CHAPTER 1

Introduction to the ODA Paradigm

ODA—pronounced with a long "O" sound ("oh-dah")—is the short way of referring to the "optimal data analysis" paradigm. This new statistical paradigm is simple to learn, the software is easy to operate, and the findings of ODA analyses have intuitive interpretations. Nonetheless, the paradigm and the software are powerful and rich. This book describes the ODA paradigm and software and demonstrates how to apply ODA in the analysis of data. Everything needed to understand ODA is contained within this book, and all ODA analyses within this book can be accomplished using the accompanying software. Only moments away from jumping directly into the fray, we pause briefly to address four questions frequently asked by beginner and expert alike as they ponder the merits of learning ODA.

What Is ODA?

ODA is a new statistical paradigm—a quantitative scientific revolution, so to speak. Perhaps the best way to illustrate what this means is by example.

The ordinary least squares (OLS) paradigm maximizes a *variance ratio* for a given sample, and includes analyses such as t test, correlation, multiple regression analysis, and multivariate analysis of variance. If one wishes to maximize a variance ratio, then the OLS paradigm is required, obviously, because that is what it does. That is, maximizing variance ratios is what the "formulas" that compute t, F, and r actually accomplish for a given sample (Grimm & Yarnold, 1995, 2000).

In contrast, the maximum likelihood (ML) paradigm maximizes the *value of the likelihood function* for a given sample. This paradigm includes analyses such as chi-square, logistic regression analysis, log-linear analysis, and structural equation modeling. If one wishes to maximize the value of the likelihood function, then the ML paradigm is required (Grimm & Yarnold, 1995, 2000).

In contrast, ODA maximizes the *accuracy* of a model. As a simple example, imagine we wished to assess whether two groups—Group A and Group B—of independent observations can be discriminated on the basis of their score on a test. ODA identifies the model that uses the test score in a manner such that it discriminates members of A versus B with theoretical maximum possible accuracy. To understand how this is accomplished, recall that the model—*any* model, *every* model—can actually physically be used to compute each observation's "score" via an equation or "formula." The resulting score is then considered with respect to the decision criteria of the particular procedure and a prediction is then made on the basis of the model. In the present example, in some instances the model (regardless of the methodology by which it was developed) will predict that an observation is from Group A. Other observations will be predicted to be from Group B. Every time the predicted group membership status of an observation is correct—the same as the actual group membership status, a point is scored. An incorrect prediction scores no points. Obviously, the largest number of points that it is possible to attain for a sample of N observations, in theory, is equal to N, the number of observations in Groups A and B that are classified by the model. Clearly, this maximum score is only possible if all observations are correctly predicted to be from A or B by the model. The minimum score possible is obviously zero points, in which case all observations are *incorrectly* predicted to be from A or B by the model.

By definition, an ODA model achieves maximum possible accuracy for a given sample of data, in the sense that no other model that is based on the test score can achieve a superior number of points. All possible alternative models are (explicitly or implicitly) evaluated to literally prove this, which is one reason why ODA is "computationally intensive." As OLS maximizes a variance ratio for a given sample of data, and as ML maximizes the value of the likelihood function for a given sample of data, ODA maximizes the accuracy of the model for a given sample of data. Of course, if different observations can be weighted by a different number of points, for example if a "natural" weighting metric, such as time, weight, or cost is available, then weighted accuracy may be maximized (or cost minimized), as may be desired by the operator.

Every type of analysis (i.e., every specific configuration of data, constraints, and hypotheses) that can be conducted in the OLS and ML paradigms can be conducted in the ODA paradigm. The ODA paradigm can conduct many analyses that one simply cannot do using either OLS or ML paradigms. ODA is much more general, and much more encompassing of different data, constraints, and hypothesis configurations, than are the alternative statistical paradigms. The ODA paradigm is, quite literally, "new and improved." Using this paradigm, and only using this paradigm, is one able to identify maximally accurate models for a given sample.

Why Is ODA Superior to Other Data Analysis Programs?

