assign people to have one or the other class of postage. The remainder of this course is about doing exactly that.

Experiments Enable Small-Scale Deployments

- Many business decisions should be fact-checked to assess whether they have the intended consequences…and no unintended ones.
- Testing out many possible scenarios on a small scale enables you to compare which is the most profitable.
- Small demonstrations of success make it easier to communicate value to stakeholders.
- Experiments answer questions of causation.

20

An experiment is a systematic procedure carried out under controlled conditions (or, as controlled as possible) in order to discover an unknown effect. When analyzing a business process, experiments are often used to evaluate which inputs have a significant impact on the output, and what level those inputs should be to achieve the desired result. Experiments can be designed in many different ways to collect this information. Design of experiments (DOE) is also referred to as designed experiments and experimental design.

1.2 Essentials of Business Experiments

Objectives

- Define experimental design concepts and terminology.
- Relate experimental design concepts and terminology to business marketing concepts and terminology.

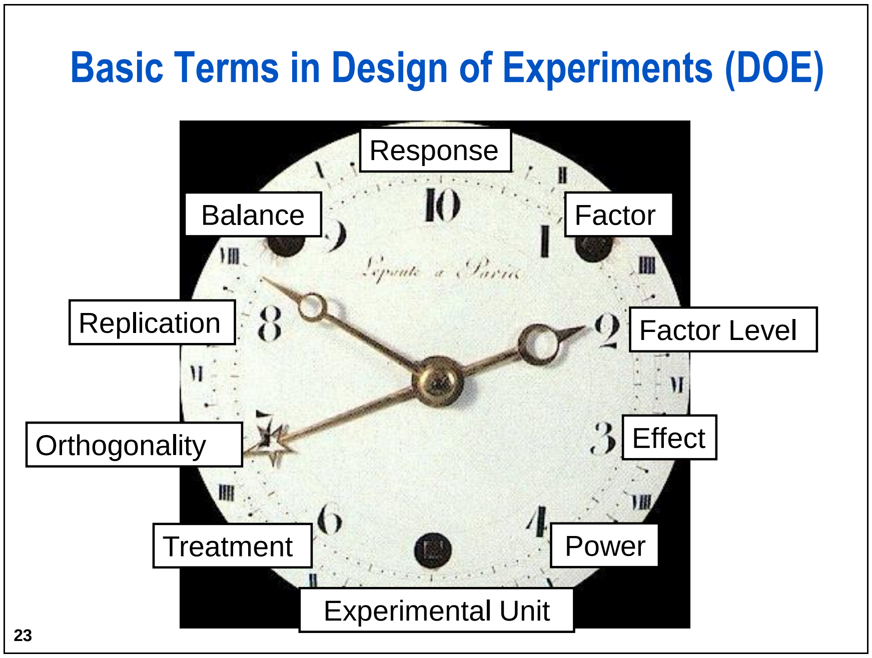
22

The purpose of this section is to introduce the concept of experimental design and how it can be applied in business. The main learning points are as follows:

- Experiments, in which you compare treatments or changes that you have made in an offer, can help you understand what things substantially influence how customers behave.
- One-factor-at-a-time (*OFAT*) experiments usually waste resources and can lead to wrong conclusions about what options offered to consumers might affect their choices.

The business objective is to ascertain features of products or services that maximize revenue or other business metrics. In the business community, experimental design is often called *test and learn*.

All technical fields have specialized terminology to enable efficient communication. Experimental design is no exception. The most important terms and their definition are presented below.



Design of experiments (DOE) can be thought of as setting parameters of data collection activity similar to setting the hands on a clock. A metric clock is chosen for this analogy because setting the hands of an experiment involves only 10 attributes, not 12.

The investigator can set the attributes of the experiment by defining the following:

- the population to which inferences are made
- the experimental units of the population
- the response variable (or variables) of interest

The investigator also must do the following:

- decide which factors and factor levels will be used in the experiment
- specify the magnitude of an effect that is meaningful for the decisions that will be made from the experiment
- specify the power for detecting meaningful effects

Finally, the investigator must decide which treatments to use (the combination of factors and their levels), the number of replications needed for each treatment, and the important balance and orthogonality for the overall efficiency of the experimental design. The investigator often has to make compromising choices to obtain the most information within his or her resource constraints.

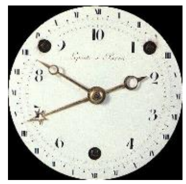
A brief discussion of the history and origin of the metric clock can be found at **en.wikipedia.org/wiki/Metric_time**.

Basic Terms in DOE: Response

A *response* is the variable of interest in the analyses.
It is sometimes called the *target* or *dependent variable.*

Examples:

- Response rate to direct mail solicitations
- Default (“bad”) rate among credit customers
- Balance transfer amount
- Fraud
- Number of items purchased from a catalog
- Spend, six months after acquisition



24

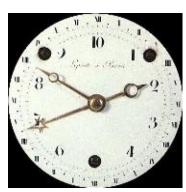
The response in an experiment is the dependent or outcome variable. This is referred to as a *target* in predictive modeling. The appropriate response is chosen for the specific purpose for which the experiment is conducted. Examples include response rate, dollar amount of purchase, balanced transferred, number of hits on a web link, and the number of purchases from a website or catalog. More than one response variable can be used when analyzing the data collected in an experiment.

Basic Terms in DOE: Factor

A *factor* is an independent variable that is a potential source of variation in the response metric.

Examples:

- Teaser or introductory APR
- Color of envelope
- Balance transfer fee
- Presence or absence of a sticker on a catalog
- First-class versus third-class mail
- Others?



25

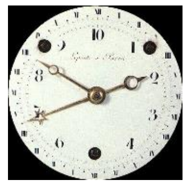
The factor is one of the attributes of a specific offer that can be controlled by the investigator. Examples include postage rate, interest rate for a credit card offer, color of an envelope, position of an advertisement on a web page, text size, and font of text in an offer or on a web page.

Basic Terms in DOE: Factor Level

A *factor level* is a particular value, or setting, of a factor.

Examples:

- 1.99% introductory APR
- White envelope
- 2% balance transfer fee
- Airline mile reward offer
- Third-class mail
- Others?

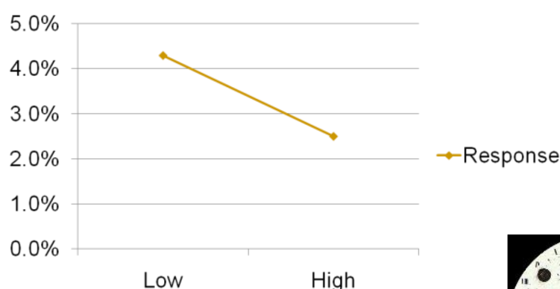


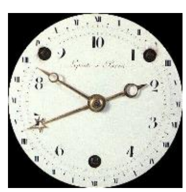
26

Because a factor is under the control of the investigator, different values or levels can be chosen and used in the experiment. Examples include two or more levels of postage, interest rates, different colors of envelopes, different positions of an advertisement on a web page, and different sizes of fonts for marketing messages.

Basic Terms in DOE: Effect

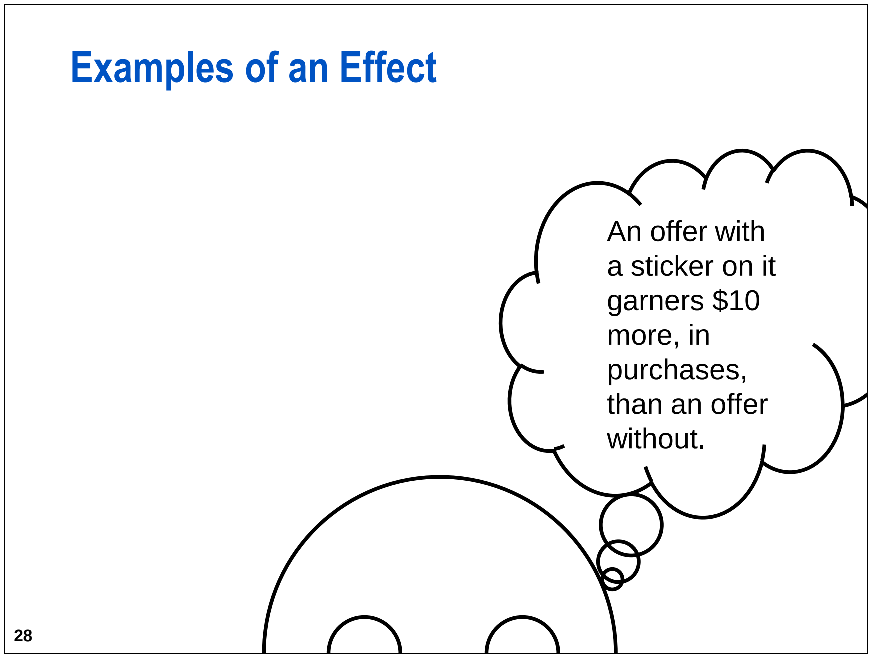
An *effect* captures and measures the relationship between ***changes in factor levels*** and ***changes the response metric***.

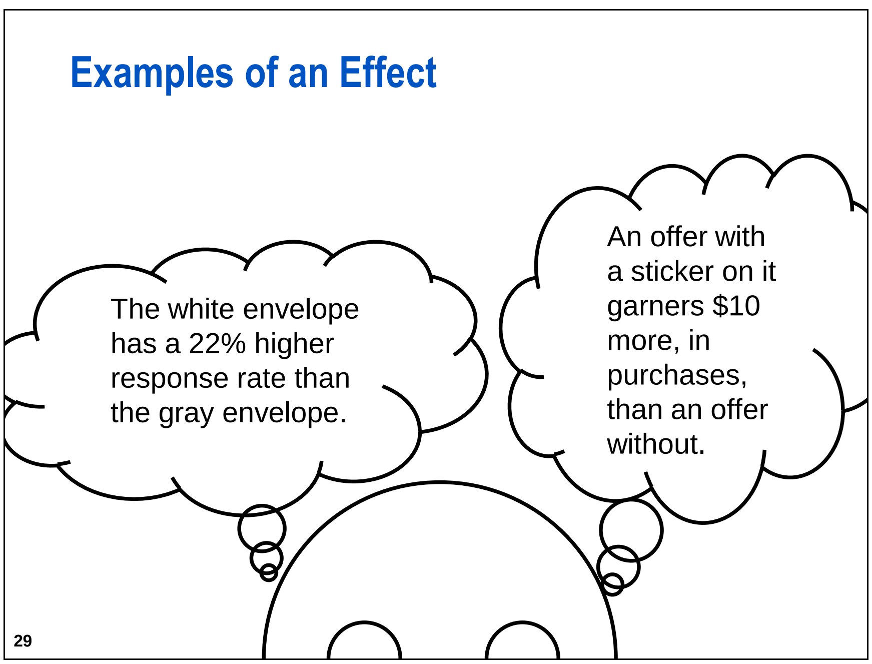


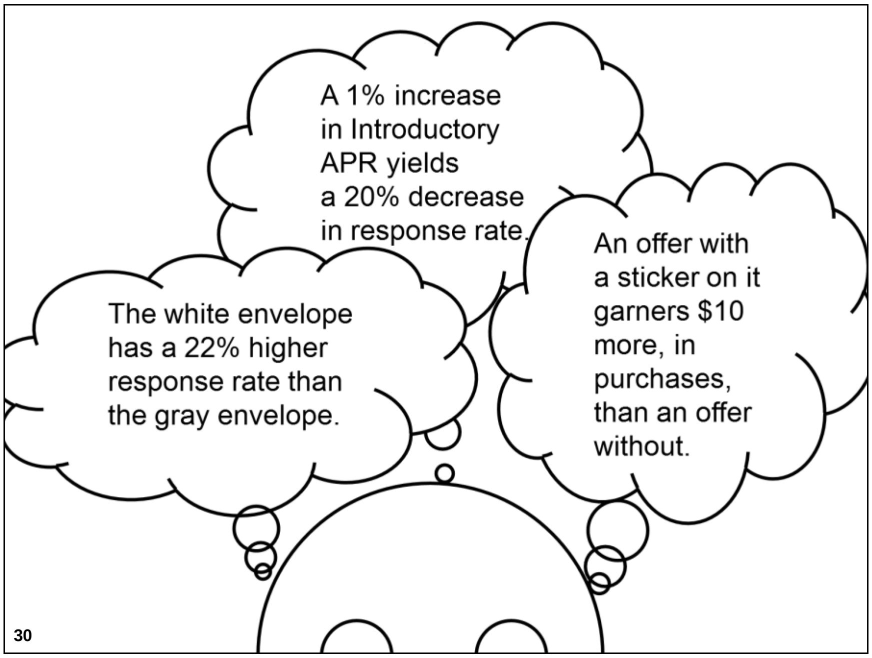


27

An effect is the difference in the response values for different levels of a factor. In the graph above, a line plot is used to show how the response (on the y-axis) changes as you change from the low level of the factor to the high level. A line plot implies that the factor could be a continuous variable, but in experiments, you typically have specific (fixed) settings for the factor levels. For example, the effect above is "People that were offered the low value responded almost at twice the rate as people that were offered the high rate."







