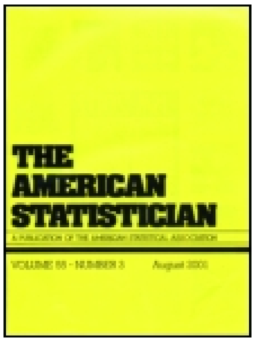

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://www.tandfonline.com/loi/utas20

Developing a New Interdisciplinary Computational Analytics Undergraduate Program: A Qualitative-Quantitative-Qualitative Approach

Scotland Leman, Leanna House & Andrew Hoegh

Developing a New Interdisciplinary Computational Analytics Undergraduate Program: A Qualitative-Quantitative-Qualitative Approach

Scotland LEMAN, Leanna HOUSE, and Andrew HOEGH

Statistics departments play a vital role in educating students on the analysis of data for obtaining information and discovering knowledge. In the last several years, we have witnessed an explosion of data, which was not imaginable in years past. As a result, the methods and techniques used for data analysis have evolved. Beyond this, the technology used for storing, porting, and computing *big data* has also evolved, and so now must traditionally oriented statistics departments. In this article, we discuss the development of a new computational modeling program that meets these demands, and we detail how to balance the qualitative and quantitative components of modern day data analyses for statistical education.

KEY WORDS: CMDA; Computational education; Curriculum; GAISE; QQQ; Statistical education

1. INTRODUCTION

Popular author, H. G. Wells (1903) stated, "it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write." Although this statement was made over 100 years ago, it is still true today and technology columnist, Steve Lohr, has one word of advice for every college graduate, "statistics" (Lohr 2009). Given the sheer throughput and rapid acquisition of data, the need for accomplished, creative, and thoughtful statisticians has never been greater. To develop and educate these statisticians, modern statistics courses within revamped analytical curricula are in demand. Alas, what "modern" and "relevant" mean is disputable among different institutions, industries, academic departments, and employees/faculty. The argument is that solutions to today's problems have computational, mathematical, statistical, and/or application-specific foundations, thus modernizing courses may focus on teaching advanced concepts in any one or subset of these disciplines. Also, "thoughtfulness" (e.g., problem solving and inquiry) is a personal process that arguably cannot be taught explicitly. After lengthy deliberations and via the collaboration of multidisciplinary faculty, Virginia Tech has developed a new program, called Computational Modeling and Data Analytics (CMDA), which integrates many aspects of analytical problem solving. CMDA graduates will have effective, technical problem-solving skills that can be applied in multiple settings. In this article, we introduce CMDA, and describe a common theme that the faculty considered in its development that parallels initiatives in the American Statistical Association (ASA) (Cannon et al. 2002; GAISE 2005; American Statistical Association Undergraduate Guidelines Workgroup 2014).

While the theoretical side of statistics is fundamental for optimizing statistical procedures, Cobb (1992) and Moore (1997) recommended using data to motivate ideas. Instead of motivating statistics from a theory-centric model, these authors advocate a focus on real applications (and all of their messy nuances), using technology and active learning in the classroom. Teaching statistics at either a practical or theoretical level has unique merits, as the goals and target audience can be dramatically different. At Virginia Tech, we have embraced the interdisciplinary nature of statistics (often referred to as analytics; see Section 4.1 for our working definition) and have developed the new CMDA program, which integrates computational aspects found in computer science, mathematics, statistics, and the physical and social sciences. One of the main focal points of CMDA is its emphasis on real data analysis. While general place holders for data, commonly denoted as X and Y in most classes, may help to illustrate technical mathematical aspects of analysis (quantitative), using real data is necessary for demonstrating the many qualitative and quantitative layers involved in data analyses. Gould (2010) stated, "today's students should (a) recognize data when they see it (b) understand how analyzing the data can help them and (c) know how to do so," which all require real data sources. Gould (2010) also provided a rather thorough list of various data types, and recommendations for incorporating these data into the classroom. Because of our emphasis on real data, our teaching motivates data analysis through a variety of techniques. However, instead of just adding new classes to our existing statistics curriculum, we have taken the opportunity to build an entirely new interdisciplinary analytics program, which combines skills from across the analytics disciplinary spectrum and are not typically deemed statistics per se.

Beyond integrating analytics, we focus on a balanced teaching approach that marries qualitative and quantitative elements of data analysis, and implements some of the ideas from Gal and Ograjenšek (2010). Specifically, most data analyses begin with a context-specific question and data collection method. Of course, the question of which data and techniques one might

Scotland Leman (E-mail: *leman@vt.edu*), and Leanna House (E-mail: *lhouse@vt.edu*), Computational Modeling and Data Analytics, and Department of Statistics, and Andrew Hoegh (E-mail: *ahoegh@vt.edu*), Department of Statistics, Virginia Tech, 410 Hutcheson Hall, Blacksburg, VA 24061. This work was supported by NSF grant DUE-1141096.

 DOI: 10.1080/00031305.2015.1090337

use to answer such questions remain. These are both inherently Qualitative issues (Q_1) and must precede any formal analysis. The second layer of analysis deals with the formal mathematics or computations required for acquiring numerical or visual summaries of the data, which are Quantitative issues (Q_2). Finally, these numerical summaries must be Qualitatively (Q_3) resummarized and assessed in a manner consistent with the questions asked in the Q_1 phase of the analysis. In turn we advocate, for each class, a more balanced $Q_1 - Q_2 - Q_3$ format (QQQ for short). Mathematical analyses are motivated only after a qualitative understanding of the primary questions. Furthermore, analyses are not considered complete until these analytical findings are concluded in a qualitative setting. This QQQ approach to teaching analytics mimics a more formal eight elements of thought for critical thinking (Elder and Paul 2007), and is specifically tailored to a data-driven analytical thought process.

This article discusses the development of CMDA, motivation for its existence, QQQ learning, and the future of interdisciplinary statistics programs at large. Section 2 details some of the challenges for motivating students in statistical education; Section 3 describes the principles of *QQQ* learning and gives an overview of the CMDA goals; Section 4 provides a thorough overview of the CMDA curriculum, and Section 5 provides concluding remarks.

2. MOTIVATING STUDENTS: AN INTERDISCIPLINARY APPROACH

Certainly statistics can be challenging, especially for those whom have not been properly prepared and motivated to think about analytical reasoning. Although it is always easier for us as instructors when students arrive prepared, we must accept that this is often not the case. In turn, we must take on the challenge of preparing *and* motivating students. Jumping straight into the analytical waters, without learning how to swim in them first, can only lead to one outcome: drowning! While we are purposefully being overly dramatic, it is true that not motivating students with a *thorough* qualitative understanding of the problems at hand is a recipe for failure. Before tackling quantitative concepts, it makes sense to motivate students with concrete qualitative ideas. A deep qualitative emphasis fosters statistical story-telling (Pfannkuch et al. 2010) for developing insights both into the analytical process and the impact of the analysis at hand. Statistical story-telling aids in moving students from understanding mechanistic concepts toward statistical thinking and conceptual understanding. In particular, focus is placed on comparative reasoning with the use of natural language for understanding and differentiating between descriptive and inferential thoughts. Albeit, one of the many endpoints in analytics is to employ analytical procedures, without a clear idea of why we are teaching topics that require quantitative skills, we need to spend a sufficient amount of time getting students to understand the problem. Thus, we advocate that students are given ample opportunity to play with real-world data before new, quantitatively based analytical ideas are taught. While Neumann, Hood, and Neumann (2013) demonstrated that real-world data applications make statistics courses more enjoyable for first year students, we maintain that this is a critical foundation that is necessary for understanding the issues that modern statisticians will encounter throughout their careers. Real applications give students insight into interesting structures in data without bombarding them with mathematical theory too early.