The ODA paradigm is vastly superior to alternative paradigms. Consider first, *conceptual clarity*. For every problem analyzed via ODA there is one precise, optimal analysis—a specific given data configuration and hypothesis dictates the exact nature of the ODA model that is appropriate. Using traditional statistics, in most applications several different analyses are feasible—all reflecting some degree of lack of fit between their required underlying distributional assumptions and the actual character of the data. Consider second, *ease of interpretation*. Every ODA analysis provides the same intuitive goodness-of-fit index: for every ODA analysis an index is computed on which 0 reflects the accuracy expected by chance for the sample, and 100 reflects perfect accuracy. Using traditional statistics, different analyses provide different goodness-of-fit indices that are non-intuitive and that are not directly comparable across procedures. Consider third, *maximum accuracy*. Every ODA analysis provides a model that guarantees maximum possible accuracy. Using traditional statistics, no analysis provides a model that explicitly guarantees maximum possible accuracy. Consider fourth, *valid Type I error*. No ODA analysis requires any simplifying assumptions, and p is always valid and accurate—a permutation probability derived via Fisher's randomization method, invariant over any monotonic (i.e., transformed values either always increase or always decrease) transformation of the data. Traditional analyses require simplifying assumptions (e.g., normality), p is only valid if the required assumptions are true for one's data, and p may be inconsistent over transformations of the data.

An obvious advantage of ODA software is *availability* (the first and currently the only software available that performs ODA). There are many good packages available for performing OLS and/or ML analysis, and many are an order of magnitude more expensive than the ODA book/software. Comparing software across paradigms, the ODA software is superior to software of earlier paradigms for two important reasons. Consider first, *ease of learning and teaching*. Everything needed to understand the ODA paradigm and analyze data is discussed in this book. Many courses, books, and articles are needed to understand traditional statistics and to correctly operate associated software, requiring years of study. Consider second, *ease of use*. Most types of ODA analyses require the same basic set of seven programming commands. Using traditional procedures requires learning of numerous—hundreds—system-unique programming commands.

Who Is the Audience for This Book and Software?

A "must-have" tool for all quantitative researchers, ODA is the only general-purpose statistics software that explicitly maximizes predictive accuracy. Many articles were published during the software's beta testing in areas such as developmental psychology, pediatrics, social psychology, physiology and physiological psychology, allergy-immunology, cardiology, emergency medicine, clinical psychology, criminal justice, education, industrial–organizational psychology, political science, social work, sociology, economics, AIDS research, women's studies, biology, general internal medicine, psychiatry, management science, rehabilitation medicine, neurology, pharmacy, marketing, and oceanography. Conceptually straightforward examples in the book represent a myriad of substantive areas drawn from many disciplines, and many examples can only be solved using ODA. Regardless of discipline, researchers ranging from new students to seasoned professionals will marvel at how easily and rapidly ODA theory and software can be mastered—and can help them master their data. Knowledge needed to conduct and interpret a cornucopia of different statistical analyses is provided in easy-to-follow steps in the book—the "no formula" presentation is engineered to maximize conceptual clarity. Similarly, the "consistent-across-design" software encourages ease of learning as well as efficiency—because most analyses require the same basic set of seven commands. The combination of ultrarapid learning and maximum accuracy results, made possible only by ODA, promises levels of efficiency that researchers only dream of. Welcome to the new revolution in statistics!

It is recommended that the person wishing to master this book and software system have familiarity with some basic statistical concepts (e.g., what is a variable or what is a sample of independent observations), and some basic computer skills (e.g., how to copy files from a CD to a new directory). No prior knowledge of any statistical procedures is required to master this book and software. On the other hand, the deeper one's understanding of alternative procedures and software systems, the greater one's appreciation for the ODA paradigm and software. For all users, the more one uses the software, and the more different types of data that one analyzes, the more one will appreciate the conceptual clarity and ease of use of ODA.

How Should the Reader Use This Book?

In our opinion, the best way to learn and master the ODA paradigm is to read this book systematically (begin at the beginning), work each sample problem when it is presented, and then pause before continuing in order to use your own (real or artificial) data to practice each particular method. If you do not have data for the example at hand, another option is to manipulate the data that are provided for the example and note the effect of your manipulation on the results. This book is linear—every page builds on ideas presented in prior pages. At every point at which an analysis technique is discussed, each technique may, of course, be used to analyze unpublished scientific data, and the findings published in leading scientific journals in any substantive area (we do plenty of this). Of course, data already published using other methods may be reanalyzed using ODA, and comparisons of the findings published in leading scientific journals, in substantive applied journals, or in applied statistics and methodology journals (we do plenty of this). Some people will apply this software to data in ways that are not yet covered in this book, but that *will* be covered in a future edition of this book once we read their papers.