A single experiment can uncover more than one effect.

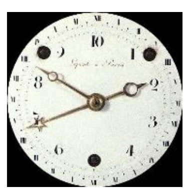
Basic Terms in DOE: Treatment

A *treatment* is a combination of all of the factors, each at one level. In a typical marketing context, a treatment constitutes a unique *offer*.

Examples:

- 2.99% Intro Rate, in a White Envelope, 4.99% Goto rate
- 0% Intro Rate, in a Gray Envelope, 7.99% Goto rate
- 2.99% Intro Rate, in a Gray Envelope, 7.99% Goto rate
- 0% Intro Rate, in a White Envelope, 4.99% Goto rate

There are eight possible ***treatments*** when you have three ***factors***, each at two ***levels***.



31

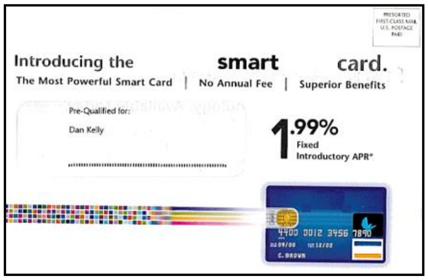
A treatment is a specific combination of the levels of the factors in an experiment. In this example, there are three factors (Intro Rate, Balance Transfer Fee, and Envelope Color) and each has two levels. All possible combinations of the levels of the factors define eight possible treatments (two rates x two fees x two colors = eight treatments).

Basic Terms in DOE: Treatment

A *treatment* is a combination of all of the factors, each at one level.
In a typical marketing context, a treatment constitutes a unique *offer*.

Examples:

- 1.99% Intro Rate, in a White Envelope, 4.99% Goto rate
- 0% Intro Rate, in a Gray Envelope, 4.99% Goto rate
- 1.99% Intro Rate, in a Gray Envelope, 7.99% Goto rate
- 0% Intro Rate, in a White Envelope, 4.99% Goto rate



There are eight possible ***treatments*** when you have three ***factors***, each at two ***levels***.

32 ...

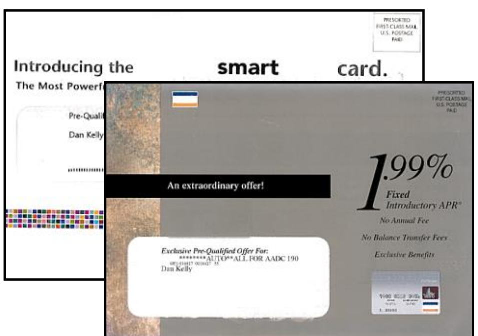
The unique offer represented by the envelope shown above contains a specific combination of the levels of the factors. It is an example of a treatment.

Basic Terms in DOE: Treatment

A *treatment* is a combination of all of the factors, each at one level.
In a typical marketing context, a treatment constitutes a unique *offer*.

Examples:

- 1.99% Intro Rate, in a White Envelope, 4.99% Goto rate
- 0% Intro Rate, in a Gray Envelope, 4.99% Goto rate
- 1.99% Intro Rate, in a Gray Envelope, 7.99% Goto rate
- 0% Intro Rate, in a White Envelope, 4.99% Goto rate



There are eight possible ***treatments*** when you have three ***factors***, each at two ***levels***.

33 ...

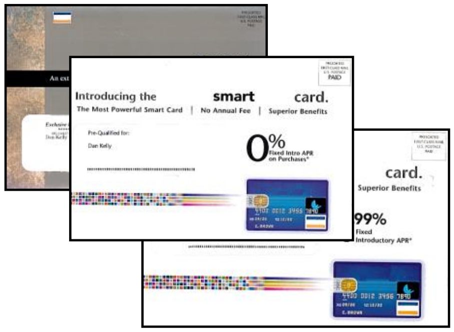
This example shows two different offers, or treatments, as two envelopes.

Basic Terms in DOE: Treatment

A *treatment* is a combination of all of the factors, each at one level. In a typical marketing context, a treatment constitutes a unique *offer*.

Examples:

- 1.99% Intro Rate, in a White Envelope, 4.99% Goto rate
- 0% Intro Rate, in a Gray Envelope, 4.99% Goto rate
- 1.99% Intro Rate, in a Gray Envelope, 7.99% Goto rate
- 0% Intro Rate, in a White Envelope, 4.99% Goto rate



There are eight possible ***treatments*** when you have three ***factors***, each at two ***levels***.

34

This example shows three different offers, or treatments, as three envelopes.

Idea Exchange

Web-based experiments are popular because they are relatively inexpensive to implement and they can be modified in real time.

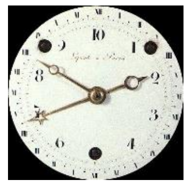
Suppose that you are designing a web-based experiment to compare different promotional offers on a retail site.

What types of factors might influence whether a customer purchases a product and the amount a customer spends? Think of at least two factors that you would want to design into your study. Write them down to use as ideas with your group.

In your groups, describe the experiment in detail, including levels of the factors, the specific definition of the response, and how you would implement the experiment.

Other Terms in DOE

- An *experimental unit* is the smallest unit to which a ***treatment*** can be applied.
- *Replication* occurs when more than one ***experimental unit*** receives the same ***treatment***.
- *Power* is the probability that you will detect an ***effect***, if one exists.

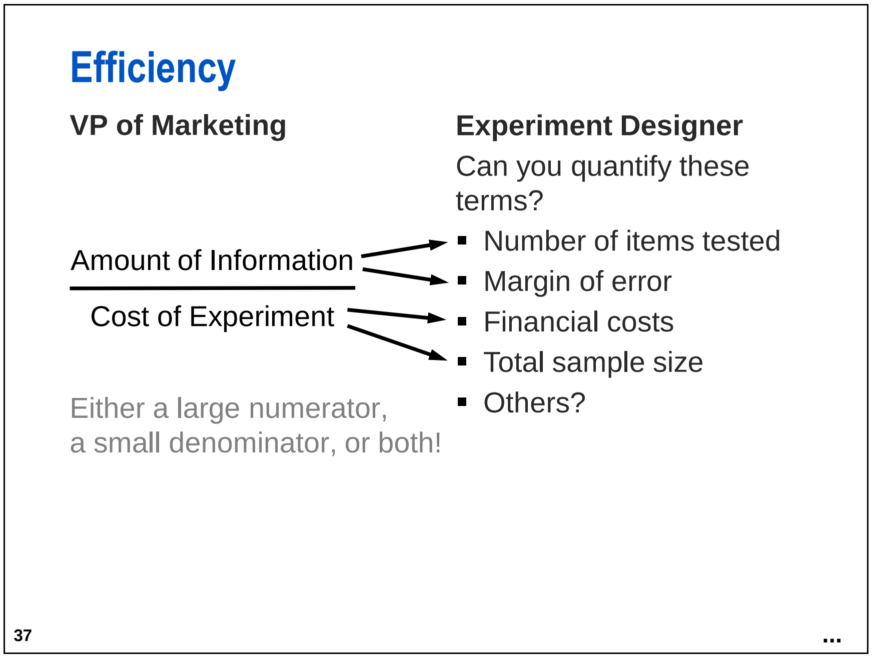


36

The appropriate experimental unit depends on the purpose of the experiment. Targeted marketing experiments usually are addressed to individuals or households. Experimental units for website experimental designs are usually visits because the problems of identifying return visitors make it difficult or impossible to identify an individual person.

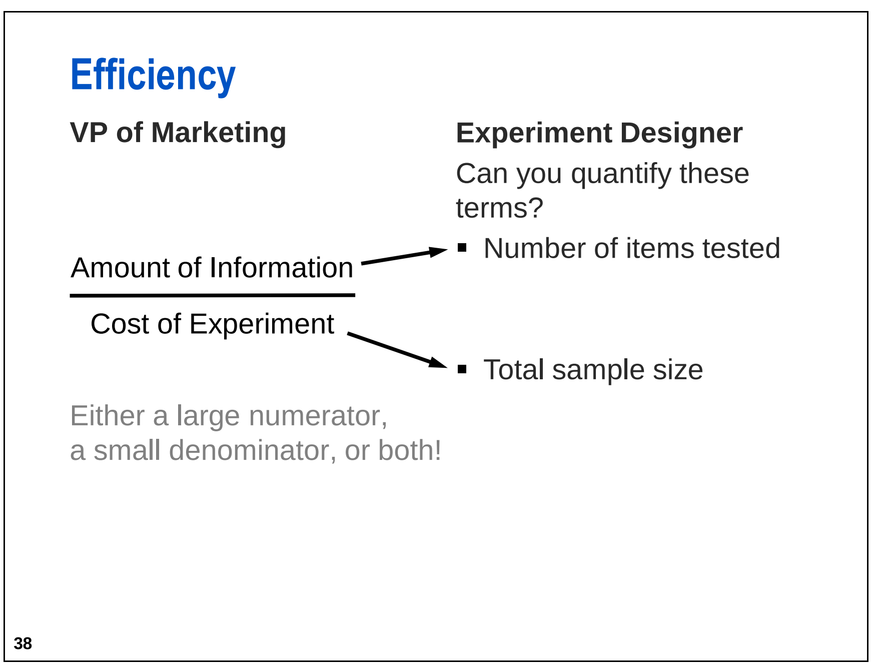
A thorough discussion of power is beyond the scope of this course, although it is helpful to consider what the appropriate number of observations should be for your experiment in order to have adequate power to detect the effects of your factors. A consulting statistician can help with these calculations. A common heuristic for large-scale business experiments (such as direct marketing campaigns) is 50,000 observations per factor level in the experiment.

Two other terms that are often associated with experimental designs are ***blocks*** (referred to generally as ***blocking***) and ***covariates***. These concepts are discussed later in the chapter.