It would be entirely uncommon for an instructor to provide no qualitative motivation in a statistics class, however, many instructors are adopting a perhaps lopsided quantitative approach in their classrooms. Consider the following scenario: *An instructor motivates* t*-testing by discussing a two sample comparison (e.g., treatment vs. control). The instructor discusses the null hypothesis (e.g., treatment group is the same, on average, as the control), selects an* α*-level, shows the class how to compute the relevant* p*-value, and finally discusses rejection of such hypotheses.* Does this sound familiar? In this case, the instructor has adopted an unbalanced quantitative paradigm. This unbalanced paradigm fails to exploit the allure of posing scientific questions and solving them with data, which ultimately differentiates statistics from mathematics. Deming (1940), in response to Hotelling, stated, “Above all, a statistician must be a scientist . . . they must look carefully at the data, and take into account the conditions under which each observation arises.” The modern day PPDAC (Problem, Plan, Data, Analysis, Conclusions) cycle (Wild and Pfannkuch 1999) echoes this sentiment, which the ASA guidelines has adopted as “the scientific method.”

The heart of statistical disciplines is steeped in probability theory, which enables a precise mathematical quantification of uncertainty. However, the current teaching blueprint seems to place far too much emphasis on *mathematical* derivations, and not enough emphasis on *statistical* problem solving. This skews the perception of the field of statistics and limits the attraction of new talent. At times, the perception is that statistics is solely focused on mathematical derivations for seemingly trivial problems, and fails to show the utility of statistics for solving complex, relevant problems. Accordingly, students that may otherwise be drawn to the applied nature of statistics are distracted by abstract mathematical characterizations without any relevant applied foundation. So, we as academic statisticians must retool our field with practical skills that balance their theoretical counterparts. As Nolan and Temple Lang (2009) mentioned, we must be “unconstrained by legacy” and be willing to “eliminate” unnecessary and irrelevant layers of mathematical abstraction. These authors identify missing components in the current statistical training regime, which include: computation, statistical thinking, and experience.

In this era of massive, multifaceted, and complex data, the need for computation and data visualization far outweighs classical hypothesis testing, since they are able to provide much greater insights (Saraiya, North, and Duca 2005; Saraiya et al. 2006; Leman and House 2012). To remain relevant and attract students, we need to instill the principles of statistical and analytical thinking (Wild and Pfannkuch 1999; Elder and Paul 2007) and data analysis that transfer beyond textbook examples to modern problems. If we accept that the field of statistics is *the science of learning from data, and of measuring, controlling, and communicating uncertainty* (Davidian and

Louis 2012), then the greatest gift we can give students is the ability to problem solve and think creatively to compose a mathematically founded, data-driven solution. In essence, problem-based learning is more effective than simply teaching through the lens of quantitative theory itself (Hemlo-Silver 2004). Hardin et al. (in press) described this process as "thinking with data," where the authors outline how to implement data science ideas (computation, data acquisition, algorithmic thinking) into a modern statistical framework. Wild and Pfannkuch (1999) detailed the concepts of "statistical thinking" at large. While the process of statistical thought is individual, artistic, creative, with substantial quantities of quantitative rigor, the process can at times be vague and nonlinear. Wild and Pfannkuch (1999) loosely stated that statistical thinking is the "statistical incarnation of common sense." They deemed experience as a critical component of developing effective statistical thinking; however these authors also note that experience in itself is not always enough. Rather, this (qualitative) experience must be melded with a (quantitative) theoretical understanding in which to make sense of this experience.

With the CMDA program, we create a modern scholastic program with a core of computation and data analysis. Rather than competing with other fields, we meld tools and techniques for solving data-driven problems from the fields of mathematics, computer science, and social/physical sciences with statistics. We de-emphasize purely theoretical aspects of individual tools and techniques, and focus on combining and marrying various practical skills for solving application-specific problems. CMDA graduates will be well equipped for graduate study, careers in government, or industry and will have a place in what ASA deems "The Big Tent for Statistics" (American Statistical Association 2012). To borrow a line from a popular blog, a future statistician will need to be "problem first, not solution backward" (Leek 2013). That is, to remain relevant, we need to teach from the beginning of problems and use our qualitative skills in conjunction with our quantitative skills to find solutions.

3. QUALITATIVE-QUANTITATIVE-QUALITATIVE LEARNING: A CORNERSTONE OF CMDA

Beyond the skill sets we teach our students, one of the driving philosophies of the CMDA curriculum is the QQQ paradigm, which first instills a Qualitative understanding of the problem (Q_1), second develops Quantitative methods for answering such questions (Q_2), and third teaches students to Qualitatively explain and disseminate their results (Q_3). While not restricted to statistical education, QQQ is particularly well suited to problem solving. In fact, QQQ parallels the statistical analysis process, so we are "teaching what we do." QQQ partitions education into three components that flow into each other (typically bidirectionally in practice). The QQQ paradigm can be applied to curricula as whole, individual classes, or a single lesson.

3.1 Sequential and Balanced *QQQ* Learning

Creating completely balanced QQQ classes is one of our ultimate goals. However, this is not entirely possible, or even recommended for every analytics course. Some courses are inherently more qualitative or quantitative. For instance, a first course in probability theory is going to be more quantitative than some, but the qualitative aspects of developing such thought processes should never be ignored. For these traditionally Q_2 centric courses, we have developed the Integrated Quantitative Science (IQS, see Section 4.2.2 for details), which helps to mitigate teaching unbalanced Q_2 principles without their flanking Q_1 and Q_3 counterparts. On the other hand, more elementary classes might not be positioned to teach some of the more advanced Q_2 concepts. For this reason, we advocate for sequential QQQ classes, with the primary goal of developing a balanced program.

Valid sequential QQQ classes can be one of the three breeds: $Q_1 - q_2 - q_3$, $Q_1 - Q_2 - q_3$, or $Q_1 - Q_2 - Q_3$, where the large Q's indicate a strong emphasis, and the small q's denote a milder emphasis. As the emphasis of statistics is moving from asymptotic analysis (inherently quantitative) to more computationally intensive modeling-based procedures, we believe that qualitative aspects of analysis have become more important. For instance, given an unwieldy and complex dataset, old fashioned asymptotic analyses are often inappropriate, since they typically focus on a confirmatory analysis. However, in such settings, learning about structure in the dataset and which features are driving the analysis, and how models should be appropriately parameterized are crucial skills that require a strong qualitative understanding of the problem at hand. Acquiring such skills requires a lot of practice.