At this point we are ready to begin. We hope that you have decided to take this journey, as we have little doubt that you will be pleased when you complete it. We begin at the beginning.

Basic Steps and Key Concepts

The first step of any ODA analysis is to define what it is that you wish to predict. In ODA, a *class variable* is any random variable that may attain two or more levels: the levels reflect the phenomena that one desires to predict. In conventional statistics, class variables are often referred to as dependent variables. Example class variables include health status (sick, healthy), socioeconomic status (lower, middle, upper), or change in price of a commodity futures contract (lower, unchanged, higher). The *category level* of a class variable is the number of different values or levels that the class variable may attain. Thus, as defined above, health status is a two-category class variable, and socioeconomic status and investment outcome are three-category class variables. During analysis, class categories are identified via *dummy-codes*: Class 0, Class 1, Class 2, and so forth.

In the best of all worlds, the class categories should represent qualitatively distinct phenomena, conditions, or states. For example, biological sex—male versus female—involves two qualitatively different categories, and is a highly stable class variable. Other variables, however, are less ideally suitable. For example, imagine that one wished to predict mortality status: alive versus dead. Further imagine that, to do this, a sample of patients was prospectively followed for one year, after which time patients were classified as either dead or alive. It is possible that, had the study been continued for one additional day, some of the patients classified as being alive would instead have been classified as dead. Another example of an imperfect class variable is age category. For example, suppose that we are interested in comparing geriatric (65 years or older) versus non-geriatric (younger than 65 years) people.

Further imagine that a participant in our study will turn 65 years of age in one more day—or in two more days. The greater the potential instability or unreliability of a class variable, the more "fuzzy" it is. Instability in the class variable that occurs near the cutpoint (e.g., 65 years in this example) is problematic, theoretically limiting the upper bound of accuracy that it is possible for an ODA model to attain.

Finally, pragmatically speaking, it is a good idea for those learning the ODA paradigm to begin by studying class variables having two category levels: so-called dichotomous or binary class variables. This is because ODA is computationally intensive—problems become much more difficult to solve as the number of category levels increases. Furthermore, in the event that one's ODA model is unable to perfectly predict or nearly perfectly predict a binary class variable—which unfortunately is usually true—one does not want to increase the complexity of the problem by including additional "shades of gray." Of course, it is possible that adding an intermediate category—"undecided"—might enhance model performance in those instances in which some participants cannot be reliably classified into either type of the dichotomy (Dr. Loretta J. Stalans, personal communication, 2002).

The second step of any ODA analysis involves defining the set of potential predictor variables. In ODA an *attribute* is any random variable that can attain two or more levels. Attributes are used to predict the class variable. In conventional statistics, attributes are often referred to as independent variables. In addition to defining the attributes, one must identify their *metric* (cf. Velleman & Wilkinson, 1993). For ODA the primary distinction is *qualitative* versus *ordered* attributes. Qualitative attributes may be *binary*—such as smoking status (smoker, non-smoker), or *polychotomous*—involving three or more qualitatively distinct, unordered categories: occupation (unemployed, clerk, lawyer) or investment decision (buy, hold, sell), for example. Ordered attributes—for which increasing scores indicate increasing values of the phenomenon, may be *ordinal* (e.g., a rating made using a 7-point Likert-type scale), *interval* (e.g., score on a college board examination), or *ratio* (e.g., time). Finally, it is important to determine whether there is an *a priori hypothesis* relating the attribute and the class variable (discussed ahead). Of course, as is true in conventional statistics, the decision concerning which variable serves as an attribute and which serves as a class variable is usually arbitrary. For example, in studying substance abuse in schizophrenia, Mueser et al. (1990) treated substance abuse as the independent variable, and Mueser, Yarnold, and Bellak (1992) treated substance abuse as the dependent variable.

The third step of any ODA analysis involves specification of appropriate *weights:* weights are used so that an obtained ODA model mirrors reality. For example, if we wished to obtain a model to guide stock market investment decision-making, then we should weight the observations (i.e., the different days in the study period) by the amount of money that the stock went up or down—because we wish the model to be most accurate on days that the price changes substantially—in order to maximize profit. Without weights, the ODA model would simply maximize the number of correct decisions: the nonweighted model might get more buy/sell decisions correct than the weighted model, but the weighted model could still be much more profitable by emphasizing accuracy on highly volatile days.