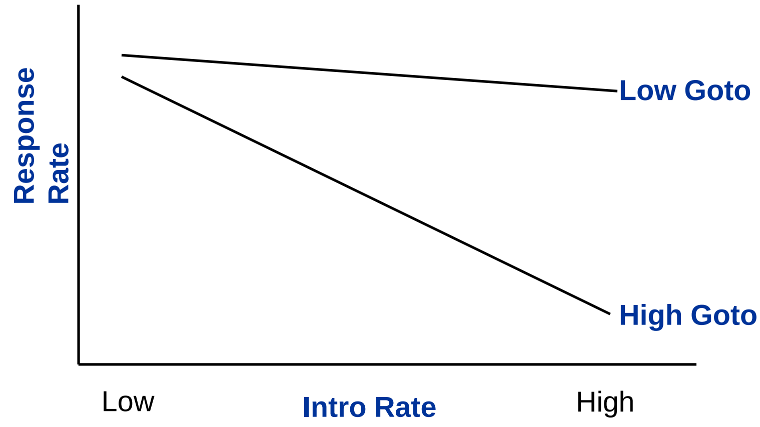


Efficiency means different things to different people, depending on their viewpoint and technical knowledge. In this context, efficiency is represented by a high amount of information relative to the cost of the experiment. However, the amount of information can have different meanings, and the cost of the experiment can mean different things. Regardless of one's viewpoint, more efficient is better. The efficiency can be improved if the numerator is increased or the denominator is decreased, or both.

The executive vice-president of Marketing might think in terms of the number of items tested and the total financial cost. A statistician might quantify the amount of information as the margin of error of estimated effects. Smaller margins of error imply more information. The main driver of financial cost is the total sample size. For this discussion, efficiency is defined to be the number of items tested divided by the total sample size.



Detecting Interactions between Factors



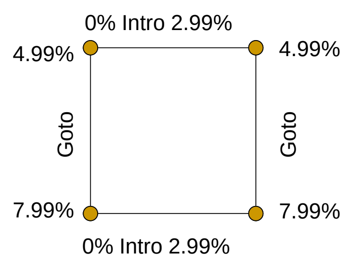
39

If there is an interaction between the two factors, then the conclusions that you reach will be different, depending on which of the four scenarios is conducted. Consequently, the results might be misleading. For example, if **Goto** is tested at the low level of **Intro**, you might conclude that there is little effect. However, if **Goto** is tested at the high level of **Intro**, you might conclude that there is a substantial effect. Similarly, if **Intro** is tested at the low level of **Goto**, you might conclude that the **Intro** effect is not large. If **Intro** is tested at the high level of **Goto**, then you might conclude that the **Intro** effect is large.

How can you avoid being misled by a sequence of OFAT tests?

The answer is to conduct a multifactor experiment as shown below.

Factorial Arrangement of the Treatments



- Permits the testing and estimation of an interaction term
- Increases the precision of estimates for the same test volumes
- Can use every individual in every test

Combinations of factor levels provide replication for individual factors.

40

A multifactor experiment enables you to test more items with a smaller sample size. The bullet points above are explained below.

Randomization

After an experiment is defined, the next step is to randomly assign treatments to experimental units.

A typical approach to randomization of *N* customers to k treatments involves the following steps:

1. Define the population of interest.
2. Select a simple random sample from the population equal to the total sample size, *N* (for example, *N*=100,000).
3. Randomly partition the sample into *k* equal groups (for example, *k*=4 groups of 25,000 each).
4. Assign each group to one of the four treatments.

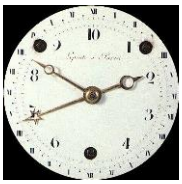
41

Randomization avoids bias from factors that are not explicitly accounted for in the design. Factors that can be controlled and factors that cannot be controlled are part of the experiment. Some factors that might affect the response cannot be controlled because they are attributes of the experimental units, not the treatment structure. Methods to control for attributes of the experimental units, if they are known in advance, are discussed later in this chapter under the topic of blocking.

Orthogonality

Another ideal property of an experimental design is orthogonality among the elements of interest. There are at least three ways to think about the importance of this property:

- Algebraic interpretation – Matrices behave well.
- Geometric interpretation – Pictures look nice.
- Statistical interpretation – Estimates have low variance.



42

When analyzing designed experiments, computer software codes the levels of factors using, for example, +1 and -1. On the next slide is the usual coding for the analysis of an experiment that has two factors, each of which has two levels.

Two-Level Full Factorial Coding

I	A	B	AB
+1	+1	+1	+1
+1	+1	-1	-1
+1	-1	+1	-1
+1	-1	-1	+1

43

The I column, all +1s, represents the intercept term, the overall mean of the response. The A column represents the two levels of the A factor (+1 or -1). The B column presents the two levels of the B factor (+1 or -1). The AB column represents the interaction between the two factors (+1 or -1). Notice that the values of the AB column are the product of the values in the A and B columns. Coding the data in this manner produces a matrix (the design matrix) that has very desirable properties. The matrix is *orthogonal*. Orthogonal is a multivariate equivalent of perpendicular in two dimensions. You can show that a matrix is orthogonal by taking the inner product (dot product) of each column with the others. If inner products equal zero (0), then the matrix is orthogonal (for example, the inner product A and B = (+1)(+1) + (+1)(-1) + (-1)(+1) + (-1)(-1)) = 1-1-1+1 = 0).

Balance also ensures orthogonality for a factorial arrangement of treatments. Although it is possible for unbalanced designs to be orthogonal (with some amount of effort in creating the design, or by using specialized experimental design software), all balanced factorial designs are orthogonal. Orthogonality implies ***statistical independence***. This is desirable because statistical independence gives the lowest variance estimated for the given sample size.

Effect of Factor A

I	A	B	AB
+1	+1	+1	+1
+1	+1	-1	-1
+1	-1	+1	-1
+1	-1	-1	+1

44

The effect of A is estimated by taking the difference in the mean of the response between the rows of the data where A is at one level (+1) (the shaded rows) minus the rows of the data where A is at the other level (-1) (the unshaded rows). Because the A column and the B column are orthogonal and the data are balanced, the test for A is independent of B. Notice that in each pair of rows the levels of B and AB are balanced between +1 and -1.

Effect of Factor B

I	A	B	AB
+1	+1	+1	+1
+1	+1	-1	-1
+1	-1	+1	-1
+1	-1	-1	+1

45

Similarly, the effect of B is estimated by taking the difference between the mean of the response for the two levels of B, the shaded rows versus the unshaded rows.

Interaction Effect AB

I	A	B	AB
+1	+1	+1	+1
+1	+1	-1	-1
+1	-1	+1	-1
+1	-1	-1	+1

46

Similarly, the effect of the interaction is estimated by taking the difference between the mean of the response between the two levels of the interaction term, the shaded and unshaded rows.

Factorial Arrangement versus OFAT

Factorial Treatment Structure

Pros

+ ***Reuses*** observations (more power for fewer exp units)
+ Tests for ***interactions***
+ Guarantees balanced and orthogonal treatment plans
+ Is an efficient way to test many factors

Cons

- Can be more complicated to set up
- Can be more complicated to sell to a nontechnical audience

47

There are many practical and theoretical advantages to using a factorial arrangement of treatments for experimental testing in a business environment. The most important among these is the ability to test for interactions, although the improved efficiency in using data is also a major advantage. So why aren't all businesses using multifactor experiments? Of course, there are always negatives. One of the barriers to implementing these types of designs is that others might not understand them. Another is that many are reluctant to make changes to methods that were used in the past, especially if they do not understand the new approach. Ignorance of analytical methods can be a serious impediment in the modern marketplace.

Factorial Arrangement versus OFAT

One-Factor-at-a-Time Tests

Pros

+ Are easy to set up (A/B and Champion/Challenger tests are typical in many industries.)
+ Might yield lower per-unit printing costs
+ Have clear “control” offer, clear test offers
+ Do not require users to learn new words such as “balance” and “orthogonality”!

Cons

+/- Permit simple analysis that could be done with a pencil and paper!

- Do not allow a test for interactions
- Represent an ***inefficient*** use of experimental units

48

The advantages of OFAT are merely practical. The challenge to an advocate of multifactor methods is to convince decision makers of the advantages and to convince them that those advantages outweigh the disadvantages.

Factors, Blocks, and Covariates

It is typical to use the same statistic to test
(H0:$p_{\text{men}} = p_{\text{women}}$) as
(H0:$p_{\text{red envelope}} = p_{\text{blue envelope}}$).

Are these factors equivalent from the perspective of experimental design?

49

In the slide above, *p* refers to the probability of a response. The test that evaluates whether a factor that you can control affects the response is often the same type of test that evaluates whether an element that you cannot control affects the response. What is the difference?

Measured Variables

You ***can*** control features of the offer that you make:

- creative
- color
- pricing
- duration of offer

Any restrictions are typically self imposed.

These are usually factors in the test.

You ***cannot*** control features of your experimental units:

- risk profile
- responsiveness
- geography
- age
- gender

Restrictions here are typically features of the population of interest and are often treated as blocks or covariates.

50

At the beginning of the chapter, you saw examples of nuisance variables. Recall that accounting for these nuisance variables reduces the background noise in detecting a difference between experimental treatments. Sometimes features of the experimental units that cannot be controlled are known in advance and can be incorporated into the design. One method of doing this is called *blocking*, which entails randomizing the assignment of treatments within blocks. For features that are not known before the experimental design, they can be measured during the experiment and used in the analysis. The first situation involves constructing randomized block designs. The second method involves analysis of covariance.

Blocking

Blocks are groups of experimental units that are homogeneous in some way. Typically, they represent nuisance variability.

- Examples: region, school, company
- Blocks might or might not be randomly selected.
- Because units exist in blocks, rather than being assigned to them, blocks reflect a restriction on the randomization in an experiment.

51

Covariates

Covariates are characteristics of the experimental units that are measured but cannot be assigned or imposed upon them.

- Examples: age, gender, income
- Covariates are usually not selected, but are characteristics of the units that have been selected.
- Covariates often represent nuisance variability.

52

What Is the Difference?

Blocks and covariates are sometimes treated interchangeably, and the distinction is not always clear.

- Blocks are typically categorical and can be thought of as groups of experimental units. Blocks are often used as random effects in models.
  - For example, a neighborhood is a group of households.
- Covariates can be continuous or categorical and are characteristics of the individual experimental unit. Covariates are often used as (fixed) predictors in models.
  - For example, a person is male or female and is of a particular age.

53

Idea Exchange

Imagine you are designing an experiment to find the best seating arrangement for a restaurant. Your factor is seating arrangement (A, B, or C).

These are your responses:

Total food and drink sales per party

Customer satisfaction ratings

In addition to the factor in the experiment (seating arrangement), list two or three possible block variables and/or covariates that might be considered in this experiment.

In your groups, plan this experiment and create a storyboard for the design how to implement the experiment. What do you expect to learn? How could this be useful for your business?

Other Designs

Data scientists use different types of experiments to answer questions.

- Factorial
- Response surface
- Split-plot (usually in time or geography)
- A/B (or Champion-Challenger)
- Discrete choice
- and many more

55

Focus on Designs: A-B Tests

CHAMPION

CHALLENGER

56

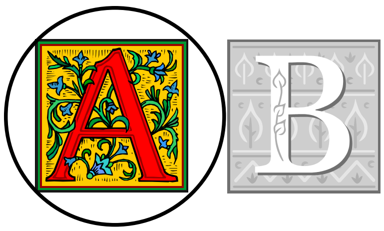
A/B tests, or champion-challenger, are very simple designs that are commonly used in website design and e-commerce. In these designs, rather than decomposing a candidate stimulus (such as a web page, banner ad, headline, or recommendation based on a search term) into many distinct factors, two (sometimes very different) stimuli are created and randomly assigned as the stimulus in the experiment. These designs are most useful when the experiment compares qualitatively different stimuli, such as a flash banner ad with animation on a website compared to a text-only email that is sent privately to customers. It simply would not be possible to create a factorial design of all the features that differ between these two approaches.

Focus on Designs: Complex A-B Tests

57

A variation of A/B tests can be used to compare many alternatives in a sequence. In these designs, perhaps there are many several stimuli. (The example above addresses five.) The five stimuli are randomized to a sequence (such as B→A→E→D→C, D→C→A→E→B, and so on). The first pair of stimuli is shown, and respondents are asked to select a preferred stimulus. Then the preferred stimulus is compared to the next in the sequence, and so on, until all have been shown. For each observation, a “champion” is determined.

