University of Southern California  
MARSHALL SCHOOL OF BUSINESS  
Tentative Syllabus - Summer 2011  
Course Guidelines & Syllabus

IOM 528 – DATA WAREHOUSING, BUSINESS INTELLIGENCE AND DATA MINING -16274 (Wednesday 6:30-9:30 p.m. in OCC.) and two Saturdays (6/4/11 and 6/18/11 - 6:30-9:30 p.m. (time may change)) – Note: The final exam is Wednesday July 13th. I may add one more class session.

Instructor: Dr. Arif Ansari  
Office: BRI 401 R (Bridge Hall 401R)  
Office Hours: Wednesday 5:30-6:30 p.m. and by appointment  
Office phone: (213) 821-5521  
Email: aansari@usc.edu  
Emergency Contact number: 213-740-0172

COURSE OBJECTIVES

• To develop an understanding of the various concepts and tools behind data warehousing and mining data for business intelligence.
• To develop quantitative skills pertinent to the analysis of data from huge corporate data warehouses.
• To develop industry level data mining skills using SAS enterprise Miner and desktop level data mining skills using SAS JMP software.
• To develop Data warehouse architecture from Business point of view.
• To develop new business strategies based on DW, BI and DM.
• To develop Critical thinking and Modeling Skill Set.

• This year I will emphasis on New Direct Marketing methods using Data Mining. See the attached Content section from the Book “The New Direct Marketing: How to Implement A Profit-Driven Database Marketing Strategy”

COURSE STRUCTURE

➢ 60% of the class will be focused on Data Mining  
➢ 20% on Business Intelligence  
➢ 20% on Data warehousing.

Overview:

This course is about how companies apply two new technologies, data warehousing (DW) and data mining (DM, including business intelligence, BI) to empower their employees, and build and manage a customer-centric business model. Besides learning the strategic role DW and DM plays in an enterprise, you will also get a close-up look at DW and DM by working on cases and gaining hands-on experience using software tools.
Students taking this class will get an overview of the technologies of DW and BI/DM from a managerial perspective.

Fortune 500 companies such as American Express, Wells Fargo and Wal-mart have accumulated a great deal of data from their day to day business. Data warehouse is the technology that integrates the data collected from various sources that include transaction processing systems and e-commerce data collecting systems. Collecting and integrating data is just the first step. What are really critical are information, knowledge and insight. So the questions are what is the utility of the data? How can one use data in managing customer relationship and empowering employees? How can one uncover patterns and relationships hidden in organizational databases? These issues are addressed by a fast growing body of research and applications, broadly known as business intelligence and Data Mining (BI/DM). These technologies draw their strengths from the fields of information technology, statistics, machine learning and artificial intelligence.

In summary, managers need to understand the strategic values of their company's information assets. DW, BI and DM are cornerstones

**COURSE GOALS:**

After taking this class, students should be able to:

- Understand the basic terms that are used in DW, On Line Analytical Process (OLAP), BI and DM
- Communicate to Information Technology workers their business perspective in terms of the language of DW and DM
- Choose appropriate tools for specific purposes of storing, integrating and analyzing data (business consideration, and technical consideration).
- Use tools provided in class to perform simulated tasks in warehousing data.
- Use Enterprise Miner and JMP to perform DM activities on moderately large data sets.
- Articulate and present the results of their analyses and the business implications of these results
- Gain inference from your analysis, from Business and Statistical point of view.

1) In Data Mining you will develop in-depth skill set to do desktop Data Mining.
2) In Data Warehousing part you will learn, why companies need DW, advantages of DW and how to create a DW schema that an executive will understand, I will not teach the hands-on programming part, programming part is made available through Teradata student network and you can learn on your own.
3) In Business Intelligence you will learn what current BI can do, how to develop the requirements of a BI system for a company. I will not teach the hands-on programming part, programming part is made available through Teradata student network and you can learn on your own.
4) You will learn how the 3 parts are interconnected and integrated to form the basis of corporate knowledge system. How to leverage them to convert your company to almost real-time corporation.
Structure of lectures:

IOM 528 will be organized in a way that includes some combination of the following: lectures, case-based class discussion, group project, computer lab work, and guest lectures.