There are two types of weights. The first type of weight is called *prior odds:* analogous to the use of *antecedent probability* or *base rate* in Fisher's discriminant analysis, weight all n_c observations in class category c by the value $1 / n_c$ (e.g., Greenblatt, Mozdzierz, Murphy, & Trimakas, 1992; McLachlan, 1992; Meehl & Rosen, 1955; Rorer & Dawes, 1982; Widiger, 1983). For example, imagine that we wished to predict whether a patient hospitalized with pneumonia would survive. If we obtained an ODA model for predicting mortality attributed to pneumonia, but failed to consider the base rate for mortality among hospitalized pneumonia patients, the model might overestimate the number of patients who died (most cases of common pneumonias are not fatal). Later it will be shown how one can estimate the classification performance obtained by one's final model for all possible base rates (this is known as assessing the *efficiency* of one's model). Then, to determine how well the model will perform in a given sample (e.g., zip-code-defined geographic area), one need only consider the base rate for that sample and consult the curve derived in the efficiency analysis.

The second type of weight is a quantitative assessment of the value or importance of the attribute to the decision-maker. For example, in an application involving predicting daily movement in the price of a stock, we would weight observations by the dollar value of the change in stock price. Were we to construct an ODA model for this application without considering the return, the model would maximize our ability to predict the direction of movement of the stock correctly. In the absence of a return weight, we might be correct, for example, 85% of the time that we predict a movement in stock price—and yet still lose money because the model misclassified the days on which the price of the stock changed the most. However, specification of a return weight would obtain an ODA model that maximized the amount of dollars correctly predicted: although overall predictive accuracy might decrease (e.g., to 40% correct predictions), the model would seek correct prediction of the times that the stock value changed substantially, and thereby maximize profit.

Functionally, ODA models are maps between values on the attribute and predicted class memberships. An example of an ODA model for a qualitative attribute is: If the person is a smoker, then predict disease; if the person is a non-smoker or an ex-smoker, then predict health. An example of an ODA model for an ordered attribute is: If the person smokes 4 or more cigarettes per day then predict disease; otherwise predict health. Essentially, an ODA model is a decision rule for predicting class membership status on the basis of the attribute. For a given (*training*) sample of data, this rule yields the theoretically maximum-attainable level of (weighted or nonweighted) *percentage accuracy in classification* (abbreviated as PAC). An observation is *correctly classified* when the actual and predicted class membership are the same. For example, the model predicts death and the person dies. An observation is *misclassified* when the actual and predicted class memberships differ. For example, the model predicts death and the person lives. The number of correctly classified observations in the sample is called the *optimal value*. The theoretical maximum possible value for both weighted and nonweighted PAC is 100%.

Historical Perspective

An application involving a single attribute is referred to as a univariable ODA (UniODA) problem, and an application involving more than one attribute is known as a multivariable ODA (MultiODA) problem. Initial exploration of UniODA occurred in the 1950s, as researchers began to address the problem of how to best assign observations into one of two mutually exclusive categories on the basis of their score on a single test or on a single composite index based on several tests. As is true of modern UniODA methodology, early research began by considering that for a two-category, single attribute problem, observations are assigned to one or the other category on the basis of a cutting score (CS). Observations with scores on the test that exceed the CS are assigned to one category, and observations with scores that fail to exceed the CS are assigned to the other category. An optimal cutting score (OCS) has the property that, compared to other possible CSs, the OCS results in the maximum proportion of correct decisions.

The earliest programmatic discussions of the definition and use (for classification purposes) of the OCS that we have located were presented by Duncan, Ohlin, Reiss, and Stanton (1953) and by Meehl and Rosen (1955). They addressed modern UniODA concepts, such as maximization of overall classification accuracy; maximization of the sensitivity of the model for a user-specified category of the class variable; the need for weighting by prior odds in imbalanced applications involving different numbers of observations in the different categories of the class variable; and constrained optimization (see also Alf & Abrahams, 1967; Blumberg, 1957; Dawes & Meehl, 1966; Luce & Raiffa, 1957; Rosen, 1954). However, the computational effort required to obtain an OCS by hand rendered this procedure infeasible for most real-world samples, so a suboptimal normal-based heuristic for obtaining an OCS was developed (Cureton, 1957; Darlington & Stauffer, 1966a) that allowed weighting either by prior odds (Darlington & Stauffer, 1966b; Dawes, 1962; Golden & Meehl, 1979; Rorer & Dawes, 1982; Rorer, Hoffman, LaForge, & Hsieh, 1966) or misclassification cost (Rimm, 1963; Rorer, Hoffman, & Hsieh, 1966).