Focus on Designs: Complex A-B Tests

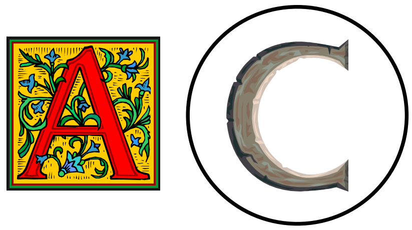


58

Focus on Designs: Complex A-B Tests

59

Focus on Designs: Complex A-B Tests

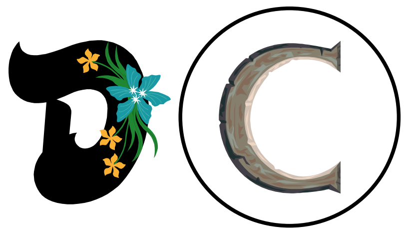


60

Focus on Designs: Complex A-B Tests

61

Focus on Designs: Complex A-B Tests

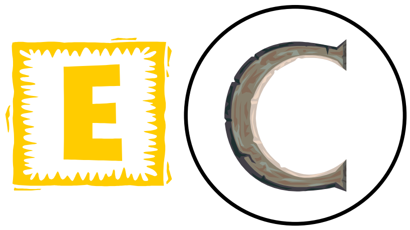


62

Focus on Designs: Complex A-B Tests

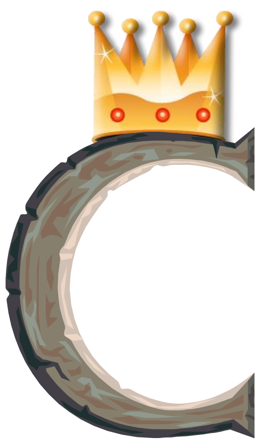
63

Focus on Designs: Complex A-B Tests



64

Focus on Designs: Complex A-B Tests



65

As with any design, A/B tests have drawbacks. The primary concern with them is that you cannot determine which feature of the champion (compared to the challenger or challengers) made it preferable. This makes it difficult to generalize your findings to other design projects. However, in fast-paced testing environments such as e-commerce, the speed of being able to obtain empirical customer preferences for every new problem negates the need for more complex (and more generalizable) experimental designs.

Chapter 2 Incremental Response Models

2.1 Introduction

Objectives

- Explain the difference between a model prediction and the action taken based on the model.
- Discuss how experiments and models together can distinguish between the effectiveness of a model and the impact of an action.
- Explain incremental response models and how they can identify cases that are most responsive to an action.

2

You cannot have “data science” without “science,” so it is important to follow scientific methods to get accurate results. But the other part is “data,” and when we put those together, a quote from Computer Science pioneer (and Rear Admiral in the US Navy) Grace Hopper really applies here: “One accurate measurement is worth more than a thousand expert opinions.” Successful organizations rely on data to prove that something will work – instead of hunches about what might possibly work – in order to make decisions.

Data can prove whether something is successful… or not. Danger comes when you become personally attached to having only one outcome. Your goal as a data scientist should be to do whatever will be most effective, not to be right 100% of the time.

The models in this chapter are described more thoroughly in Lee, Zhang, Meng, and Ryan (2013).

Deploying a Model

- You know how to build a predictive model.
- You know how to score new data.
- Now you must take action based on those scores.
- How do you know whether the deployment is effective?
  - Did you identify the right cases?
  - Did the action have the desired consequence?

Experiments are an effective way to separate the impact of the model from the impact of the message.

5

Evaluate Model Deployment

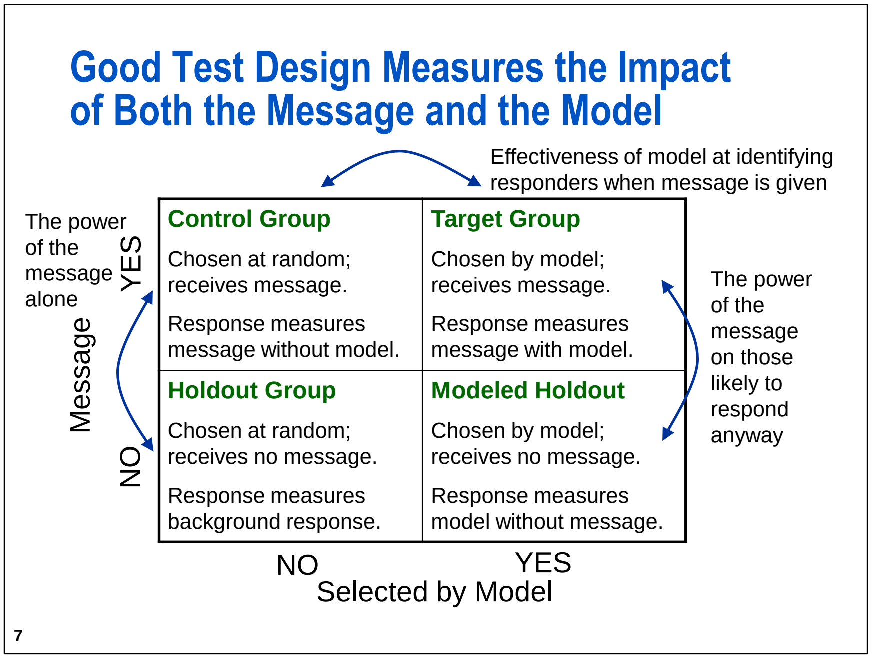
- Compare actual results against expectations.
- Compare the challenger's results against the champion's.
- Did the model find the right people?
- Did the action affect their behavior?
- What are the characteristics of the cases most affected by the intervention?

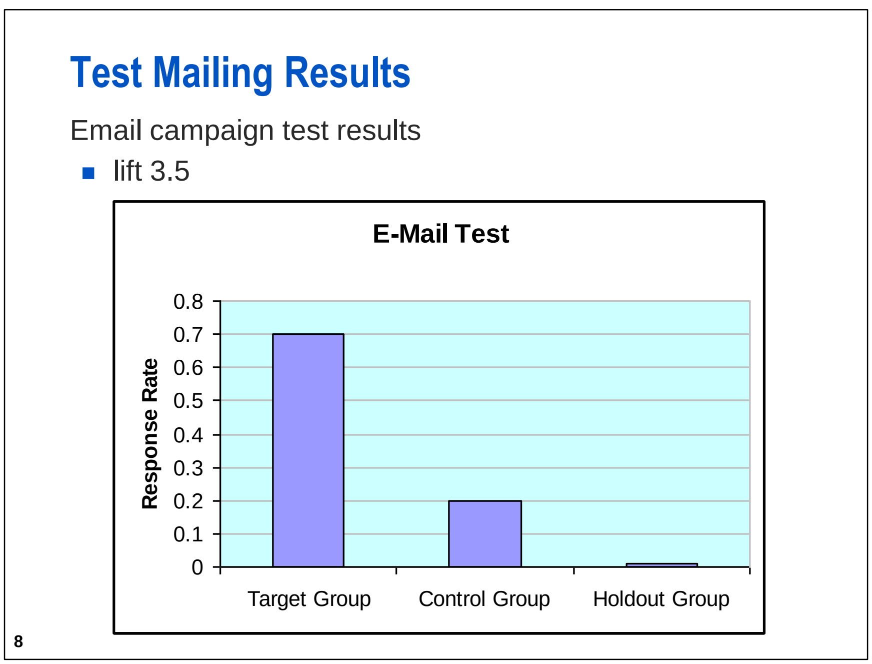
6

The real test of data mining comes when you can measure the value of the actions that you have taken as a result of the mining. Measuring lift on a test set helps choose the right model. Profitability models based on lift will help decide how to apply the results of the model. But, it is very important to measure these things in the field as well. In a database marketing application, this requires always setting aside control groups and carefully tracking customer response according to various model scores.

When thinking about designing an experiment to assess results, the following are some things to keep in mind:

- the right treatment size for the test
- being sure that any test groups are chosen randomly and get the same treatment (same message, same offer, same timing, similar customers)
- being sure that operational systems can handle the process





This graph is taken from the assessment of an email campaign by a bank. The email was sent to customers who gave permission to be contacted in this manner. The email suggested that they sign up for a particular product. Ten thousand people were selected at random to receive the email. This is the control group, which had a 0.2% response rate. Another ten thousand people were sent the email because they were scored by the model as likely to be interested in the product. This group responded at the rate of 0.7%, or three and a half times the rate of the control.

Clearly, the model did something. It found a group of people more likely to sign up for the product than the average customer. Was it the email that caused them to sign up, or did the model simply identify people who were more likely to sign up with or without the email? To test that, the bank also tracked the take-up rate of people who did not get the email. It was virtually nil, so both the model and the message had an effect.

What this picture does not show is the response rate for people who were chosen by the model but did not receive the message (the modeled holdout). Such a group would enable you to evaluate the effect of the message.

The Data Science Lab

Carefully designed experiments enable you to evaluate many factors, interactions, and scenarios to identify key drivers.

These experiments are successful by virtue of the data scientist’s ability to carefully control the factors, similar to a laboratory setting.

Data science labs can be rich with many experiments, trying out ideas on a small scale. Successful experiments can lead to changes in how the business is run.

9

Out of the Lab

Although the deployment has left the lab, it is important to continue to learn from any decisions made as a result of experiments in the lab.

Justification to continue to experiment outside the data science lab:

- unintended consequences
- population drift
- incremental response

10

Unintended Consequences

Plan:	Product recommendation to generate new sales.
Unintended consequence:	Cannibalized sales of another product, net loss of profit.
Plan:	Solicit customers that the model predicts are likely to respond.
Unintended consequence:	Your marketing approach turns off customers and drives them away.

11

Deploy on a Small Scale

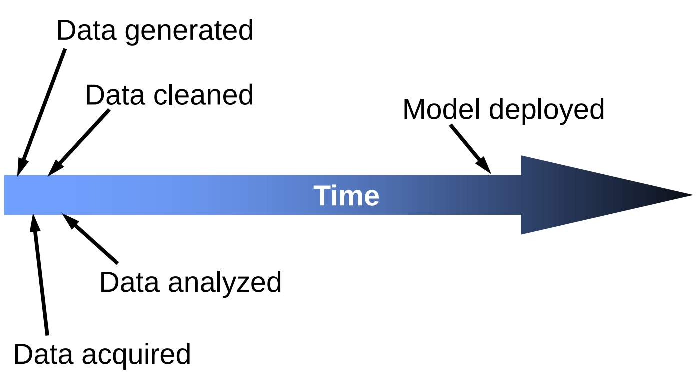
Smart move: Phase in new actions to evaluate unintended consequences without compromising the business.

One example of phasing in an intervention:

Week	Old Action	New Action
1	99%	1%
2	95%	5%
3	90%	10%
4	75%	25%
5	50%	50%

12

Scoring Pitfalls: Population Drift



13

Having an independent test set does not guarantee generalization. The population is dynamic. The test set might not adequately reflect the population at future times. Predictive models should be monitored, revalidated, and periodically refitted.

The Offer and the Response

When an action is taken to drive customer spend, there is typically some cost involved.

- Some people respond regardless of the offer. (Example: brand-loyal customers)
- Some people respond because of the offer. (Example: many competing brands)
- Some people do not respond regardless of the offer. (Example: irrelevant product)
- Some people are less likely to respond because of the offer. (Example: phone call during dinner)

14

Customers can be assumed to belong to one of the following four groups:

- Persuadables
- Sure Things
- Lost Causes
- Sleeping Dogs

The Persuadables are the customers, who must be persuaded to buy. That is, these are the customers, who respond ***only*** if they are targeted by a marketing action. The Sure Things are the customers who would respond anyway—that is, with and without the marketing action. Thus, the marketing action has no effect on the response rate here. The Lost Causes are the never-responding customers. That is, they do not respond even with the marketing action. But if the marketing action does not take place, they also do not respond. Thus, the marketing action has no effect here. Finally, we have the Sleeping Dogs, also called Do Not Disturb Customers. These customers show a lower probability of responding if they are targeted by a marketing action than if no action takes place.