This class is designed in such a way that only limited mathematical and statistical (Descriptive Statistics, Hypothesis testing and Regression) background is required. I will give a brief review on the above mentioned topics. Learning and understanding underlying DW concepts, studying cases, applying DM ideas and methods to business data, and communicating ideas and solutions will be our main theme. Technical details of selected DM methods will be discussed. Students are expected to use Data Mining software for various business applications.

COURSE REQUIREMENTS

1. **Class Attendance & Participation.** I strongly suggest that you attend all classes. I strongly encourage, as well as expect, questions during the lectures. I am always accessible by e-mail, and will be more than happy to speak with you before or after class or during office hours. I will keep you updated about a course TA/grader who will have office hours to assist you.

2. **Homework.** I feel that written homework is crucial to the learning process, I will assign homework problems. Each assignment is due at the beginning of class on the date indicated. The due dates are listed in the Course Schedule. So late assignments will not be accepted. You should indicate clearly how you obtained your solution and make sure that you have given a valid explanation. Computer printouts without commentary are not adequate for full credit. Working with other students on homework problems will be allowed, but direct copying is **NOT** allowed. You are required to prepare and submit your own homework answers. Your homework must contain your own printout(s) if values off a printout are used to answer a homework question. **NO PRINTOUT = NO CREDIT.** No photo copies of computer printouts are allowed. The solutions to the homework problems will be available through Blackboard.

Homework assignments will be distributed via blackboard (Blackboard.usc.edu). If you believe that an error has been made in the grading of your homework you may ask to have it regarded. Please be specific about the problem. If you are still concerned after this process you may come and see me.

If you do not agree with the TAs grading, you may appeal your solution to me. Note, however, that I will review your entire assignment and will include in my assessment of your grade your oral arguments as well. **I am a tougher grader than the TA, so be prepared when you see me. I reserve the right to adjust your grade up or down as I see fit.**
3. **Mini Cases.** We will analyze mini cases during the semester. The mini cases will be evaluated and will be part of the class work points.

4. **Group Project.** I strongly believe the students learn the most during the project. The Group will consist of 3 or 4 students. Learning to work in teams is essential and to get different perspective and will greatly enhance your learning. The project points will be based on the following criteria:

   a) **Selection of the project and submitting the proposal** - 5%
   b) **Submission of the Data set and descriptive statistics** - 10%.
   c) **Preliminary report with Analysis and further direction of the project** – 25%
   d) **Final Project report and Presentation** – 60%

   A word document of the Final project report is required as well as a hard copy of the Final project report. The groups will also do peer evaluation of the group. The final report will include an *Executive Summary* write-up that translates the quantitative findings into a real-world analysis. You will be expected to participate in the discussion of your project during the semester to share your methodologies and interesting findings.

5. **Midterm and Final Exam.** The midterm will also take place at the beginning of class approximately one hour and 45 minutes. You may bring **two sheets (four pages)** containing formulas, definitions etc., to the midterm except solved problems and solved multiple choice questions. For the final, you may bring **four sheets** (eight pages) containing formulas, definitions etc., except solved problems and solved multiple choice questions. **No make-ups of mid-term or the final will be given.** You will receive a grade of zero for each missed exam unless you have a written excuse from your doctor or the professor. In case of emergency or approved absence, the professor may decide to give a make-up exam or redistribute the points.

There will be 1 midterm and 1 final exam. They are close-book.

<table>
<thead>
<tr>
<th>Grading</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>20</td>
</tr>
<tr>
<td>Class Work</td>
<td>20</td>
</tr>
<tr>
<td>Midterm</td>
<td>30</td>
</tr>
<tr>
<td>Final</td>
<td>15</td>
</tr>
<tr>
<td>Project</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

**Review Session.** There will be a review session before the final exam.
Course Materials. The following items will be necessary for completion of reading assignments and homework and successful completion of the course.

1. **Online Resources**

   ➢ **Sign up with Teradata University Network**

   Teradata University Network (www.TeradataUniversityNetwork.com) is a free learning portal designed to help faculty to teach, learn, and connect with others in the fields of data warehousing, DSS/Business Intelligence, and database.

   Teradata University offers web-based courses and related web sites on data warehousing, DSS/BI and database. They have a library of Teradata white papers. Students can become Teradata certified. We will use their material and software in the class particularly for the Business Intelligence and Data warehousing part of the class.