Research in these areas declined steeply in the 1970s and remained dormant until the late 1980s. Although the reason(s) underlying this decline are not fully understood, we speculate that two important contributing factors included (a) the difficulty involved in hand-computing the OCS, particularly in weighted applications; and (b) a shift in focus—as computers became more accessible and capable—to the dramatically different and more difficult problem of obtaining MultiODA solutions (see Epilogue). Because of the enormous computational resources required to solve MultiODA problems, most of the latter research involved relatively limited experimental comparisons of the classification performance achieved using a MultiODA model versus suboptimal multiattribute classification models such as might be obtained using logistic regression or Fisher's discriminant analysis. It is thus not improbable that the failure of prior research to discover the broader implications of the OCS concept—that is, as concerns the existence of the ODA paradigm—was due in large part to the unavailability of an efficient methodology for solving ODA problems.

The recent discovery that exact statistical distributions underlie ODA models—which are so flexible that the models may be specified to precisely reflect the finest details in any experimental design—paved the way for the discovery of ODA as a general statistical paradigm. However, the basic methodology used to identify ODA models—that is, models that specifically maximize return or minimize cost for a given scenario—is used in a variety of academic disciplines. For example, the basic technical concept involved in identifying the optimal (most accurate) ODA model is known as maximum feasible subsystems of linear inequalities (Max FS), along with the closely related concept of minimum irreducible infeasible subsystems (Min IIS). Both Max FS and Min IIS have been appearing with increasing frequency in the literature of operations research (Chinneck, 2001; Chinneck & Dravnieks, 1991; Gleeson & Ryan, 1990; Mangasarian, 1994; Marcotte & Savard, 1995; Soltysik & Yarnold, 1994b; Van Loon, 1981), computer science (Amaldi, 1995; Amaldi, Pfetsch, & Trotter, 2003), machine learning (Bennett & Bredensteiner, 1997; Parker & Ryan, 1996), and perceptrons and neural networks (Hastad, 2001; Mattavelli & Amaldi, 1995). Examples of recent applications include radiation therapy planning (Sadegh, 1999), speech translation (Kussner & Tidhar, 2000), computational biology (Wagner, Meller, & Elber, 2002), and digital television broadcasting (Rossi, Sassano, & Smriglio, 2001). Clearly, the stage is now set for the application of this new paradigm in statistical analysis to a host of empirical data.

Thirty Hypothetical Applications

What types of problems can ODA solve? To begin to answer this question and illustrate the flexibility of the approach, consider the following hypothetical applications for which ODA is the optimal analytic methodology.

Astrology

Imagine that one is interested in optimally predicting personality differences between people born under different astrological signs. To address this issue, one might conduct a study in which a large random sample of persons whose zodiac signs were known all take a series of personality tests. With these data ODA can determine, with maximum possible PAC, (a) which signs of the zodiac are associated with relatively high or low scores on the different personality tests relative to other signs of the zodiac, and (b) exactly what values on the personality tests to use as an operational definition of relatively high versus low scores. Other attributes, such as dominance or nurturance, have also been hypothesized to differentiate people born under different astrological signs and could be investigated. ODA can also determine whether the findings of this research are consistent for different data samples, such as reflected by data collected from different races, genders, religious orientations, and socioeconomic status levels. If the results are not consistent across all samples, ODA can determine which samples (if any) share which ODA model.

Astronomy

When will there be relatively heavy meteorite showers? To address this question, one might conduct a study in which data are collected from Hawaiian observatories for three years. Each day, the presence or absence of unusually active meteorite showers is recorded. For the sake of illustration, imagine that the criterion for unusually heavy showers is more than 15,000 meteorites per hour in one's observation area. In addition, the level of solar flare activity is recorded (a time lapse might be appropriate) as the attribute. With these data ODA can determine what level of solar flare activity predicts heavy meteorite showers with maximum PAC. If the actual number of meteorites per hour is recorded, ODA can determine what level of solar flare activity optimally predicts the number of meteorites per hour. If the mean mass of the meteorites can be validly estimated, ODA can determine the level of solar flare activity that optimally predicts meteorite mass per hour. Other attributes, such as the level of sunspot activity, Earth's magnetic activity, or the distance of the Earth from the sun, moon, other planets, or asteroid belts, might be investigated. ODA can also determine whether findings generalize to other observatories, and can identify groups of observatories that share a consistent ODA model.