We see that the only customer group that should be targeted by the marketing promotion is the group of Persuadables. Many marketing modeling tools concentrate on overall predicting which customers have the highest probability of responding. This means that many Sure Things automatically are also targeted, which is unnecessary.

It should be noted that the techniques in this chapter are useful only if Sure Things actually exist. If a prospect could never know that a product exists (for example, due to lack of marketing), then there would be no response at all without the offer. In this case, the incremental response model is not usable.

Models and Actions

Traditional predictive models identify the customers who are most likely to respond. However, some of these customers would respond anyway.

Ideally, action is taken exclusively on those unlikely to buy if there is no promotion, but are very responsive to offers/incentives/contacts.

15

Control Group

A control group is a randomly sampled group of candidate cases (customers, prospects, store locations, transactions, and so on) that are held back from a policy change, offer, or action. In medical research, the control group often receives a placebo.

This might seem unfair, at first, to hold back a potentially beneficial action. However, the control group is critical to fact-based decision making and smart business practice.

16

The Value of the Control Group

By comparing the results of every business modification or offer to a control group, you can do the following:

- differentiate correlation from causation regarding changes to customer behavior over time
- identify customers who would have responded the same regardless of the business modification
- identify business processes that are detrimental to business, customer experience, or profitability
- deploy many possible changes on a small scale and determine objectively which is best
- eliminate the effect of HiPPO (Highest Paid Person's Opinion)

17

Control Group: Never Leave Home without It

- Best practice is to hold out a control group for every deployment (action, offer, policy change, and so on).
- This should be a randomly selected subset of cases that are candidates to receive the action (promotion, campaign, and so on).
- With a control group, you can estimate the impact of the action, or the *incremental response*.
- Information about incremental response can save unnecessary marketing costs and improve the customer experience by eliminating unwanted communication.

18

Example

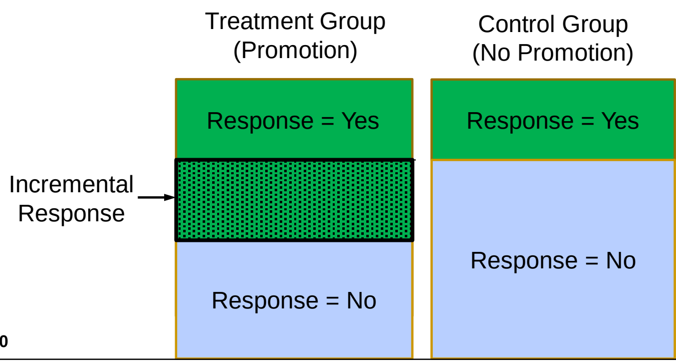
Which customers should be selected for a promotional offer to maximize net profit?

"I know half the money I spend on advertising is wasted, but I can never find out which half."
-John Wanamaker (1838-1922)

19

Example

- Treatment: the group that receives the offer. You assume that an incremental response exists, but you do not know for which customers.
- Control: the group that receives no offer. You assume that there is no known incremental response.



20

What Is an Incremental Response?

- A measure of the true effect of an action taken.
- The ***additional*** value that, in absence of the action, would not be realized.
- The most common application of incremental response modeling is in targeted marketing campaigns.
- Many other application areas are possible.

21

The idea behind incremental response modeling is to compare the difference in response rate between a control group (no marketing action applied) and a treatment group (marketing action applied).

Four Types

Persuadables: Respond only with offer

Loyal Customers: Offer irrelevant, Likely to respond

Lost Causes: Offer irrelevant, Unlikely to Respond

Do Not Disturbs: Less likely to respond with offer

RESPONSE:	YES	NO
Offer = YES	Persuadables + Sure things	Do Not Disturbs
Offer = NO	Sure things	Lost Causes + Do Not Disturbs

22

Incremental Response / Sales Modeling

Only binary target variable:

- Probability of response for both scenarios predicted

Both binary and interval target variables:

- Probability of response for both scenarios predicted
- Difference in amount of revenue for the two scenarios predicted

23

For a binary target variable (customer responding / not responding), incremental response modeling predicts the probability of response for two scenarios (treatment, control) for each customer.

For data sets with both a binary target variable (customer responding / not responding) and an interval target variable (amount of revenue if customer responds), incremental response modeling also predicts the difference in amount of revenue for the two scenarios (control and treatment).

Data Structure

- Treatment variable (required):
  - Binary
  - Does observation belong to treatment or control group?
- Response variable (required):
  - Binary target variable
  - Does the customer purchase or not?

24

Data Structure

- Outcome variable (optional):
  - Interval target variable
  - Amount of purchase
- Input variables

25

In the Incremental Response node, the data structure is as follows:

- A binary treatment variable is required. This variable tells whether an observation belongs to the treatment or control group.
- A binary target variable, called response variable, is required. This variable tells whether a customer purchases..
- An optional interval target variable, called *outcome variable*. This variable contains the amount of the purchase.
- Input variables.

Variable Selection

Built-in tool for variable selection based on Net Information Value (NIV):

- Measuring the difference in information value between control and treatment groups

26

The Incremental Response node in SAS Enterprise Miner has a built-in tool for variable selection. This tool is based on measuring the difference in information value, called Net Information Value (NIV), between the control and treatment groups for each of the potential input variables.

Variable Selection

- Important because incremental response modeling is relying on a double calculation
- Incremental effect = Treatment result – Control result
- Net Information Value (NIV) method by Larsen (2010):
  - intuitive
  - easy to implement
  - flexible
- Modification of the concepts of weight of evidence and information value

27

Weight of Evidence

- Y is binary (0/1) and Y=1 is the event
- Input variable X partitioned into I bins
- Weight of Evidence (WoE):

$$WOE_i = \log \frac{P(X = x_i | Y = 1)}{P(X = x_i | Y = 0)} \quad \text{for } i = 1,2,\ldots,I$$

28

The input variables are partitioned into an appropriate number of bins and then the WOE is calculated for each bin. The bins are often quantile-based with equal sizes. The WOE is analyzed over the bins to find out about the strength of the input variables.

Actually, the formula above means that X belongs to the set of elements in bin i.

Information Value

$$IV = \sum_i \left(P(X = x_i|Y = 1) - P(X = x_i|Y = 0)\right) \cdot WOE_i$$

29

The IV measures the correlation between the binary response variable and a particular input variable. If IV < 0.02, the variable is not predictive, and with an IV > 0.3, you call it a strong predictive variable as a heuristic (Siddiqi 2005).

Net WOE and Net IV

- Treatment = T
- Control = C

$$NWOE = \log \frac{P(X = x_i|Y = 1)_T / P(X = x_i|Y = 0)_T}{P(X = x_i|Y = 1)_C / P(X = x_i|Y = 0)_C}$$

$$NIV = \sum_i \left(P(X = x_i|Y = 1)_T P(X = x_i|Y = 0)_C - P(X = x_i|Y = 0)_T P(X = x_i|Y = 1)_C\right) \cdot NWOE_i$$

30

This is a log-odds ratio comparing the treatment response odds with the control response odds.

In SAS Enterprise Miner, the NIV statistic is used to rank the input variables. A fixed number or a percentage of the total number of variables is used to select the number of input variables. No rule for setting the threshold is available.

$$NIV = \sum_i (P(X = x_i|Y = 1)_T P(X = x_i|Y = 0)_C - P(X = x_i|Y = 0)_T P(X = x_i|Y = 1)_C) \cdot NWOE_i$$

Penalized Net Information Value

- Penalty for the difference in NWOE between training and validation data
- Variables that improve model stability are selected
- Difference:

$$\omega = |NWOE_{train} - NWOE_{valid}|$$

- PNIV = NIV – Penalty
- PNIV is automatically used in SAS Enterprise Miner if a validation data set exists.

31

The calculated penalty is then as follows:

$$Penalty = \sum_i | P(X = x_i|Y = 1)_T P(X = x_i|Y = 0)_C - P(X = x_i|Y = 0)_T P(X = x_i|Y = 1)_C| \cdot \omega_i$$

Difference Score Model

- The difference between the predictions for the treatment group model and the control group model is measured.
- The differences are called ***difference scores***.
- The difference scores are binned and ranked.
- The top subset (for example, 10%) is used as Persuadables.

32

Construction of the Predictive Model

Two alternatives exist:

- two separate models for the treatment and the control group
- one combined model for both groups

33

Two Separate Models

- Difference scores:

$$\widehat{DS}_{\,i} = (\hat{Y}_T - \hat{Y}_C)_{\,i} \;\; \text{for } i = 1,2,\dots,n$$

- Two alternatives to determine the Persuadables:
  - At least, the value of the difference score should be positive
  - More careful considerations with ranks:

$$\widehat{DS}_{\,(i)} = (\hat{Y}_T - \hat{Y}_C)_{\,(i)} \;\; \text{for } i = 1,2,\dots,n$$

34

Two separate models are built for the treatment and control groups:

$$\hat{Y}_T = X_T\hat{\beta}_T$$

$$\hat{Y}_C = X_C\hat{\beta}_C$$

The Combined Model

- Introduce an indicator variable T:
  - Treatment group: T=1
  - Control group: T=0
- Combined model:

$$Y = X\beta + T\gamma + (XT)\varphi$$

- Difference scores:

$$\widehat{DS}_i = \hat{Y}_T - \hat{Y}_C = \hat{\gamma} + X\hat{\varphi} \quad \text{for } i = 1,2, \ldots n$$

35

Stepwise Regression Model

- Variable selection techniques:
  - None (default), backward, forward, stepwise regression
- Selection criterion:
  - AIC, SBC, validation error, cross validation error

36

The Incremental Response node is based on a stepwise regression model. This means that you can choose between some standard automatic variable selection techniques: backward, forward, and stepwise regression. As default, no variable selection is performed. That is, all candidate input variables are used in the regression model. To select variables for the model, the following model selection criterion are offered: AIC, SBC, validation error, and cross validation error. As usual in stepwise regression, there are possibilities to tune the regression process by, for example, choosing stay and enter significance levels others than the standard levels. Another possibility to tune the regression model is to select to include two-way interactions in the candidate models in the variable selection process of the stepwise regression modeling.

Including two-way interactions means that all two-way interactions for class variables and all second-order polynomials for interval variables are included as possible input variables.

Stepwise Regression Model

- Tuning possibilities:
  - Stay significance level, enter significance level
  - Two-way interactions
- Combined model:
  - Treatment variable used as predictor

37

The Incremental Sales Model

- An incremental response model taking care of two targets.
- One target binary:
  - Response or not
  - Response model
- One target interval:
  - Amount of response
  - Outcome model
- Identify customers likely to spend incrementally ***with*** marketing action.

38

The Incremental Sales Model

- Selection bias due to unobserved target of amount if no response.
- Correction by use of the Heckman two-step method based on the inverse Mills ratio.

39

For more information about the Heckman method, see Heckman (1979).

Model Result for the Incremental Response Model

- The treatment and control data sets are ranked in descending order by the difference scores.
- The ranked observations are partitioned into bins.
- The average predicted interval response is predicted for each bin, once for the treatment model (t_ave) and once for the control model (c_ave).
- Predicted incremental response: t_ave - c_ave.

40

If we use 10 bins, we would use the deciles.

Model Result for the Incremental Sales Model

- Treated analogously to the incremental response model.
- Difference scores in sales prediction are considered.