   Students register for www.TeradataStudentNetwork.com and login using the current password: ***** (last year’s password: *****)

   ➢ **Sign up with DMreview.com**

   DMreview is another portal with many resources for DM/BI/DW professionals.

2. **Text Books and Class notes**

   ➢ The first book is a standard book for Data Mining, the book talks about the various techniques and it is written from computer science perspective. *(Recommended)*


     Note: The book is written from a Computer Science and it will help you to understand the data mining techniques but it is does not have real world business application – Buy the book if you want to understand Data Mining Algorithms. My PowerPoint slides will cover the data mining topics but not in depth.

   ➢ The second book is a standard book for Direct Marketing, the book talks about the various techniques and it is written from Marketing perspective. *(Recommended)*

     The New Direct Marketing: How to Implement A Profit-Driven Database Marketing Strategy
     Hardcover: 736 pages
     Publisher: McGraw-Hill; 3 edition
     ISBN-10: 0070580561
The third book is from SAS – The world’s leading Data mining software company. This book introduces you to industry level Data mining software – SAS Enterprise Miner. (Recommended)


website: www.sas.com

The fourth book is from SAS (I am currently reviewing this detailed technical book). This book introduces you to industry level Data mining software – SAS Enterprise Miner. (Recommended)

Data Mining Using SAS Enterprise Miner by Randall Matignon, John Wiley and Sons publishing

website: www.sasenterpriseminer.com

Class notes. Class notes for this class will be available on blackboard. You should familiarize yourself with these notes before they are covered in class.

Recommended (If you want to concentrate on Data Warehousing)


Recommended (If you want to concentrate on Analytics, Business Intelligence and Forecasting)

1. Information Dashboard Design: The Effective Visual Communication of Data (Paperback) –by Stephen Few (Author)

http://www.amazon.com/Information-Dashboard-Design-Effective-Communication/dp/0596100167/ref=sr_1_1?ie=UTF8&s=books&qid=1240440281&sr=1-1

2. Balanced Scorecards & Operational Dashboards with Microsoft Excel (Paperback) by Ron Person (Author)

http://www.amazon.com/Balanced-Scorecards-Operational-Dashboards-Microsoft/dp/0470386819/ref=sr_1_1?ie=UTF8&s=books&qid=1240440389&sr=1-1

3. Competing on Analytics: The New Science of Winning (Hardcover) by Thomas H. Davenport (Author), Jeanne G. Harris (Author)

http://www.amazon.com/Competing-Analytics-New-Science-Winning/dp/1422103323/ref=sr_1_1?ie=UTF8&s=books&qid=1240442477&sr=1-1
Important dates: (Refer to Schedule of classes for up-to-date information)

Midterm exam:
June 15, 2011, Wednesday 6:30-9:30 P.M.

Final Exams:
July 13, 2011, Wednesday 6:30 P.M.-9:30 P.M.

Academic Integrity. Academic dishonesty of any type will not be tolerated in this class. Students who find this statement ambiguous should consult the Student Conduct Code, page 83, of the USC SCampus handbook.

A comment about writing the assignments up individually and working in teams: You can work together in teams to discuss the problems and concepts. However, you are required to write up the assignments individually. This means that all the words in your assignments are your own, and you generate all of your own computer output and graphs.

Now, while correct solutions will have very similar or even the same computer output, no two answers should be phrased the same way. If I find two or more assignments that are highly similar, I will at a minimum give the homework a zero, and may refer the incident to the Dean. Do not test me on this policy.

STUDENTS WITH DISABILITIES

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. DSP is located in STU 301 and is open 8:30 am - 5:00 pm, Monday through Friday. The phone number for DSP is 213 740-0776.

Tentative Schedule:

- The course will start with Data Mining. The Data Mining part of the class will be quantitative and the following topics will be covered in it.
  1. Standard Data Mining techniques:
     a. Classification
     b. Clustering
     c. Association
d. Prediction
e. Text Mining, Link Analysis, Visual Data mining (if time permits)

Using various appropriate techniques,
   i) Bayesian Estimation
   ii) Neural Networks
   iii) Decision Tree
   iv) Similarity Measures
   v) Other techniques like Boosting, Bagging (if time permits)

2. Statistical Model Building using Regression and Logistic Regression. Depending on the project other topics may be covered.

- The second part of the course will be Business Intelligence Software. You will be introduced to software used as Business Intelligence software.