Nolan and Speed (1999) advocated laboratory lessons that teach theory through applications by using a series of intensive case studies or labs. Labs are divided into five main parts: (i) introduction, (ii) data description, (iii) background material, (iv) investigations, and (v) theory. This framework is quite similar to QQQ, and instills a balanced qualitative/quantitative approach to learning statistics. In Appendices A and B, we provide a teaching example ("The Consumer Spending Problem") with discussion that highlights the QQQ paradigm. As a starting point, theoretics and methodologies are not described. Rather, the problem is motivated with a concrete objective and dataset, and illustrates how to combine analytics skills to solve challenging data-driven problems.

3.2 Yesterday's Versus Today's Students

The field of statistics as a whole has developed and defined methods for experimental design, point estimation, hypothesis testing, probabilistic inferences, and asymptotics. Ultimately, these procedures were created for the purposes of knowledge confirmation, where one wishes to test whether two (or more) groups (e.g., test vs. control) differ and estimate their effect sizes. In general, the individual performing the testing will have some idea whether there is a meaningful effect and is able to confirm their idea(s) through collected data.

Modern day analytics places emphases on predictive modeling, computational inferential procedures, and pattern recognition, which are heavily steeped in algorithmics. Although yesterday's procedures are centrally focused on knowledge confirmation, today's procedures are oriented around *knowledge* *discoveries*. The confirmation or rejection of hypotheses is still relevant, but today's analysts must also explore data before knowledge can be confirmed.

The changing analytical trends are a direct consequence of the fact that a majority of the data collected today are vast in size and often unstructured. Emphasis has changed from finding optimal testing procedures to ascertaining the structure of data through flexible modeling procedures. Since mathematical procedures are often specific to subclasses of problems, computational methods have become commonplace based on their relative ease, usefulness, and adaptability. For example, estimating quantities from a distribution (parameter, quantiles, and interval estimates) analytically can be arbitrarily complex, and the complexity of the mathematics involved depends strictly on the distribution at hand. However, using Monte Carlo methods for estimating such quantities is relatively simple, instills the basic ideas in statistics, and provides a powerful tool that can be used regardless of mathematical aptitude.

Many mathematical results are still relevant and important to analysts; however, gaining practical experience often outweighs the need to exhaustively research the mathematical properties of relatively simpler methods. While we do not dismiss the relevance of training graduate students to understand the mathematical nuances of statistical procedures, for undergraduate students, we believe it to be more important to teach from a real-world data-driven perspective, embracing flexible computational methods, and model validation procedures. This application-oriented approach to teaching reflects the applied perspective under which the field of statistics was originally created (Box 1987).

3.3 Program Goals: What Kinds of Students are We Developing?

The ultimate goal of any collegiate degree program is to develop a set of skills that students can draw on well into their futures. In essence, modern analytics programs need to develop creative and flexible problem solvers. While understanding theory is essential, it is also essential to be able to put such theory to use. Creating students who can solve complex, data-driven problems starts with teaching students to ask the right questions. Because evaluating creative thought processes can be a challenge, we leverage rubrics and journal writing to instill and assess these qualities in our students (see Section 4.4 for details). For the example problem supplied in Appendix A, we provide a sample assessment rubric in Appendix C.

Analytics is not simply about quantitative theory; it is about a start-to-finish process of asking questions, finding solutions, and generally iterates between the two until well-formed solutions are found. Extending the QQQ metaphor, these iterations create a feedback circuit, constituting a bi-directional flow in the QQQ hierarchy as shown in Figure 1 illustrates a rigorous, yet flexible and creative process. The process is initialized by a scientific inquiry and seeks initial exploratory insights into the data (Q_1). This phase of analysis will attempt to solidify the questions that require answering, and confront various data related issues (such as identifying potential biases, identifying simple

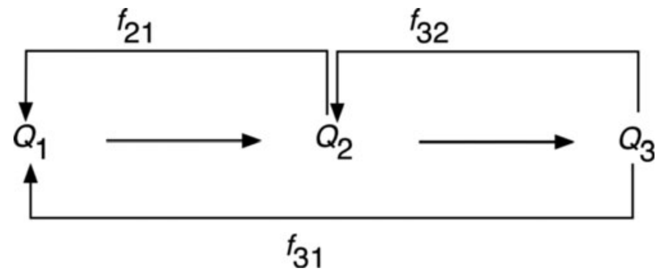

Figure 1. The QQQ paradigm, with feedback loops. f_{ij} represents feedback from Q_i into Q_j.

variable relationships, etc.). After which, the process adapts to asking more data-driven questions that provide an understanding of what needs to be analyzed, and how ($Q_1 \rightarrow Q_2$). This phase of the analysis connects skills and techniques across the analytical spectrum. The final phase of QQQ process is achieved by assessing the methodology, interpreting results, and disseminating results in the original qualitative language the problem was introduced ($Q_1 \rightarrow Q_2 \rightarrow Q_3$). At every given phase, an analyst may want to impose feedback, and refine a previous step of the analysis. As described in Figure 1, f_{ij} represents feedback from the Q_i to Q_j phase of the analysis. Examples of feedback include: model validation (f_{32}), assessing model/data limitations (f_{21}), and steering analyses to reassess and redirect scientific inquiries (f_{31}). Accomplishing this last type of feedback is the holy grail for any applied analyst. In Appendix B, within the context of the consumer spending problem, we provide some further examples of these feedback loops. The *QQQ* process teaches students how to seamlessly move between qualitative and quantitative tasks, and balance all the traits of a skilled data analyst.

To train the data analysts of tomorrow, we need to develop critical thinkers that have strong qualitative and quantitative skills. Beyond this, we must instill in our students strong technical skills that include: operating databases (e.g., managing large datasets and assembling them), exploratory data skills (e.g., visualizing multiple aspects of the data), data cleaning skills, conducting simulation studies (Monte Carlo), and becoming fluent in mathematical theory (but not constrained or stifled by it). Yesterday's students had the luxury of being provided *clean* datasets, because emphasis was simply on conveying statistical concepts. However, today's students must learn skills that help them set up datasets, and prepare them for analysis. In practice, it is not uncommon for statisticians to help guide the scientific process, and help with data collection and management. In the big data era, problems are down-right messy, and a simple Q_2 approach to analysis will no longer do. By sheltering students from these messy (and technical) issues, we are not preparing them to succeed when presented with unfamiliar datasets.

Finally, students must learn to communicate effectively in each phase of the QQQ process. Within the Q_1 phase, this is absolutely critical. Communication must be bi-directional, in that students must learn to listen and respond with thoughtful, well-formed questions. Within the Q_2 phase of the analysis, communication is often much more difficult, particularly when communicating with those without technical mathematical skills. Thus, it is necessary to teach students how to effectively communicate the *spirit* of the analysis. Also, a well-formed simulation study can help to overcome such communication issues; well-formed simulation studies demonstrate the effectiveness of analytical solutions in hypothetical cases where the truth is known. Such studies show when and where methods are expected to capture the truth, and when they might fail. Instead of speaking with a purely mathematical vocabulary, simulation studies help to transition the conversation to a procedural approach, which is often easier to grasp. Within the Q_3 phase of analysis, students must learn to tell the story, which completes the analytical process. Without making strong conclusions about the problem that was analyzed, the analytical process is not complete. Hence, quantitative conclusions must be translated into a language that is compatible with the original problem statement.