Beer Brewing

What is the best recipe for beer? To address this issue, one might conduct a study in which data are collected from a random sample of beer drinkers. Each observation is randomly assigned to try one of three different brews of beer that differ only in the weight of hops added to the recipe: one pound, two pounds, or three pounds of hops. Observations rate whether the beer tastes good or bad. With these data ODA can determine, with maximum possible PAC, which brew(s) are associated with good versus bad ratings, and which brew(s) maximize the percentage of the sample responding with good evaluations. If observations also provide a numerical rating, say on a 10-point scale, of the desirability or goodness-of-taste of the beer, ODA can determine which brew(s) are associated with maximum overall satisfaction. If the amount of beer consumed by the observations is recorded, ODA can determine which brew(s) are associated with maximum consumption. Other attributes that might predict taste ratings, such as the type of water used, method of fermentation, length of storage, effect of different combinations of other ingredients in the recipe, and alcohol content of the beer could also be investigated. ODA can also determine whether the findings generalize to other data samples, such as different races, genders, and ages. If the findings are inconsistent, ODA can determine which samples share consistent ODA models.

Bird Watching

Imagine that one is interested in determining the optimal time(s) in the morning for spotting endangered birds. To address this issue, one might conduct a study in which data are collected on the hour between 5 a.m. and noon every day for two months. Each hour a survey of a wildlife refuge is conducted. If at least one endangered bird was seen, the hour is deemed a success; if no endangered birds were seen, the hour is deemed a failure. With these data, ODA can determine the optimal time(s) in the morning to spot endangered birds. If the total number of endangered birds seen each hour is recorded, ODA can determine the time(s) in the morning in which the maximum total number of endangered birds can be seen. If the total number of different endangered species seen each hour is recorded, ODA can determine the time(s) in the morning in which the maximum variety of endangered birds can be seen. Other attributes that might be investigated include the air temperature, wind velocity and direction, presence of precipitation, presence of other animals, or the amount of ground cover. ODA can also determine whether findings are consistent for different genders or species of bird for different seasons. If the findings are inconsistent across different samples, ODA can determine which samples (if any) share which ODA model.

Credit Collection

Imagine that one is interested in accurately predicting which of a pool of observations might best be hounded for past-due bill payment. To address this issue, one might conduct a study in which data are collected for a comprehensive sample of all customers of a large department store over a span of eight consecutive months. The class variable is whether the observation paid the outstanding bill by the end of the study period. Imagine that the observation's age is the attribute. With these data, ODA can determine the age(s) that best predict bill payment. If the amount of the bill is recorded, ODA can determine the age(s) that pay the greatest amount or percent of money. Other attributes, such as type of solicitation message, marital status, income, or past credit history might also be analyzed. Combinations of these and other attributes may be used to define new class variables, or be treated as multiple samples to evaluate consistency.

Credit Screening

Imagine that one is interested in minimizing the need for credit collection, and desires an optimal model for determining the credit-worthiness of applicants for credit cards or loans. To address this issue, one might conduct a study in which data are collected for a comprehensive consecutive sample of 200 applicants for a credit card, all of whom were accepted for the purposes of the study and then followed for one year. At the end of the year it is determined whether each observation had been a positive or negative revenue source for the lender (class variable). Following the lead of the preceding example, the attribute is the age of the observation. With these data ODA could determine the age(s) that attain positive or negative status with greatest relative frequency. If the amount of the profit, or the absolute value of the loss, associated with each observation is recorded, ODA can determine the age(s) that return the greatest amount (or percentage) of money. Other attributes, such as gender or ethnicity, marital or parental status, religious or political affiliation, or income or past credit history might also be analyzed—so long as it is legal and ethical. Combinations of these and other attributes could also be used to define new class variables, the generalizability of the findings could be evaluated, and consistent samples could be identified.