  - We will discuss Forecasting methods to identify trends, seasonality etc.,

- The third part of the course will be Data Warehousing. You will be introduced to Data Warehousing from business perspective, how to create Data Warehouse Architecture.

*Note: This year, I may use Enterprise Miner or another software*
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Reading from textbooks</th>
<th>Reading from Class notes</th>
<th>Due/Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 May</td>
<td>Introduction to Classification - Bayesian Classification and Distance Based Algorithms / Personalization and Customization/</td>
<td>JM 1-26, JM 36-40, JM 359-362, JM372-375, JM285-290, JM 310-318, JM 347-350, SAS 39-67</td>
<td>Dr. Ansari Notes</td>
<td>Bring in Laptop with JMP software loaded</td>
</tr>
<tr>
<td>25 May</td>
<td>Classification Methods – Decision Tree Based Methods</td>
<td>JM 291-306, SAS 19-36</td>
<td>Dr. Ansari Notes</td>
<td>Bring in Laptop Group List Due</td>
</tr>
<tr>
<td>1 June</td>
<td>Review of Statistical Model Building and Introduction to Enterprise Miner</td>
<td>SAS 39-67, SAS 67-81</td>
<td>Dr. Ansari Notes</td>
<td>Turn in your Class work</td>
</tr>
<tr>
<td>4 June</td>
<td>Visual Data Mining/Logistic Regression/New methods</td>
<td>JM 358-359, JM 327-336, JM 358-359, JM 327-336</td>
<td>Dr. Ansari Notes</td>
<td>Bring in Laptop HW1</td>
</tr>
<tr>
<td>8 June</td>
<td>Logistic Regression / Neural Network/Clustering and Association</td>
<td>JM 384-414, JM 227-234, JM 384-414, JM 227-234, SAS 91-104, SAS 105-109</td>
<td>Dr. Ansari Notes</td>
<td>Bring in Laptop</td>
</tr>
<tr>
<td>15 June</td>
<td>Midterm/Business Intelligence</td>
<td>TUN: Microstrategy BI Information/Continental Airlines Takes Off with Real-time Business Intelligence</td>
<td>Dr. Ansari Notes</td>
<td>Midterm/ Turn in your Class work4</td>
</tr>
<tr>
<td>18 June</td>
<td>Business Intelligence/ Dash Boards</td>
<td>TUN relevant information</td>
<td>Dr. Ansari Notes</td>
<td>Bring in Laptop</td>
</tr>
<tr>
<td>22 June</td>
<td>Lecture DW1: Data Warehousing(I): Strategic View</td>
<td>JM 105-114, JM 127-134 and TUN relevant information</td>
<td>Dr. Ansari Notes</td>
<td>Turn in your class work5</td>
</tr>
<tr>
<td>29 June</td>
<td>Lecture DW3 : Dimensionally Designed DW (I)</td>
<td>JM 114-123 and TUN relevant information</td>
<td>Dr. Ansari Notes</td>
<td>Turn in your Preliminary report</td>
</tr>
<tr>
<td>6 July</td>
<td>Lecture DW5: OLAP and Business Intelligence Lecture DW6: Web-Based OLAP and Business Reporting/Presentation/Review</td>
<td>JM 135-137, JM 144-152 and TUN relevant information</td>
<td>Dr. Ansari Notes</td>
<td>Oral Presentation of Project and Turn in your Final Report</td>
</tr>
<tr>
<td>13 July</td>
<td>11</td>
<td><strong>Final Exam/Project Presentation</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONTENTS

SECTION ONE
MARKETING 1

Chapter One
An Overview of the New Direct Marketing 2

Introduction 2
The New Direct Marketing—What Is It? 3
The Premise of the New Direct Marketing 4
An Overview of Some of the Advantages and Practices of the New Direct Marketing 6
  Lead Grading 6
  Customized Targeting at the Right Time 6
  New Information and Past Results Help Formulate Strategies 7
  Information Can Drive New Programs and Fuel New Revenue Sources 7
  Information Can Foster New Services and Generate Repeat Orders 8
  Ongoing Communications Efforts Can Increase Customer Loyalty 8
Summary 9