41

Model Diagnostics

- The top deciles should have a high incremental rate.
- The bottom deciles should have a low incremental rate.
- At all deciles, the differences between predicted and observed incremental response rates should be small.

42

Model Diagnostics 1

Table: Incremental Response Model Diagnostics

Data Role	Predicted Treatment	Predicted Control	Observed Control	Observed Treatment	Predicted Increment	Observed Increment
Train	0.485987	0.168801	0.01042	0.60704	0.317186	0.596621
Train	0.357948	0.142841	0.12323	0.53333	0.215107	0.410101
Train	0.356049	0.186744	0.16895	0.42188	0.169304	0.252925
Train	0.347986	0.222134	0.24064	0.37891	0.125852	0.138265
Train	0.333149	0.245582	0.24299	0.36570	0.087567	0.122705
Train	0.31458	0.259651	0.27511	0.27431	0.054928	-.000795
Train	0.291697	0.267655	0.32474	0.30046	0.024042	-0.02428
Train	0.265571	0.273816	0.33333	0.23011	-0.00825	-0.10323
Train	0.241743	0.285697	0.36207	0.19298	-0.04395	-0.16909
Train	0.282256	0.465978	0.51579	0.16234	-0.18372	-0.35345
Validate	0.482879	0.172487	0.03306	0.52349	0.310391	0.490432
Validate	0.35953	0.149461	0.13725	0.50000	0.210068	0.362745
Validate	0.361889	0.196729	0.16970	0.46667	0.16516	0.29697
Validate	0.348372	0.225965	0.17763	0.40336	0.122408	0.22573

43

Model Diagnostics 2

Table: Incremental Outcome Model Diagnostics

Data Role	Predicted Treatment	Predicted Control	Observed Control	Observed Treatment	Predicted Increment	Observed Increment
Train	119.51	24.8766	4.033451	140.0723	94.63345	136.0389
Train	59.30056	16.33156	9.177628	93.43466	42.969	84.25704
Train	56.42121	25.2598	19.15788	63.17019	31.16141	44.01231
Train	54.60632	32.47519	25.91406	36.21836	22.13112	10.30431
Train	51.96969	38.20378	41.05792	30.34887	13.76591	-10.7091
Train	48.06628	42.51167	47.60998	34.34119	5.554608	-13.2688
Train	42.21048	45.19649	47.21389	33.85993	-2.98601	-13.354
Train	34.59262	46.16476	63.15998	36.81531	-11.5721	-26.3447
Train	26.48922	47.88364	76.13358	29.24344	-21.3944	-46.8901
Train	24.27894	69.42282	106.9503	26.46063	-45.1439	-80.4897
Validate	119.5379	25.31658	6.53354	89.79183	94.22128	83.25829
Validate	60.51594	18.8233	9.399437	73.84595	41.69265	64.44652
Validate	57.65127	26.83521	15.25732	29.69147	30.81606	14.43415
Validate	54.54719	32.56887	25.55356	57.99962	21.97831	32.44605

44

Incremental Revenue Analysis

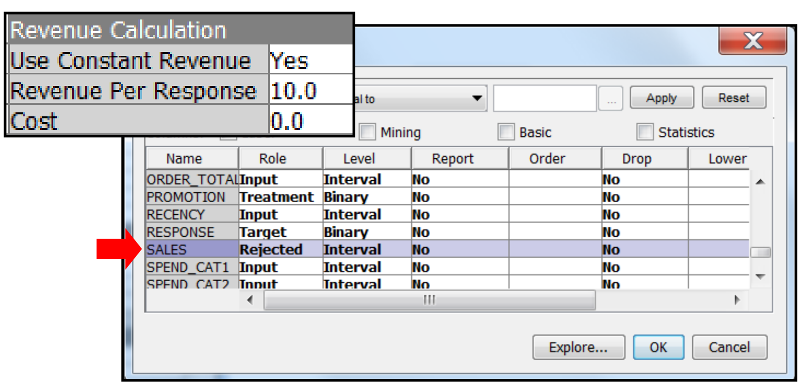
- Constant revenue and constant cost
- Variable revenue and constant cost
- Variable revenue and variable cost

45

The default setting is that the expected revenue is estimated for each customer from the model. You can override this by specifying that a constant revenue shall be used for each observation. To do this, set the property **Use Constant Revenue** to *Yes* and specify the constant revenue in the property **Revenue Per Response** (which has the default value *10*). The property **Cost** specifies the cost of the marketing action for each customer contact. If the incremental revenue is greater than the direct cost, a customer is said to be profitable. The property **Cost** has the default value *0*.

Constant Revenue and Constant Cost

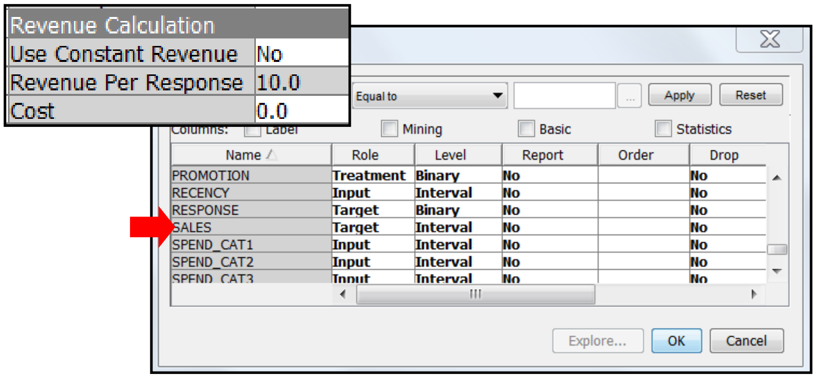
- No use of a cost variable
- No use of an interval target variable (but the property **Use Constant Revenue** would override it)



46

Variable Revenue and Constant Cost

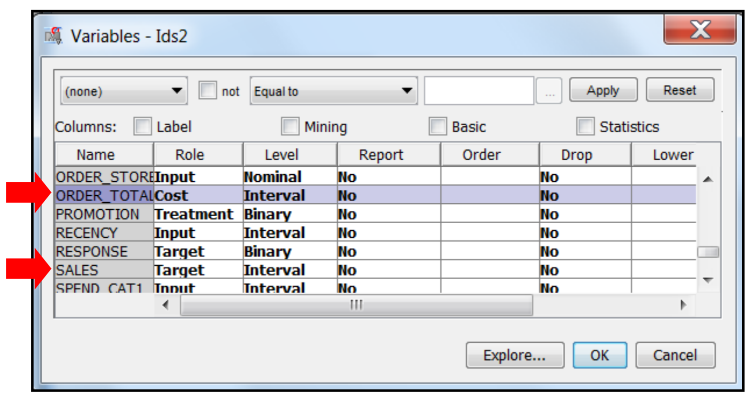
- No use of a cost variable
- Use of an interval target variable



47

Variable Revenue and Variable Cost

Set the role of the cost variable to **Cost** in the metadata definition.



48

Further information about incremental response modeling can be found in Lee et al (2013) and Larsen (2010).

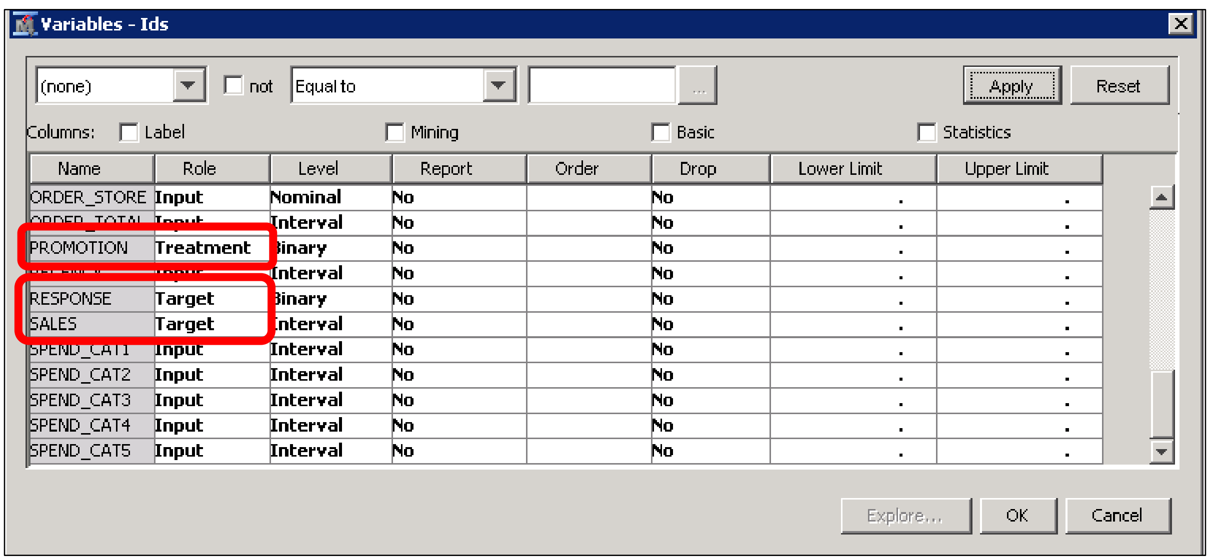
Incremental Response Modeling

This demonstration illustrates incremental response modeling. The data set **DMRETAIL** contains the binary target variable **RESPONSE** and the interval target **SALES**. Furthermore, the treatment variable **PROMOTION** is contained as well as some input variables.

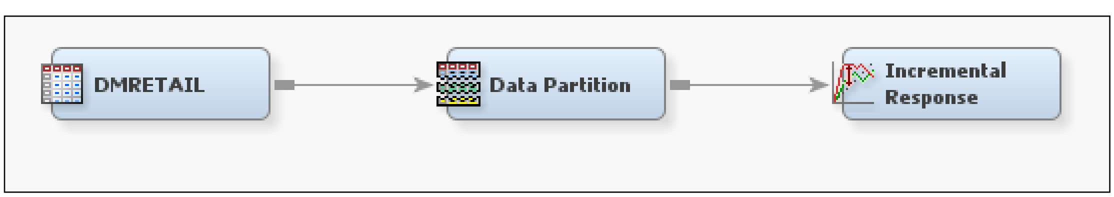
Data Source Configuration

1. Create a new diagram. Name it **Incremental Response Modeling**.
2. Create a new data source for the data set **DMRETAIL** in the **XDS** library.
3. Use the Advanced Metadata Advisor, and change the metadata roles for the following variables. Set **Promotion** to a **Treatment** role, and set **Response** and **Sales** to **Target** roles. Accept all the remaining defaults as you complete the data source creation.
4. Drag the data set **DMRETAIL** into the diagram workspace.
5. Right-click the data source and select **Edit Variables**.

   Look at the Column metadata. The role of the variables **RESPONSE** and **SALES** is set to **Target**. The role of the variable **PROMOTION** is set to **Treatment**.