By providing students with these skills, we aim to create the complete analyst. These students will match theoretical and practical skills, be able to handle all technical aspects of the analytical process, and communicate their results effectively. These skills exceed the more constrained learning objectives of yesterday's students, and balance the qualitative and quantitative aspects of today's analyses.

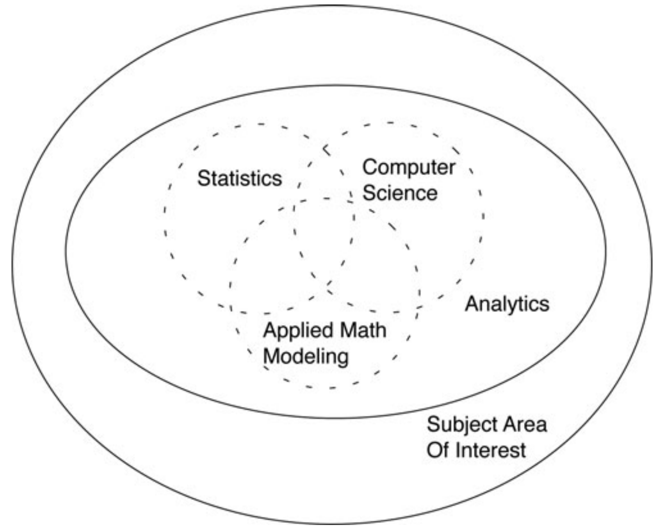

Figure 2. The analytics sphere. Analytics combines skills from statistics, computer science, applied mathematics, and varied application fields.

4. DEVELOPING A NEW MAJOR: COMPUTATIONAL MODELING AND DATA ANALYTICS

Because of the computation and analytical demands that inundate all data-driven problems, we at Virginia Tech have created a new degree program. While we consider this to be an "analytics" program, CMDA very much embraces the goals of a modern, technology-driven statistics program that combines computational skill sets from statistics, computer science, and mathematics, and integrates these cutting edge data applications. Before describing the structure and organization of this program, we discuss our choice of branding this as analytics, rather than statistics.

4.1 What is in a Word? Data Analytics Versus Statistics

The word *statistics* has been in use since the mid 1700s, and has seen a wide variety of uses. From the description of raw data, the analysis of such data, to the design of experiments, the term has been loosely used to mean the science of data analysis. Academic statistics departments are now struggling to define their field in the presence of modern analytical devices (Brown and Kass 2009).

While many have their own definitions of statistics and analytics, Figure 2 shows how the authors and CMDA envision the connections between the disciplines. The sub-disciplines (which comprise the core departments involved in CMDA) are illustrated by dashed circles, denoting the weak boundaries between these fields. While there is a large degree of overlap between the disciplines, each specializes in a particular aspect of the analytics domains. However, analytics is bigger than the sum of its parts. That is, integrating these concepts is more than just offering a varied course curriculum that selects classes from their respective fields. Rather, the field of analytics is all about how one might connect the ideas in these fields to solve real data-driven problems. Subject areas of interest typically guide which analytics tools and measures answer questions developed in the Q_1 phase, hence analytics should always be taught in the context of real data problems.

While the integration of applied concepts could easily take place in statistics departments alone, completely revamping existing statistics departments in a single stroke is both impractical and worrisome (again, see Brown and Kass 2009). However, being able to leverage our analytical peers (math, computer science, and other varied engineering and science programs) is an attractive alternative. CMDA is an interdisciplinary program that stresses skills across the analytics spectrum, from deterministic to stochastic methods. This program is not modular; rather it is a true integration that teaches skills in a manner that is useful for actual data analysis. While this is a labor intensive initiative, we believe the efforts will best serve our new analytics students, and prepare them for the new frontier in analytics (House and Leman 2014) and help them choose and combine techniques that actually help them to solve real-world problems.

4.2 CMDA Curriculum

When developing a new degree, there are logistics to consider in addition to selecting required courses and content within, including number of credit hours for graduation and personnel available to teach classes. At Virginia Tech, students are required to have at least 120 credit hours; approximately 25% of which are allocated toward a general education, which we refer to as the Curriculum for Liberal Education (CLE). All Virginia Tech students, regardless of major, must complete courses that satisfy the CLE requirements. Thus, any major may require as many as 90 credits or approximately 30 classes (a typical class is 3 credit hours). We consider majors to be highly restrictive when they require classes that sum to 90 credits because students do not have room to take free electives, unless the students

Table 1. List of required courses for majoring in CMDA with associated credits per class in parentheses. This list does not include general education courses that are required by Virginia Tech (as defined in Section 4.2). To complete prerequisite courses, students must take the first three courses listed and choose either Option A or B. Option A (IQS) integrates concepts from all courses listed in Option B to prepare students for future classes

Prerequisite courses (23–28 credits) *Students select option A or B*		Core courses (21 credits)	Select 4 electives (12 credits) *Below is a subset of possible classes (3 credits each). Students select 4.*
Univariate Calculus (6)		CMDA Computation (6)	Intermediate Math Modeling
Linear Algebra (2)		Data Analytics (6)	Comp. Stoch. Modeling †
Intro. to Software Design (3)		Mathematical Modeling (6)	Numerical Analysis
Opt A:	Opt. B:	Capstone Project (3)	Applied Multivariate Analysis
IQS (12)	Multivariable Calculus (3)		Intro. to Statistical Genomics
	Vector Calculus (2)		Experimental Design
	Intro. Statistics (6)		Regression
	Intro. Differential Eq.*(3)		Applied Bayesian Analysis
	Prob. and Dist.+ (3)		Intro. Computational Physics
			Database Management Systems
			Data Structures and Algorithms
			Statistical Computing
			Bioinformatics Methods
			...

NOTE: * Intro. to Differential Equations.
+ Probability and Distributions.
† Computational Stochastic Modeling.

take more than 120 credits before graduation. Less restrictive majors require fewer than 90 credits and offer opportunities for students to take free electives, double major, and/or earn minors. For example, computer science is a more restrictive major than statistics at Virginia Tech because, after accounting for the CLE and completing major-required courses, computer science students have only 15 open credits to take free electives, whereas statistics students have 30 credits. Thus, statistics students have more opportunity to double major or minor in other areas.

With these details in mind, CMDA integrated concepts from computer science, mathematics, and statistics courses to create an analytical degree of comparable flexibility to Statistics that requires 61 credits of courses. The new degree includes current courses from the three contributing majors and new courses that we developed specifically for CMDA so that students efficiently gain an integrated expertise in areas relevant for analytics. For explanation and comparison, we provide Tables 1 and 2 to list CMDA and Statistics major requirements, respectively. Loosely, CMDA students take prerequisite courses during their freshman and sophomore years and upper-level, CMDA-specific classes during their junior and senior years. Now (Section 4.2), we state the topics and courses selected and/or designed for CMDA majors. Subsequently, in Section 4.3, we describe how the progression of required classes in CMDA follows the *QQQ* paradigm.