Criminal Justice

Which attributes accurately predict whether a defendant is convicted in a criminal trial? To address this issue one might conduct a study in which data are collected from random samples of 50 criminal trials from each of 10 counties in a single state. The attributes might include gender and age of the prosecutor, defender, and client, type of evidence emphasized, presentational manner (e.g., cold and calculating versus animated and dramatic), prior record of the defendant, characteristics of the judge or jury, and type of crime. With these data, ODA can determine the ability of each attribute to accurately predict the disposition of the trial. At the user's discretion, models may be determined that obtain optimum PAC over all 10 counties when they are considered as a single sample, or, alternatively, that obtain the optimum mean PAC when applied to each of the 10 counties individually. Were the time taken to decide on the guilt or innocence of the defendant recorded, or to conduct the trial, ODA could find the model that best predicted relatively fast or slow decisions. Were cost data available, ODA could find the model that best predicted expensive, inexpensive, or cost-effective (defined by the user) trials. Were sentencing information available, ODA could find the model that best predicted the severity (length) of sentences. Were recidivism (returning to a life of crime after being released from prison) data available, ODA could find the model that minimized the relative frequency of recidivism, or the weighted cost of the recidivism (e.g., in dollars for crimes committed against property; in lives or rapes for crimes committed against people). And, were data available concerning reintegration of the survivors of this study into mainstream society, ODA could help identify variables that best predict success and that best predict failure in this context.

Dating

Imagine that one is interested in discovering factors that best predict one's personal dating bliss. To accomplish this, one might conduct a study in which a person collected data concerning all dates (observations) occurring over some designated time period. The class variable is an indication of whether one desired to redate the observation. A host of personally salient variables might serve as attributes. ODA can determine the ability of each attribute to optimally predict the desirability of one's date. Also, dates may be rated by any desired subjective or objective measure, and ODA used to discover the attributes that best predict the weighted satisfaction. Note that this methodology can be used to determine an optimal model for any desired personally salient issue. Also, investigators, counselors, educators, and other professionals who work with single cases could use this method to discover statistically reliable markers of change, for example, in scale scores or behavioral observations of behavior, in single-case longitudinal series (cf. Yarnold, 1992).

Direct Mail Advertising

How can one determine the most productive people to whom to direct-mail an advertisement? To address this issue, one might conduct a study in which data are collected from a random sample of people about whom psychoethnographic data (attributes) were available, and who were mailed an advertisement. Six months after the mailing, observations would be coded as either having responded or not responded to the advertisement. ODA could be used to optimally evaluate the ability of each attribute to predict whether observations responded to the advertisement. If the amount of money each observation spent is recorded, ODA could be used to optimally evaluate the ability of each attribute to predict the return of the mailing. Data from different samples, such as reflected by different product lines, could be used to evaluate the generalizability of the findings, and samples for which consistent findings emerged could be identified.

Driver Licensing

Imagine that one is interested in improving the quality of licensed automobile drivers (to the extent possible) by optimally determining the value constituting the minimum passing score (MPS) on the written examination section of the licensing application. To address this issue, one might conduct a study in which data are collected from all people who lived and worked in a given county, and who applied for and received a driver's license during the past year. Each observation would be tracked using mailed questionnaires, telephone interviews, and police and insurance files for one year after receiving the license. It would be recorded whether, during the year, a parking violation or moving violation citation, accident, injury or fatality report had been generated naming the observation as the culpable party. Using these class variables, ODA could determine the value of the MPS that resulted in optimal prediction of any of the class variables, or in optimal prediction of any of the categories constituting any of the class variables (e.g., specifically, fatalities). Had the number of each event reflected by the class variables, and/or their corresponding direct (repairs, litigation) or associated (time off from work, physical and emotional trauma) costs been recorded, ODA could determine the value of the MPS that resulted in optimal prediction of the number or cost of any of the class variables/categories. Other attributes that might predict the class variables, such as age, geographic region (urban, rural), or prior driving record might also be investigated. Were data available from multiple samples, such as reflected by data collected from multiple counties or from professional drivers, ODA could optimize the models for the pooled data or simultaneously and separately across the samples and could identify the samples for which the findings were consistent.