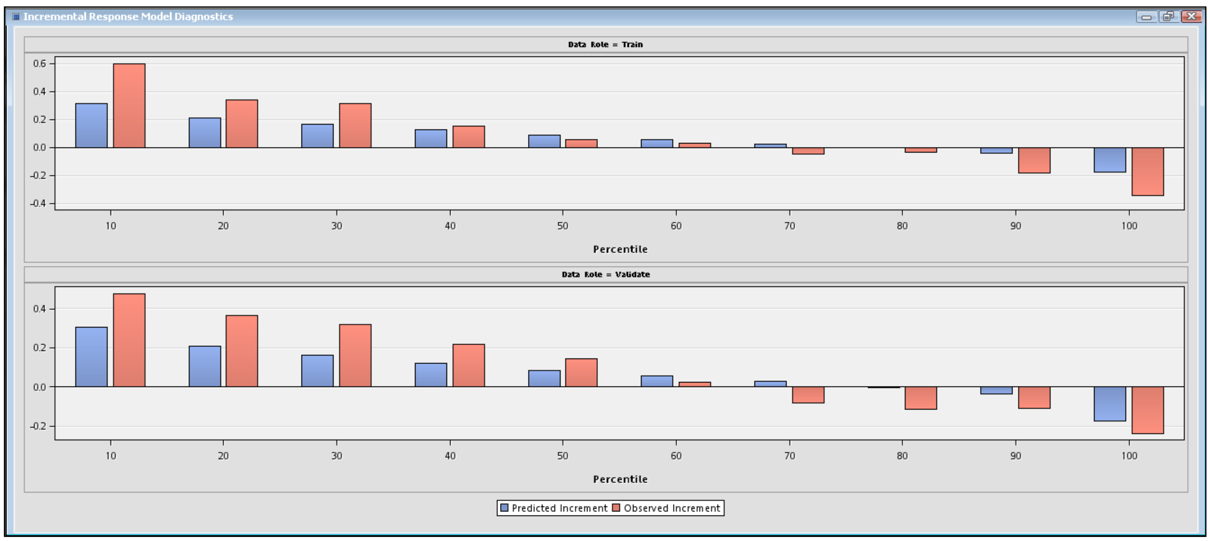
6. Click **OK** to exit the Column metadata.
7. Drag a **Data Partition** node from the Sample tab and an **Incremental Response** node from the Applications tab onto the Incremental Response Modeling workspace and connect the three nodes as follows:



8. Click the **Data Partition** node in the diagram to select it. Modify the Data Set Allocations properties as follows:
   - Set the **Training** property to **70.0**.
   - Set the **Validation** property to **30.0**.
   - Set the **Test** property to **0.0**.
9. Right-click the **Incremental Response** node in the diagram and select **Run**.
10. Select **Results** in the Run Status dialog box. The Results – Node: Incremental Response Diagram opens.

Results of the Response Model

11. Examine the Incremental Response Model Diagnostics window.



This plot shows incremental response by deciles. There is a declining pattern across deciles to the predicted incremental response rate (blue). There is a similar pattern to the observed incremental response rate (red). The top 10% of customers in the training data show an observed incremental response rate of 59.7%. The expected incremental response rate at this decile is 31.35%. These rates are much higher than the average incremental rate of 7.57%, which can be seen from the **Response Outcome Summary** plot discussed later.

This means that the model has the potential to pick up a significant portion of customers who are the Persuadables. It provides a guideline for targeting this group, justifies the expense of the marketing campaign, and shows the ability to minimize the risk of negative responses to the promotion.

12. Select **View** ⇨ **Table**. The Incremental Response Model Diagnostics table appears.

Table: Incremental Response Model Diagnostics

Data Role	Predicted Treatment	Predicted Control	Observed Control	Observed Treatment	Predicted Increment	Observed Increment	Cumulative Observed Increment	Cumulative Predicted Increment	Percentile
Train	0.474417	0.160884	0.022508	0.619497	0.313533	0.596989	0.596989	0.313533	10
Train	0.355064	0.144162	0.143426	0.480315	0.210903	0.336889	0.466939	0.262218	20
Train	0.356197	0.191084	0.135831	0.445545	0.165112	0.309713	0.41453	0.22985	30
Train	0.347965	0.22426	0.234807	0.389513	0.123705	0.154706	0.349574	0.203313	40
Train	0.334187	0.246241	0.266904	0.324713	0.087946	0.057809	0.291221	0.18024	50
Train	0.314876	0.259386	0.293103	0.324937	0.05549	0.031834	0.24799	0.159448	60
Train	0.293226	0.268273	0.308458	0.257009	0.024954	-0.05145	0.205213	0.140235	70
Train	0.268763	0.273453	0.285714	0.247289	-0.00469	-0.03843	0.174758	0.122119	80
Train	0.245841	0.285909	0.381215	0.196429	-0.04007	-0.18479	0.134809	0.104098	90
Train	0.285916	0.462008	0.514344	0.167785	-0.17609	-0.34656	0.086672	0.076079	100
Validate	0.471262	0.166718	0.078125	0.556338	0.304544	0.478213	0.478213	0.304544	10
Validate	0.359112	0.14863	0.111111	0.47619	0.210482	0.365079	0.421646	0.257513	20
Validate	0.356829	0.19317	0.179775	0.5	0.163659	0.320225	0.387839	0.226228	30
Validate	0.348441	0.227862	0.165517	0.384	0.12058	0.218483	0.3455	0.199816	40
Validate	0.333326	0.247272	0.233333	0.38	0.086054	0.146667	0.305733	0.177064	50
Validate	0.315135	0.259857	0.317308	0.343373	0.055278	0.026066	0.259122	0.156766	60
Validate	0.294907	0.267464	0.333333	0.251366	0.027443	-0.08197	0.210395	0.138291	70
Validate	0.273569	0.274216	0.329114	0.21466	-.000647	-0.11445	0.169789	0.120924	80
Validate	0.252495	0.289031	0.306818	0.197802	-0.03654	-0.10902	0.138811	0.103429	90
Validate	0.288379	0.462872	0.464455	0.229508	-0.17449	-0.23495	0.101435	0.075636	100

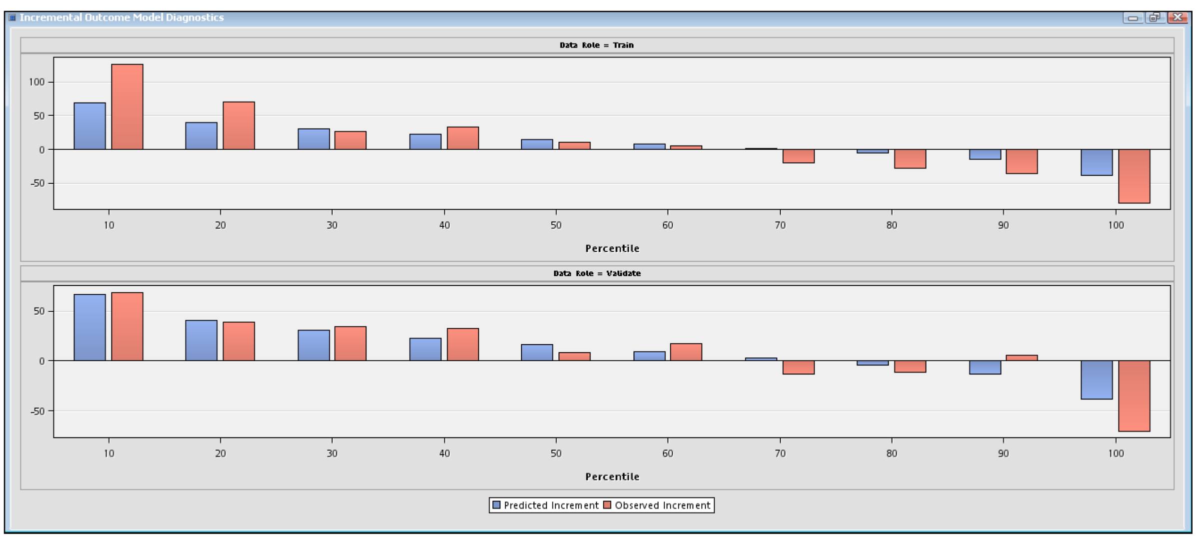
This is the table that the Incremental Response Model Diagnostics plot is based on. For example, in the top decile for training data, we see that the predicted increment (0.313553) equals the predicted treatment (0.474417) minus the predicted control (0.160884). Further, the observed increment (0.596989) equals the observed treatment (0.619497) minus the observed control (0.022508).

The table also displays the values from the Treatment Response Model Diagnostics and the Control Response Model Diagnostics plots, which are shown next.

13. Close the Incremental Response Model Diagnostics table.

Results of the Outcome Model

14. Examine the Incremental Outcome Model Diagnostics window.



This plot displays model diagnostics from the sales perspective. When a customer makes a purchase during the promotion period, the amount is used as a target variable. A two-stage model estimates the outcome. The top decile (training data) shows an observed incremental outcome (of sales) of $126.64, and the expected incremental outcome for this group is $69.10. These rates are both much higher than the average incremental outcome of $3.82, which can be seen from the Response Outcome Summary plot discussed later.

With these results, we can identify customers who display the maximal increment in sales from a much larger group that includes customers who make a purchase regardless of the marketing promotion being offered.

15. Select **View** ⇨ **Table**. The Incremental Outcome Model Diagnostics table appears.

Table: Incremental Outcome Model Diagnostics

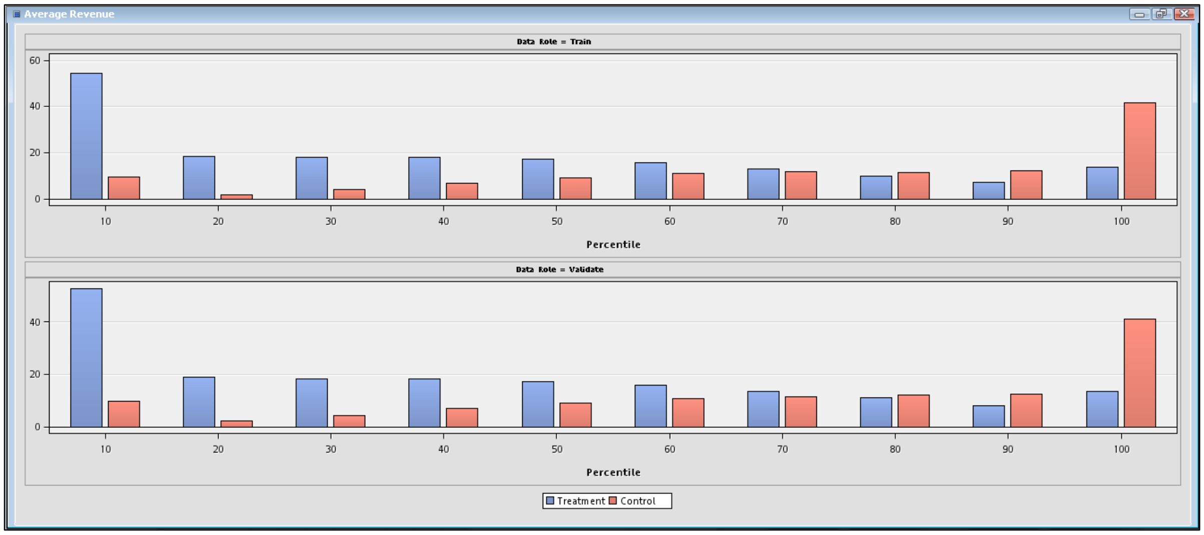
Data Role	Predicted Treatment	Predicted Control	Observed Control	Observed Treatment	Predicted Increment	Observed Increment	Cumulativ e Observed Increment	Cumulativ e Predicted Increment	Percentile
Train	91.81694	22.7212	3.202761	129.8398	69.09574	126.637	126.637	69.09574	10
Train	60.21919	19.64386	10.38884	80.38041	40.57533	69.99157	98.31429	54.83554	20
Train	57.60081	27.0508	21.3741	48.65896	30.55001	27.28486	74.63781	46.74036	30
Train	55.4475	32.86548	27.68892	61.53113	22.58202	33.84221	64.43891	40.70078	40
Train	52.56422	37.54267	39.21269	49.90357	15.02155	10.69088	53.6893	35.56493	50
Train	49.38149	41.15621	34.70161	39.99415	8.225285	5.292532	45.62317	31.00832	60
Train	44.89814	43.38089	57.11823	37.18226	1.517251	-19.936	36.25758	26.79531	70
Train	39.53536	45.10624	59.03249	31.1955	-5.57088	-27.837	28.24576	22.74954	80
Train	31.3057	45.3708	67.27773	31.84267	-14.0651	-35.4351	21.17011	18.65902	90
Train	34.8271	73.02681	97.15932	18.73677	-38.1997	-78.4225	11.21085	12.97315	100
Validate	90.54792	23.5306	10.49939	79.44898	67.01732	68.94959	68.94959	67.01732	10
Validate	62.39381	21.7753	12.56848	51.77039	40.61852	39.20191	54.07575	53.81792	20
Validate	58.62647	27.89107	20.19052	54.92778	30.7354	34.73727	47.62959	46.12374	30
Validate	55.71768	32.74674	15.81719	48.66505	22.97094	32.84786	43.93416	40.33554	40
Validate	53.20126	37.12359	41.81914	50.47232	16.07766	8.653181	36.87796	35.48397	50
Validate	50.03576	40.73885	24.17115	41.27812	9.29691	17.10697	33.5828	31.11946	60
Validate	46.50925	43.82915	59.42589	46.34436	2.680099	-13.0815	26.91646	27.05669	70
Validate	41.06227	45.39215	40.96541	29.26356	-4.32988	-11.7019	22.08917	23.13337	80
Validate	31.98132	44.71709	36.98484	43.13208	-12.7358	6.14724	20.31785	19.14791	90
Validate	34.65902	72.99827	94.54187	24.58809	-38.3393	-69.9538	11.29068	13.39919	100

This is the table that the Incremental Outcome Model Diagnostics plot is based on.