Chapter Two
Contact Strategy 11

Introduction 11
Contact Strategy—What Is It? 11
RFM: The Workhorse Variables of Both the New and the Old Direct Marketing 12
Using Purchase History Information to Develop Contact Strategies 12
Using Promotion History to Define Contact Eligibility and Propensity to Respond 14
Using Demographic Data to Create Actionable Market Segments for the Purpose of Contact Strategy 15
Combining Both Purchase History and Demographic Data Can Lead to New Opportunities 15
Using Attitudinal Data to Create Actionable Market Segments for the Purpose of Contact Strategy 16
Specific Examples of the Role of Contact Strategy 17
Summary 23
  Marketing Program Strategy Guideline Checklist 23

Chapter Three
Buzzwords 29

A Word about Buzzwords 29
Buzzword 1: Generation of Languages 29
  Procedural Languages 30
Buzzword 2: SQL 31
Buzzword 3: Flat Files and VSAM File Structures 31
Buzzword 4: Indexes 32
  How Indexes Are Used 32
Buzzword 5: Productivity Tools 35
Buzzword 6: Logical and Physical Design 36
Buzzword 7: CPU 36
  Channel 37
  Channel Contention 37
  Direct Access Storage Device (DASD) 37
  Seek Time 37
Buzzword 8: Summary Files 38
  Normalization 38
Buzzword 9: Types of Database Products 38
  Hierarchical 38
  Inverted File 38
  Relational 38
  Proprietary Database Systems 39
  Other Database-Related Topics 39
Buzzword 10: Internet Buzzwords 40

Chapter Four

Sources and Uses of Marketing Data 42

Customer and Prospect Data 42
  Customer Data 42
  Prospect Data 44
  Nontransactional Data Sources 46

Examples of the Type of Data Collected from Three Types of Direct Marketing Business 49
  Financial Services Marketers 49
  Examples of Questions Asked by Financial Services Marketers 49
  Negative Options and Continuity Marketers 50
  Examples of Questions Asked by Book and Music Clubs 50
  Insurance Marketers 51
  Using Questionnaires to Gather Data 52
  Using Survey Data to Assign Customers and Prospects to Segments 52

Data Purchased from Third-Party Sources 54
  National Databases as Sources of Data for File Enhancement 57
  National Databases as Sources of Names 62
  Using Acquired Data to Predict Segment Membership 63

Chapter Five

Relationship Marketing and How It Relates to the New Direct Marketing 64

Some Background and Attempts at Definitions 64
Mass Markets 64
Today's Markets—The Impact of Technology 65
Today's Markets—The Impact of Demographic Trends 65
Relationship Marketing = Customer Retention + Share of Customer 66
Chapter Nine

Building a Marketing Database  107

A Framework for Implementing a Database Project  107
Will the Database Be Created and Maintained In-House or at an Outside Service Bureau?  108

Cost  108
Control  109
Customization  109
External Vendors  110
A Combined Approach  111

What Data Is Needed to Perform the Required Functions?  112
Identify the Files to Be Included in the Database  112
Review the Data Elements Contained in Each Contributing File  113
Select the Data Elements That Are Needed from Each File  113
Define Data Elements That Must Be Created during the Update Process  113

How Will Data Get into the Marketing Database?  113

In-House Processing Implications  114
Service Bureau Processing Implications  115

Decide Which, If Any, Data Enhancements Files Will Be Used  115
Consolidating Records  115

Duplicate Identification  116
Address Standardization  117
Matching Issues  117
Scrubbing  117
Householding  118

Develop a Preliminary Database Design  120
Frequency of Update or Replacement  120

Creating Separate Databases for Marketing to Avoid Conflict  121
Levels of Detail of Data  131

Will the Database Be Updated or Replaced?  131
Return to the Business Needs and Determine the Adequacy of the Plan for Meeting Those Needs  132

Chapter Ten

Using Data Hygiene to Identify Individuals and Households  134

Real-World Issues  134
Data Hygiene  135

Other Benefits of Data Hygiene  137
Applying Hygiene  137
Name Standardization  137
Address Standardization  138
Postal Address Standards  138
File Scrubbing  140
Determining Deliverability  140
Comparison among Files  141

Duplicate Identification  141
Householding  142
Chapter Eleven

Campaign Management 147

Campaign Phases 147
Management Processes 148
The Campaign Planning Phase 149
The Campaign Development Phase 150
The Campaign Execution Phase 152
The Campaign Analysis Phase 152
Profile Analysis 153
Response Analysis 153
Campaign Content and Structure Data 154
Actuals Data 155
Conclusion 155