Table 2. List of required courses for majoring in Statistics with associated credits per class in parentheses. This list does not include general education courses that are required by Virginia Tech (as defined in Section 4.2). There are clear differences between CMDA and Statistics in the core and elective requirements

Prerequisite courses (25 Credits)	Core courses (27 Credits)	Select 3 electives (9 credits) *Below is a subset of possible classes (3 credits each). Students select 3.*
Univariate Calculus (6)	Theoretical Statistics (6)	Nonparametric Statistics
Multivariable Calculus (3)	Advanced Calculus (3)	Applied Bayesian Analysis
Vector Calculus (2)	SAS programming (3)	Intro. Computational Physics
Linear Algebra (2)	Regression (3)	Data Analytics I
Intro. Statistics (6)	Experimental Design (3)	Data Analytics II
Prob. and Dist.+ (3)	Statistical Computing (3)	Bioinformatics Methods
Intro. to Software Design (3)	Statistics Communication (3)	Contingency Tables
	Technical Writing (3)	Applied Multivariate Analysis
		Intro. to Statistical Genomics
		Sample Survey Methods
		Comp. Stoch. Modeling †
		Econometrics
		...

NOTE: + Probability and Distributions.
† Computational Stochastic Modeling.

4.2.1 Core Classes

Central to CMDA are three two-course sequences (i.e., 6 courses that total to 18 credits) and a capstone project. The sequences are in (i) Mathematical Modeling, (ii) Data Analytics, and (iii) CMDA Computation (computational skills that are particularly relevant for data management and simulation). Five of the six courses were developed specifically for CMDA. Table 3 highlights key concepts that are covered in each of the six courses. These key concepts were discussed by a multidisciplinary team and selected carefully from current courses and modern areas of research in analytics. To any one researcher in a discipline, some concepts might seem out of place, but to others, they are crucial. For example, to a statistician, it might seem unnecessary that CMDA covers differential equations. However, there are a diverse sets of tools that span across fields to assess physical processes. Statisticians may assess a process between inputs and outputs using data as boundary conditions, but mathematicians may impose theoretical conditions in differential equations instead. CMDA integrates both statistical and mathematical approaches. Furthermore, in accordance with suggestions made by Utts (2014), we advocate that ethical standards (treatment of human/animal participants, assurance of data quality, appropriate statistical analyses, and fair reporting of results) be imposed throughout the curriculum. These standards must be applied and reinforced continuously.

At the conclusion of these courses, students have a skill set to design computer simulators, manage large and small datasets, and assess both simulated and real datasets using supervised and unsupervised methods. They demonstrate these skills with a semester-long project in the capstone course. Because of time constraints, students must complete at least two of the three core sequences to enroll in the capstone course; ideally they will have taken all three. During the capstone course, students identify a research question, develop a comprehensive solution, and present results. Ideal projects provide publishable results and students may continue research as an independent study, if they take this course during their fall semester, senior year. One faculty member will lead the class, but the students will be encouraged to identify other CMDA faculty or partner from industry with whom to work.

4.2.2 Prerequisites and Integrated Quantitative Science (IQS)

The two-course sequences and capstone course discussed in Section 4.2.1 have prerequisites that span linear algebra, calculus, differential equations, probability, hypothesis testing, simple linear regression, and introductory programming. Introductory courses that cover these topics are not too different from those required of typical statistics majors, as well as overlap with early requirements of computer science and mathematics majors. Students may satisfy the prerequisites by taking eight traditional classes that are readily available (Option B in Table 1). However, we introduce a new approach for teaching some of the prerequisite material to improve the education of CMDA students and promote integrated analytical thinking from the start. In this approach (Option A in Table 1), students take four, rather than eight, classes, including a new 12 credit sequence (typical sequences are 6 credits) that we call Integrated Quantitative Science (IQS). Not all students will be able to take IQS, such as those who transfer into the CMDA major, but it is the recommended method for meeting the CMDA prerequisites, whenever possible.

IQS is team-taught by a mathematician and statistician and streamlines the education of CMDA in that the students' prerequisite credit load is reduced from 28 to 23 credits. This is because the content covered in IQS is considered equivalent to the key concepts covered in all the classes listed under Option B in Table 1. The motivation for IQS is to start early in promoting the idea that answers to research questions may not be discipline specific. For example, to identify patterns in a physical process, mathematicians might design a simulator of the process based on differential equations; computer scientists might use machine learning methods with exhaustive permutation strategies; and statisticians might sample data and apply the central limit theorem to make probability-supported inference. CMDA graduates will have the technical skills to complete any one solution individually, but also the problem-solving skills to either assess which option is best or merge the options to develop a "best" solution. To do so, CMDA students must appreciate both the similarities and differences in the disciplines, as well as how they can reinforce one another. When students take IQS, they may start to develop their appreciation as they learn, for example, matrix manipulations and decompositions (e.g., QR and singular value decomposition) with regression; multivariable integration with probability and Monte Carlo; and/or deterministic with stochastic models.

4.2.3 Technical Electives

In addition to the prerequisites, three core sequences, and capstone, CMDA students further develop their technical skills in areas of their choosing by taking four restricted elective courses. That is, approximately 20 courses have been identified as suitable for CMDA students; students must take four of these 20 classes. Some of the options were developed specifically for CMDA, but most are classes that are readily available in typical computer science, math, and statistics departments. For example, students may take Advanced Mathematical Modeling, Introductory Stochastic Processes, Bayesian Statistics, and Parallel Processing as their four electives.

An interesting feature of CMDA is that students may also integrate their education even more with other sciences, such as, Biology, Physics, and Neuroscience. Several applied disciplines have expressed interest in working with CMDA to design science-based computational courses that are accessible to both CMDA and science majors. In fact, professors from physics were part of the CMDA development team and designed a physics track within CMDA. The purpose of such tracks is to prepare application-oriented students for computational positions within industry or academia upon graduation.

4.3 CMDA and *QQQ*

The *QQQ* paradigm in teaching data analytics may apply to individual lessons, isolated courses, and entire degrees. As mentioned in Section 3, each application may emphasize all *Q*'s

Table 3. Prerequisite and core, seminal courses for the CMDA program are listed in this table. The courses are labeled as either a prerequisite (Pre) or required (Req) course and as part of a sequence, when applicable. The sequences are highlighted in Section 4.2.1 and, in this table, Seq. 1 references Mathematical Modeling, Seq. 2 references Data Analytics, and Seq. 3 references CMDA Computation. Also, which Q's of the QQQ paradigm are emphasized in the courses are stated in the right most column