16. Close the Incremental Outcome Model Diagnostics table.

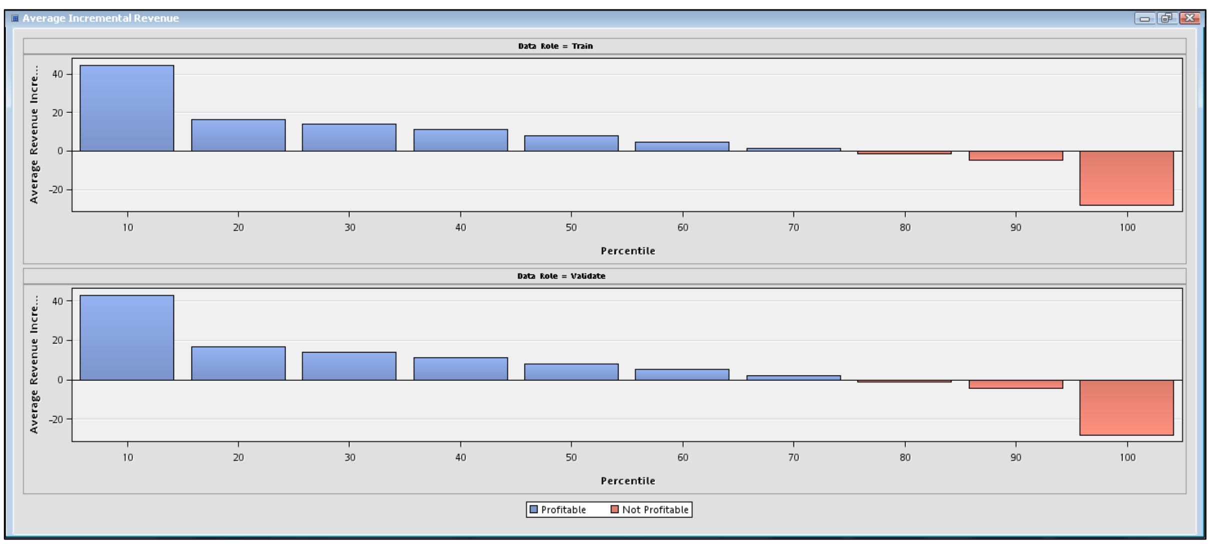
Results: Revenue Plots

17. Examine the Average Revenue window.



This plot compares the average revenue for treatment versus control groups by deciles. The most profitable deciles are on the left, and the least profitable are on the right.

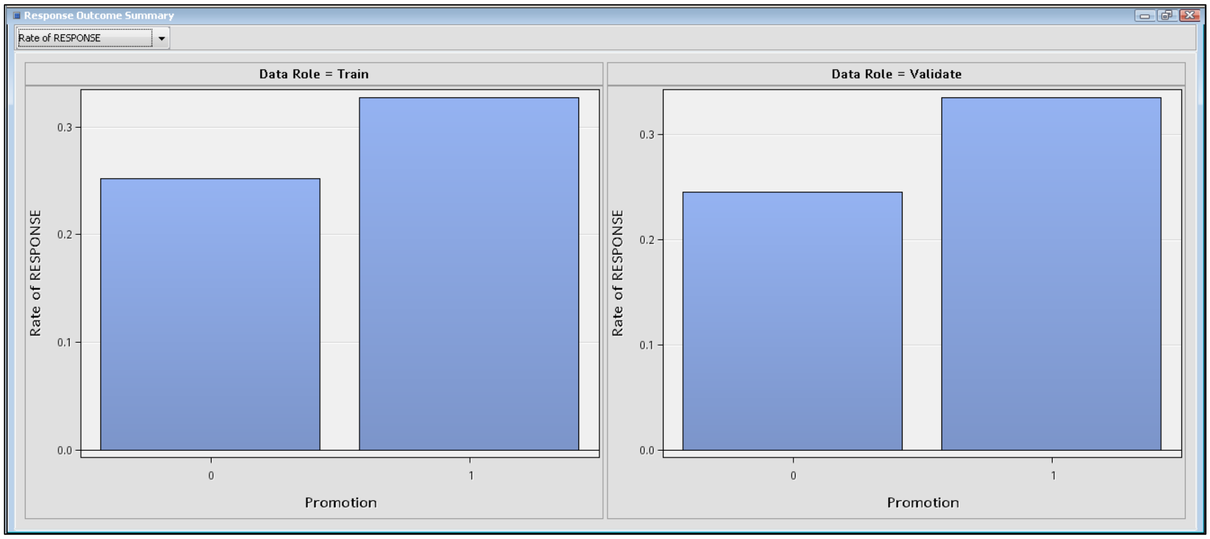
18. Examine the Average Incremental Revenue window.



A customer is profitable when the incremental gross revenue is greater than the contact cost. Here, the first 60% of customers are profitable because they show positive net revenue increments. For example, the top decile of customers show an expected revenue increment of $44.83 (training data). A negative effect of the promotion on “Do Not Disturb” customers is shown by negative values. This group of customers should be omitted from the promotion.

Results: Other Plots

19. Examine the Response Outcome Summary window.



This plot shows the response rate of purchases on the vertical axis for the control and treatment groups. The purchase response rate for the training and validation data seems to be nearly identical, which indicates little variation in the characteristics of the two data sets.

The control group (no treatment) shows a 25.21% response on training data. This means about 1 in 4 customers made a purchase with no marketing treatment. The group that did receive a marketing treatment shows a response rate of 32.78%, so about 1 in 3 made a purchase. (Some customers in the treatment group made a purchase because of the marketing treatment, whereas others would have made the purchase regardless. The customers in this group are not homogeneous.) We can calculate the response rate of the customers in the treatment group that were “true” responders to the marketing treatment by taking the difference between the response rates of the control and treatment groups.

So here, approximately 7.57% (32.78%-25.21%) of the customers purchased because they received the marketing treatment. This value, 7.57%, represents the average incremental response rate.

This chart can also show results for Average Sales by changing the pull-down selector in the upper right corner. The chart for Average Sales (not displayed here), shows the average sales for the treatment group (**PROMOTION** = 1) on training data is $157.82, and the average sales for the control group (**PROMOTION** = 0) is $154.00. The resulting difference of $3.82 represents the average incremental sales rate.

20. Select **View** ⇨ **Model** ⇨ **Selected Variables by NIV**. The Selected Variables by NIV window appears.

Selected Variables by NIV

Variable Name	Penalized Net Information Value	Net Information Value	Rank Percentile	Selection	Label
CUST_TENURE	51.21812	122.0471	3.125	Yes	Customer ...
ORDER_JAN	18.41162	20.57142	6.25	Yes	Order in Jan
SPEND_CAT4	6.049328	35.09345	9.375	Yes	Spend in ...
SPEND_CAT1	4.969308	25.51107	12.5	Yes	Spend in ...
ORDER_SEP	2.562611	4.890913	15.625	Yes	Order in S...
ORDER_NOV	2.56162	4.532496	18.75	Yes	Order in Nov
ITEM_ONLINE	2.028988	13.768	21.875	Yes	Online Ite...
ORDER_MAY	1.43928	6.118613	25	Yes	Order in May
ITEM_STORE	0.424843	8.297627	28.125	Yes	Store Items
MEMBERSHIP	-3.34E-6	1.651E-5	31.25	Yes	Membersh...
ORDER_PHONE	-0.00745	0.2711	34.375	Yes	Phone Ord...
FREQUENT_BUY...	-0.43011	0.174471	37.5	Yes	Frequent b...
ORDER_JUN	-1.62333	2.01516	40.625	Yes	Order in Jun
ORDER_TOTAL	-1.78282	29.04813	43.75	Yes	Total Orders
AVE_INTV_PURC...	-2.42918	51.24895	46.875	Yes	Average p...
ORDER_DEC	-2.82268	2.447854	50	Yes	Order in D...
RECENCY	-4.73183	28.68042	53.125	No	Recency
ITEM_PHONE	-5.27482	4.24515	56.25	No	Phone Ite...
ORDER_STORE	-5.7664	10.38299	59.375	No	Store Orde...
SPEND_CAT2	-7.51476	35.92749	62.5	No	Spend in ...
ORDER_FEB	-9.35322	3.132113	65.625	No	Order in Feb
ITEM_TOTAL	-9.65036	24.39622	68.75	No	Totoal Items
SPEND_CAT5	-10.1308	19.39982	71.875	No	Spend in ...
LAST_YEAR_SPE...	-10.4993	32.8864	75	No	Totoal Spe...
ORDER_ONLINE	-10.6958	32.59136	78.125	No	Online Ord...
ORDER_APR	-13.1355	72.77251	81.25	No	Order in Apr
ORDER_MAR	-13.2223	13.3858	84.375	No	Order in Mar
ONSALE_ITEM	-13.3768	39.84524	87.5	No	Onsale
ORDER_AUG	-13.7491	4.698188	90.625	No	Order in Aug
SPEND_CAT3	-16.0794	18.63527	93.75	No	Spend in ...
ORDER_OCT	-24.5443	22.22052	96.875	No	Order in Oct
ORDER_JUL	-27.4623	42.69237	100	No	Order in July

Setting the Prescreen Variables property to **Yes** before running the model results in the Selected Variables Table displaying the top 50% of input variables ranked by net information value (NIV). The NIV score indicates the variables that have the strongest correlation to the model response. NIV is the difference in information values between the treatment and control groups for each input variable. The Rank Percentage Cutoff property controls the proportion of variables to be displayed in the table.

21. Close the Selected Variables by NIV table and close the Results window.

End of Demonstration

Appendix A References

A.1 References

Chase, Charles W. 2009. *Demand-Driven Forecasting: A Structured Approach to Forecasting*. Wiley.

Heckman, J.J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*, 46:931-959.

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

Larsen, K. 2010. *Net Lift Models: Optimizing the Impact of Your Marketing Efforts*. SAS Course Notes. Cary, NC: SAS Institute Inc.

Lee, T., D. Duling, and D. Latour. 2009. "Predictive Models Based on Reduced Input Space That Uses Rejected Variables." *Proceedings of the SAS Global Forum 2009 Conference*, Paper 111-2009, Cary, NC: SAS Institute Inc.

Lee, T., R. Zhang, X. Meng, and L. Ryan. 2013. "Incremental Response Modeling using SAS Enterprise Miner." *Proceedings of the SAS Global Forum 2013 Conference,* Paper 096-2013, Cary, NC: SAS Institute Inc.