SECTION THREE

WHAT DIRECT MARKETERS NEED TO KNOW ABOUT TECHNOLOGY 157

Chapter Twelve

Operations and Decision Support Systems 158

A Very Brief History of Computer Applications in Business 158
Operating Support Systems for Day-to-Day Processing 159
Decision Support Applications Can Be a Drain on Mainframe Resources 160
Decision Support Systems 160
Need for Flexibility and Quick Response 160
DSS Technology 160
Integrating Operations Support Systems and Decision Support Systems 161

Chapter Thirteen

The Unique Requirements Direct Marketers Place on Their Decision Support Systems 162
Direct Marketing’s Functional Requirements 162

Chapter Fourteen

Hardware and Software Fundamentals 164

Hardware 164
Mainframes 164
Midframes 165
Workstations 166
Microcomputers 167
Variation within the Data 209
Confidence Intervals 210
  Theorem 1 212
  Theorem 2 215
Confidence Intervals and Tests of Significance for Response Rates 215
  Theorem 3 215
Tests of Significance—Two Types of Errors and Power 219

Chapter Eighteen

Relationships between Variables 222
Correlation Coefficient 225
  Correlation Coefficient for Scalar Variables 227
  Correlation Coefficient in Practice 228
Simple Regression 233
  The Regression Line 233
  Simple Regression in Practice 235
Straightening Out the Data 238
  Explanatory Power 239
  Predictive Power 239

Chapter Nineteen

Multiple Regression 241
Multiple Regression Statistics—How to Read Them 242
  Quick and Dirty Regression 245
  More Quick and Dirty Regression 251
Regression Built with Care 253
  Step 1—Examination of the Correlation Matrix 253
  Step 2—Normalization of All Variables 253
  Step 3—Checking for Linearity 254
Rerunning the Model 262
Validation—Choosing the Best Model 264
Multiple Regression—Some Odds and Ends 266
  Interaction 266
  Multicollinearity 267
  Selection of Variables 268

Chapter Twenty

Response Analysis 270
Introduction 270
Logistic Regression 272
  Concepts and Definition 272
  Logistic Regression Model 274
  Logistic Equation 274
  Summary Variables 274
  Logistic Coefficients 276
Discriminant Analysis 277
Automatic Interaction Detection—AID/CHAID 277
The SI-CHAID Analysis of the Same Data 281
Using CHAID in Regression Analysis 284
Multiple Regression: Guidelines for Building a Model 287

Chapter Twenty-one

Segmentation Analysis 288
Data Needed for Segmentation Modeling 289
Factor Analysis: What It Is and When You Should Use It 289
How Factor Analysis Works 290
Running and Interpreting Factor Analysis 292
Cluster Analysis: What It Is and When You Should Use It 293
How Cluster Analysis Works 295
Running and Interpreting Cluster Analysis 295
Using Small-Sample Survey Data to Segment a Much Bigger File 297
Concluding Example and Review 298

Chapter Twenty-two

A Closer Look Back 302
A Closer Look at r, the Correlation Coefficient 302
Review of r 302
Rematching 303
Rematch Example 304
Rematching Phi 305
Implication of Rematching 307
A Closer Look at a Model’s Power 308
An Example 308
A Closer Look at Straightening Data 311
Ladder of Powers 312
Bulging Rule 312
An Example of the Bulging Rule 313
A Closer Look at Reexpressions for Many Variables 318
A Simple PCA Example 318
Algebraic Properties of PCA 320
Model Building with PCA—Two Case Studies 322
A Continuity Example 322
A Packaged Goods Example 324

Chapter Twenty-three

Artificial Neural Networks 326
Introduction 326
When Are ANNs Used 326
Classification 327
Prediction 327
What Are ANNs? 327
Definition 327
Basic Structure 328
Basic ANN 328
Basic Architecture 329
How ANNs Are Trained 330

Supervised Training Method 331
Designing an Optimal Multilayer Feedforward ANN 331
The Black Box ANN 332
Proof of the Pudding 333
ANN Model 333
Logistic Regression Model 334
Reruns 334
Discussion 336