Purpose	Course name	Description	QQQ
Pre	Integrated Quantitative Science	Integrated topics from quantitative sciences. Topics include: probability and statistics, infinite series, multivariate calculus, linear algebra, vector geometry.	QQq
Req; Seq. 1	Mathematical Modeling I	Mathematical modeling with ordinary differential equations and difference equations. Numerical solution and analysis of ordinary differential equations and difference equations. Stochastic modeling, and numerical solution of stochastic differential equations.	QQq
	Mathematical Modeling II	Concepts and techniques from numerical linear algebra, including iterative methods for solving linear systems and least-square problems, and numerical approaches for solving eigenvalue problems. Ill-posed inverse problems such as parameter estimation, and numerical methods of computing solutions to inverse problems. Numerical optimization. Emphasis on large-scale problems.	QQq
Req; Seq. 2	Intro Data Analytics and Visualization	Basic principles and techniques in data analytics, including, what is meant by ”learning from data;” methods for collecting, storing, accessing, and manipulating standard-size and large datasets; data visualization; and identifying sources of bias. Concepts are applied to real-world case studies.	QQQ
	Intermediate Analytics and Machine Learning	A technical analytics course that will teach supervised and unsupervised learning strategies, including regression, generalized linear models, regularization, dimension reduction methods, tree-based methods for classification, and clustering. Upper-level analytical methods are shown in practice: for example, naïve Bayes and neural networks.	QQQ
Req; Seq. 3	Software Design and Data Structures	A programming-intensive exploration of software design concepts and implementation techniques. Builds on knowledge of fundamental object-oriented programming. Advanced object-oriented software design, algorithm development and analysis, and classic data structures. Includes a team-based, semester-long software project.	QQq
	CMDA Computation	Survey of computer science concepts and tools that enable computational science and data analytics. Data Structure design and implementation. Analysis of data structure and algorithm performance. Introduction to high-performance computer architectures and parallel computation. Basic operating systems concepts that influence the performance of large-scale computational modeling and data analytics. Software development and software tools for computational modeling.	QQq
Req	Capstone	Capstone research project to cultivate skills including reviewing the literature, creative problem solving, teamwork, critical thinking, and oral, written, and visual communications.	QQQ

equally or some more than others. Crucially, students are held accountable for all three to become good problem solvers and analysts.

Classes developed specifically for CMDA follow the QQQ paradigm, as stated in the last column of Table 3, starting with IQS. IQS is taken during a typical student's sophomore year and emphasizes Q_1 and Q_2. As explained above, students in IQS master technical skills in mathematics, statistics, and computer science. While doing so, the students reflect upon the skills and assess differences among them. This reflection and assimilation of new content requires intensive qualitative forms of thought. Also, as part of the reflection, student are exposed to Q_3 forms of thought when they compare implications of problem solutions. However, the students do not complete applied projects and present results. Thus, IQS is both Q_1 and Q_2 heavy, with some exposure to Q_3.

During a typical junior year for CMDA majors, students take at least two required sequences that emphasize the Q's differently (Table 3). The Mathematical Modeling sequence is QQq because it focuses on simulation techniques and reasons to use them; the CMDA Computation sequence is also QQq because the students are motivated by challenges presented by real datasets and taught tools to process and manage them; and the Data Analytics sequence is QQQ because the students are motivated by applications, learn tools to summarize data, and make inference.

Then, during the senior year of CMDA majors, students take a capstone course that also unites Q_1, Q_2, and Q_3 equally. To complete the capstone, students draw upon what they know, select a problem in which they are interested, devise a solution, and communicate their results in both written and oral forms. During the oral presentation, the students must defend their analytical choices and present implications of their findings.

4.4 Assessment in CMDA

As mentioned previously, there are logistics to consider when developing new programs. Until now, we have only discussed those that are relevant to the students, including credit hours and content covered in required courses. However, there are logistics relevant to instructors as well, including methods of assessment. Clearly, good forms of assessment provide accurate measures of whether students have met course objectives. However, when there are qualitative objectives (e.g., master skills that are inherently subjective and do not have clear definitions of right vs. wrong), developing useful forms of assessment is challenging, at best. Furthermore, in classes with many students, assessment must be time efficient and repeatable. For CMDA, we turn to resources that are readily available to assess student mastery of both quantitative and qualitative skills. For example, we consider ideas from GAISE (2005) and the Association of American Colleges and Universities (AAC&U).

Primarily, the GAISE report suggests using active learning strategies when teaching introductory statistics. Additionally, in the GAISE report, it is suggested to conduct group, rather than individual assignments; enable peers to evaluate assignments; and use recitations to develop qualitative skills. Typical recitations are much smaller than large classes so that discussions are possible across most, if not all, of the students. When we take this advice for CMDA, the number of assignments to grade decreases and the number of people available to grade increases.

Alas, Q_1 and/or Q_3-focused assignments are still a challenge to grade. For such assignments, versions of the rubrics designed by the AAC&U may apply. In particular, consider the Valid Assessment of Learning in Undergraduate Education (VALUE) Rubrics for Inquiry and Analysis, Critical Thinking, and Problem Solving (Rhodes 2010). These rubrics can be altered for assignments at hand and are accessible to both teaching assistants and professors for fair, adequate grading by multiple people. We considered rubrics for Quantitative Literacy, Oral communication, and Written Communication to create the rubric provided in Appendix C to evaluate an analytical project described in Appendix A. With the advantages of rubrics, come disadvantages. It takes time, effort, and open lines of communication among students, teaching assistants, and professors to develop effective and standardized rubrics. In fact, even after tremendous effort, rubrics may not guarantee perfect evaluations of student work. But, rubrics offer an opportunity to evaluate qualitative work well and to give relevant feedback to students on assignments, when standard methods of assessment do not (Reddy and Andrade 2010; Kenworthy and Hrivnak 2014). Thus, for courses and projects within CMDA, we support the use of rubrics.

One final suggestion is to include assignments that are graded as complete or incomplete, such as journal or blog writing (Burrows et al. 2001). Although journaling is informal, it forces students to use their own words to describe their thoughts before, during, and after quantitative methods are taught; they have an opportunity to make connections that they may not have otherwise and actively write about their problem-solving issues and critical thinking strategies (Blake 2005). For example, when teaching K-means, professors may ask students to journal about how K-means may miss clusters in data or to describe examples or characteristics of applications for which K-means may or may not be useful. These are Q_1 and Q_3 elements of analytics with K-means.

5. CONCLUSION

Today's datasets can be larger and messier than ever and require analysts to have technical, problem-solving skills that are not taught nor practiced in traditional statistics classes. Thus, there is a clear need to revamp (or create) analytical programs for developing statisticians who can respond to data-driven questions. In response to this need, we introduce CMDA, a new integrated analytics program, with a statistical core, which leverages our analytical peers in mathematics, computer science, and the physical and social sciences.

To create CMDA, (i) we constructed integrated analytics classes, with an emphasis on data related issues, and (ii) we enhanced the qualitative nature of analytics in our teaching. For example, we applied the QQQ paradigm for determining the syllabi of new courses that were specially created for the CMDA major. Some courses emphasize one Q more than others, and some emphasize all three equally. Additionally, CMDA students are trained with *real* datasets that exhibit issues they will encounter throughout their statistical/analytical careers. In the context of practical applications, students are continuously prompted to practice and refine both the qualitative and quantitative aspects of problem solving. Additionally, we have extended the technology presented to analytics students to prepare them for the frontiers of the big data era.

While the age of modern statistics began approximately 100 years ago, with an emphasis on the foundations of probability, design of experiments, and hypothesis testing, the golden age of analytics looms. Preparing students to handle real data issues is crucial for any analyst in training. For some students, a rigorous theoretical treatment of statistical models and methods will still be of interest; however, we believe that the current trends in the field suggest that future students might demand a more applied, technology-driven training. The latter students must understand qualitative data issues, and balance their skills with a sound quantitative methodology. We see integrative analytics programs as a viable alternative to fully revamping undergraduate statistics programs, which is what is underway in the new CMDA degree program.

APPENDIX A: CONSUMER SPENDING PROBLEM: STUDENT HANDOUT

Developing an understanding of consumer spending patterns is an important problem, especially for retailers. Blanket advertising campaigns are not particularly effective, since they do not discriminate between individual spending patterns. However, understanding an individual's spending patterns can be a powerful predictor in assessing their likelihood to purchase various items. Furthermore, these patterns often suggest how to effectively incentivize consumers to make purchases. Learning the propensity for a consumer to shop at an establishment allows retailers to implement targeted, data-driven advertising. The goal of this assignment is to create a strategy for predicting the number of transactions at a merchant during a 12 month period and total dollars spent at a merchant.

Data Description: The data contain historical transactions for individuals. For each transaction in the dataset, you are provided with: *a timestamp, the purchase price, a retailer identifier, the retailer type, the zip code, and a binary indicator for online purchases.* Each student is provided with six DVDs (roughly 30 GB total) that store tables pertaining to transactions over a 6-year period. Furthermore, there are two separate datasets corresponding to training customers and test customers. The training data contain years 1–3 for all customers. The test set only includes years 1 and 2, with the intent that third year will be predicted.

Prediction: For all customers, we are interested in predicting purchases at various retailers. The goal is to predict if an individual shops at a particular retailer, as well as dollars spent. Your predictions will be evaluated using the logistic loss function:

$$L(x_{ij}|p_{ij}) = x_{ij}\log(p_{ij}) + (1 - x_{ij})\log(1 - p_{ij}),$$

where x_{ij} and p_{ij} represent an indicator for a transaction between customer *i* and retailer *j* and its corresponding probability. Additionally, the predicted transaction amounts will be evaluated using the squared error loss function:

$$L(\hat{s}_{ij}|s_{ij}) = (\hat{s}_{ij} - s_{ij})^2,$$

where s_{ij} denotes the transaction amount of customer *i* at retailer *j*, and $\hat{s}_{ij}$ is the corresponding predicted spending.

Requirements: Students will work in teams of four to five students, and compete to provide a comprehensive marketing strategy based on the historical transaction records. Additionally, students will provide a 15 min classroom presentation of their work, and a written document which will include: (i) A Qualitative overview of the problem. This will detail the data handling/preparation, and qualitative insights imposed. For instance, if zip codes are included in the model, how were they used, and why might they be useful. (ii) A Quantitative description of models/tools/techniques used for prediction. This section will make strong reference to qualitative insights, and how these were quantitative encoded in your model. (iii) A Qualitative discussion of results, conclusions, and additional insights acquired.

APPENDIX B: DISCUSSION OF THE CONSUMER SPENDING PROBLEM

Here, we discuss the goals of the project through the lens of our QQQ framework, which is imposed throughout the CMDA curriculum.

Qualitative: Answering the question of *"Who shops where, when, and for what?"* is crucial for targeted incentivized marketing regimes. This problem is unique in that it is a relevant big data scenario in which each student has the requisite background/scientific knowledge to tackle the problem. Every student makes purchases and hence, can think about specific factors that may influence purchases. The Q_1 component contains the preliminary aspects of the analysis often referred to as the exploratory data analysis (Tukey 1977). In particular, we stress the importance of exploratory data visualization (i.e., visualizing the raw data). This dataset is also valuable as it places a major emphasis on data management skills (e.g., data streaming, caching, and database management skills). The physical size of the dataset itself presents challenges as it is larger than the typical working memory of a laptop computer (16GB). Each row in our dataset corresponds to a transaction, whereas for a typical analysis, or data visualization, each row should correspond to shopper. Hence, useful insights can be acquired through derived variables (variables not initially compiled). For instance, the original dataset includes zip codes, however these in themselves are not informative. However, knowing the driving distances between consumers and retailers is highly informative. Another example of a derived variable might be the frequency of times that a consumer shopped at a particular type of retailer in the last 3 months (or 6 months, etc.). These qualitative insights form the basis for an effective marketing program.

Quantitative: The quantitative aspect of this problem focuses on modeling customer behavior. While a simple linear framework might be useful for predicting the total dollars spent at a retailer, a nonlinear method (e.g., a decision tree) might be better suited. To assess model performance(s), students must attempt to forecast shopping patterns using their supplied training dataset, and measure and compare the predictive performances under various modeling iterations. Again, the authors stress visualizing modeling results; hence, visualization should also be part of the Q_2 phase.

Qualitative feedback: As Figure 1 illustrates, quantitative issues are coupled with insights developed during the qualitative phase. In turn, quantitative measurements can also provide qualitative insights. For instance, after developing a relatively simple decision tree model, students may discover heterogenous groups that exist within the population. For instance, customers that shop regularly might tend to spend more than customers that only shop periodically. Noting this, the student might reassess their qualitative understanding of the problem, and propose to stratify the data into groups and reparameterize their model accordingly (f_{32} feedback). These insights will warrant an iterative adjustment to the quantitative model through some clustering mechanism.

Qualitative: The final piece of the analysis is the qualitative closing, which returns to the question stated in the Q_1 section: *Who shops where, when, and for what?* This includes an overview of predictive performance; however, the challenge is unpacking the inner workings of the statistical machinery to explain predictions each student generates. In general, the closing returns to the realm of the problem and away from the details of the statistical methodology. Given that we are all "experts" in spending, the qualitative closing must explain the results in plain language. Decimating results using (clean) visualizations is also of paramount importance in the Q_3 phase. Upon project completion, many new questions might arise. Speculating about new questions, and understanding which data are necessary for answer those questions will help drive future inquiries (f_{31} feedback). For example, if we had data concerning customer's social networks (e.g., Facebook), we might be able to simultaneously incentivize the customer's "friends."