SECTION FIVE
ADVANCED MODELING 339

Chapter Twenty-four
Assessment of Direct Marketing Response Models 340
Accuracy 340
Decile Analysis 341
Precision 342
Separability 344
All Together Now (Discussion) 344

Chapter Twenty-five
Direct Marketing Models Using Genetic Algorithms 345
What Is Optimization? 345
What Is a Genetic Algorithm? 346
Illustration of a Simple GA 346
Reproduction 347
Crossover 348
Mutation 350
GAs Today 351
GAs: Strengths and Limitations 351
The Genetic Algorithm DMAX Model 352
DMAX Fitness Function 352
DMAX Response Model 352
Powerful Response Modeling 353
DMAX Discrete Profit Model 353
DMAX Regression 355
DMAX Improvement over Ordinary Regression 357
DMAX Guarantee 358

Chapter Twenty-six
Bootstrapping in Direct Marketing 359
Traditional Response Model Validation 359
Illustration 359
Three Questions 360
Contents

CHAID Tree Graphs 396
Market Segment Classification Model 399
Summary 401

Chapter Thirty-one

Modern Methods of Testing in Direct Marketing 402
Introduction 402
Common Direct Marketing Testing Rules 402
Test One Thing at a Time 402
Test All Things 403
Statistical Experimental Design 404
Designed Experiments 404
Experimental Designs 404
Analyses of Experimental Designs 405
Blocking 405
Efficiency of Experimental Designs 407
Multifactor Experiments 408
Factorial Experiments 408
Statistical Experimental Design Rules 411

SECTION SIX

ECONOMICS, LIFETIME VALUE, AND THE ROLE OF MODELING IN THE NEW DIRECT MARKETING 413

Chapter Thirty-two

An Introduction to the Economics of the New Direct Marketing 414
Implied versus Contractual Relationships 415
Classical Direct Marketing versus Database Marketing 416
The Search for the Perfect Control 417
Once the Customer Is on the File 417
Some Assumptions about the Economics of Multidivisional Databases 419

Chapter Thirty-three

Back to Basics: The Economics of Classical Direct Marketing 420
Solo Promotions 421
Multistep Promotions Leading to a Sale 421
Catalog Sales 422
Continuity Sales 424
Negative Option Clubs 425
Newsletters 426
Magazines 427
Performance Measurement 427
Front-End versus Back-End Performance 427
Calculating CPM for Different Direct Response Media 428
Measuring Response 433
One-Step Promotions 433
Two-Step Promotions 433
Calculating Cost per Response 434
One-Step Promotions 434
Tracking Back-End Performance 435
  Measuring Back-End Performance 436
  Attributing and Measuring Lifetime Value
  in a Financial Services Environment 444
The Return of the Chicken-or-the-Egg Problem 446
Measuring Profitability: Combining Front-End and Back-End Measurements 446
ROP and the Infamous 2 Percent Response Rate 450
Appendix to Chapter 33 The Present Value Concept 451

Chapter Thirty-four

The Role of Modeling in the New Direct Marketing 477
The Difference between Scoring Models and Forecasts 478
  Obtaining Forecasts of Response 479
  Applying Scoring Models to Forecasts 479
Implementing Scoring Models at the Individual Level 481
Response Modeling: RFM versus Regression 482
  Eliminating Arbitrary Decisions from RFM 483
  Tree Analyses Compared with RFM Analyses 484
  A Quick Review of Interaction Variables 484
  Trees as a Benchmark against Regression 485
  CHAID, the Great Equalizer 487
Using Principal Components to Model Product Usage Patterns 488
Regression Models 490
  Reminder about Ordinary versus Logistic Regression Models 492
  How to Determine the Right Number of Variables to Include in
  Your Model 492
Typical Response Model Results 493
ZIP Code Models 495
  Adjusting for ZIP Code Size 498
  Modeling Variables Other Than Response 498
Combining Models Modeling Profit 500
  A Word of Caution: Regarding Combined Models 502
  Lead-Conversion Models 503
Attrition Models 506
Adding Household-Level Overlay Data to Prospect Models 507
Adding Household-Level Overlay Data to Internal Customer Models 516
Enhancing Your Database with Internal Survey Data 519
Implementing Models 523
Modeling, Managing, and Marketing to Unique Customer Segments 523

Chapter Thirty-five

Applications of the New Direct Marketing 526
About This Chapter 526
Moving from the Old to the New Direct Marketing 526