APPENDIX C: CONSUMER SPENDING PROBLEM: RUBRIC

Component	Sophisticated	Competent	Novice	Points
Summarize Data				**15**
Data handling	Comprehensive description of the procedures to handle and import data.	Vague discussion of importing and processing data.	No mention of how data are imported and processed.	5
Variable creation	Discussion on how variables are created and why these variables are important, including any potential improvements.	Explanation of the created variables.	No mention of variable creation.	10
Create model				**50**
Logistic loss model performance	Outstanding performance judged by loss function.	Average performance judged by loss function.	Poor performance judged by loss function.	20
Squared error model performance	Outstanding performance judged by loss function.	Average performance judged by loss function.	Poor performance judged by loss function.	20
Model justification	Discussion of viable alternatives and why selected model is superior, including comparison between other models and inclusion of different variables.	Highlight strengths and weaknesses of proposed model, but no discussion of alternatives.	No model justification or discussion of alternative models.	10
Explain customer behavior				**15**
Customer behavior	Comprehensive quantitative and qualitative discussion of models explanation of customer behavior	Limited quantitative and qualitative discussion of models explanation of customer behavior	No discussion of ability to model customer behavior	10
Assess model shortcomings	Assessment of strengths and weakness of the approach for modeling customer behavior	Assessment of strengths or weakness of the approach for modeling customer behavior	No discussion of model strengths or weaknesses	5
Written and Oral				**20**
Oral presentation	Information presented in logical sequence. Presentation includes clear articulation, proper volume, eye contact with audience, and correct use of time.	Large use of "crutch words," confusing sequencing or ideas, or poor transitioning between speakers.	Presenters too quiet to hear or looking at and reading from slides. Presentation considerably too short or too long.	10
Written article	Logical sequencing with well-developed paragraphs. No major grammatical errors.	Confusing article sequencing and/or a few grammatical errors.	Poor sentence or paragraph structure, and major grammatical mistakes.	10

APPENDIX D: CONSUMER SPENDING PROBLEM: LEARNING OUTCOMES

At the end of the project students should be able to:

- Define and implement data handling techniques and variable creation procedures,
- Develop a predictive model for customer behavior,
- Justify predictive models ability to explain customer behavior, and
- Present statistical findings in written and oral forms.

[Received December 2014. Revised August 2015.]

REFERENCES

American Statistical Association (2012), "Strategic Plan, Theme: The ASA as 'The Big Tent for Statistics'," Alexandria, VA: American Statistical Association. Available at *http://www.amstat.org/about/strategicplan.cfm* [399]

American Statistical Association Undergraduate Guidelines Workgroup (2014), "2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science," Alexandria, VA: American Statistical Association. Available at *http://www.amstat.org/education/curriculumguidelines.cfm* [397]

Blake, T. K. (2005), "Journaling; An Active Learning Technique," *International Journal of Nursing Education Scholarship*, 2, 1–13. [405]

Box, J. F. (1987), "Guinness, Gosset, Fisher, and Small Samples," *Statistical Science*, 2, 45–52. [400]

Brown, E. N., and Kass, R. E. (2009), "What Is Statistics?," *The American Statistician*, 63, 105–110. [401]

Burrows, V. A., McNeill, B., Hubele, N. F., and Bellamy, L. (2001), "Statistical Evidence for Enhanced Learning of Content through Reflective Journal Writing," *Journal of Engineering Education*, 90, 661–667. [405]

Cannon, A., Hartlaub, B., Lock, R., Notz, W., and Parker, M. (2002), "Guidelines for Undergraduate Minors and Concentrations in Statistical Science," *Journal of Statistics Education*, 25, 4–12. [397]

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action* (Vol. 22), ed. L. A. Steen, Washington, DC: Mathematical Association of America, pp. 3–43. [397]

Davidian, M., and Louis, T. A. (2012), "Why Statistics?" *Science*, 336, 12. [399]

Deming, W. E. (1940), "Discussion of Professor Hotelling's Paper," *The Annals of Mathematical Statistics*, 11, 470–471. [398]

Elder, L., and Paul, R. (2007), *The Thinker's Guide to Analytic Thinking: How to Take Thinking Apart and What to Look for When You Do: the Elements of Thinking and the Standards They Must Meet* (Vol. 16), Dillon Beach, CA: Foundations of Critical Thinking. [398]

GAISE (2005), "Guidelines for Assessment and Instruction in Statistics Education, College Report," available at *http://www.amstat.org/education/gaise* [397,404]

Gal, I., and Ograjenšek, I. (2010), "Qualitative Research in the Service of Understanding Learners and Users of Statistics," *International Statistical Review*, 78, 287–296. [397]

Gould, R. (2010), "Statistics and the Modern Student," *International Statistical Review*, 78, 297–315. [397]

Hardin, J., Hoerl, R., Horton, N. J., and Nolan, D. (2015), "Data Science in the Statistics Curricula: Preparing Students to 'Think with Data'," *The American Statistician*, 69, this issue. [399]

Hemlo-Silver, C. E. (2004), "Problem-Based Learning: What and How do Students Learn?" *Educational Psychology Review*, 16, 235–266. [399]

House, L., and Leman, S. (2014), "Analytical Attire," available at *http://magazine.amstat.org/blog/2014/01/01/analytical-attire/* [401]

Kenworthy, A. L., and Hrivnak, G. A. (2014), "To Rubric or Not to Rubric: That Is the Question," *Journal of Management Education*, 38, 345–351. [405]

Leek, J. (2013), "What Statistics Should Do About Big Data: Problem Forward Not Solution Backward," available at *http://simplystatistics.org/2013/05/29/what-statistics-should-do-about-big-data-problem-forward-not-solution-backward/* [399]

Leman, S. C., and House, L. (2012), "Improving Mr. Miyagi's Coaching Style: Teaching Data Analytics with Interactive Data Visualizations," *CHANCE*, 25, 4–12. [398]

Lohr, S. (2009), "For Today's Graduate, Just One Word: Statistics," available at *http://www.nytimes.com/2009/08/06/technology/06stats.html* [397]

Moore, D. S. (1997), "New Pedagogy and New Content: The Case of Statistics," *International Statistical Review*, 65, 123–137. [397]

Neumann, D. L., Hood, M., and Neumann, M. (2013), "Using Real-Life Data When Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course," *Statistics Education Research*, 12, 59–70. [398]

Nolan, D., and Speed, T. (1999), "Teaching Statistics Theory Through Applications," *The American Statistician*, 53, 370–375. [399]

Nolan, D., and Temple Lang, D. (2009), "Approaches to Broadening the Statistics Curricula," in *Quality Research in Literacy and Science Education*, eds. M. C. Shelley II, L. D. Yore, and B. Hand, Berlin: Springer, pp. 357–381. [398]

Pfannkuch, M., Regan, M., Wild, C., and Horton, N. J. (2010), "Telling Data Stories: Essential Dialogues for Comparative Reasoning," *Journal of Statistics Education*, 18, 1–38. [398]

Reddy, Y. M., and Andrade, H. (2010), "A Review of Rubric Use in Higher Education," *Assessment & Evaluation in Higher Education*, 35, 435–448. [405]

Rhodes, T. L. (ed.) (2010), *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*, Association of American Colleges and Universities. Available at *http://aacu.org/* [405]

Saraiya, P., North, C., and Duca, K. (2005), "An Insight-based Methodology for Evaluating Bioinformatics Visualizations," *IEEE Transactions on Visualization and Computer Graphics*, 11, 443–456. [398]

Saraiya, P., North, C., Lam, V., and Duca, K. (2006), "An Insight-based Longitudinal Study of Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics*, 12, 1522–1522. [398]

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. [406]

Utts, J. (2014), *Seeing Through Statistics*, Belmont, CA: Cengage Learning. [403]

Wells, H. G. (1903), *Mankind in the Making*, London: Chapman & Hall, Ltd. [397]

Wild, C. J., and Pfannkuch, M. (1999), "Statistical Thinking in Empirical Enquiry," *International Statistical Review/Revue Internationale de Statistique*, 67, 223–248. [398]