Epidemiology of AIDS

How can one discover factors that predict progression from testing positively for HIV infection to development of AIDS? To address this issue, one might conduct a study in which data are collected from a comprehensive consecutive series of individuals who tested positively for HIV infection at a clinic. People are tracked for exactly one year following initial diagnosis, at which time it is recorded whether each person had AIDS (the class variable). The attribute of greatest interest is the number of T4 initiator cells. ODA can determine the critical number of T4 initiators that optimally predicted whether a person in this sample developed AIDS within the first year of diagnosis of HIV infection. A host of other attributes, such as drug use, blood chemistry measures, dietary, physical, and sexual behaviors, psychoethnographic measures, and so forth could also be used to predict the development of AIDS. Were data concerning hospitalizations, fatalities, or treatment costs for each person recorded, ODA could determine the number of T4 initiators that optimally predicted the number of hospitalizations or fatalities or the cost of medical care. Were data collected from multiple clinics, the generalizability of the results could be determined, and samples with consistent findings could be identified. Were data from many clinics available, ODA could be used to discover factors that optimally discriminate clinics with consistent findings in the prior analysis from clinics with inconsistent findings.

Farming

What is the optimal amount of cow manure to mix into soil to ensure that as many watermelons as possible grow to at least twenty pounds? To address this question, one might conduct a study in which the land available to conduct this research is randomly subdivided into 25 equivalently sized plots, each sufficiently large to support ten mature vines. Five different levels of manure (5%, 10%, 15%, 20%, and 25%) are to be contrasted, and five plots of land are randomly assigned to (and mixed at) each manure level. At the end of the growing season, all watermelons are harvested and weighed. If a watermelon achieves a weight of twenty pounds or more, it is classified as a success; otherwise, it is classified as a failure. ODA can determine the level(s) of manure that optimally predict success versus failure or that predict either successes or failures with maximum PAC. ODA can also determine the level(s) of manure that optimally predict the total weight of watermelons harvested. If the sales prices of the watermelons are recorded, ODA can determine the level(s) of manure that optimally predict the total return on the harvested watermelons. Other attributes, such as the amount of water or direct sunlight or the brand of seed used could also be investigated. Data from other samples (e.g., other cash crops) could be used to evaluate the generalizability of the findings, and ODA could identify samples with consistent ODA models.

Fishing

What is the optimal lure color for catching fish? To address this issue, one might conduct a study in which data are collected for 100 hours of fishing. Each hour, a gold, silver, red, yellow, or green lure is randomly selected and used: If any fish were caught during this hour, the hour is deemed a success; if no fish were caught during the hour, then it is deemed a failure. With these data, ODA can determine which lure(s) to use in order to maximize fishing success. If the number of fish caught each hour is also recorded, ODA can determine which lure(s) to use in order to maximize the number of fish caught per hour. If the weight of each fish is also recorded, ODA can determine which lure(s) to use in order to maximize the total weight of the fish caught per hour. In addition to lure color, other attributes that might be investigated include the type of lure, time of day, air or water temperature, barometric pressure, depth of the water, weight and type of fishing line used, speed of the retrieve of the lure, or any other similar attribute. ODA can also determine whether the findings are consistent for different data samples, such as reflected by data collected from different lakes, different seasons, or different species, ages, and gender of fish. Finally, if the results are inconsistent across samples, ODA can determine which samples (if any) share which ODA model. Note that this example also works well with hunting of both animals and edible wild foods.

Gambling

Imagine that one is interested in evaluating the validity of several types of information as predictors of whether a horse will finish in one of the top three places ("in the money") in a race. To address this issue one might conduct a study in which data are recorded for all races occurring over a one-year period at a nearby track. For each race, horses are classified as either finishing in the money or not (class variable). Attributes include the number of times the horse finished in the money in the last two months, the win/loss record of the jockey in the past week, and the workout speed of the horse on the day of the race. ODA could be used to optimally determine the ability of these attributes to predict whether a horse made money. If the actual amount of money the horse won was recorded, ODA could optimally determine the ability of these attributes to predict the amount of money won by a horse. ODA could also be used to assess the generalizability of these findings across different racetracks. Conceptually similar methods could be used to determine which money market or mutual fund is the best investment.

Golfing

Imagine that one is interested in improving one's putting. To address this issue one might conduct a study in which data are collected for 500 consecutive putts made at the local nine-hole golf course. The class variable is whether the attempted putt was successful, and the attribute is the type of putter used (standard versus long-handled). For each putt, the putter used is randomly selected. ODA could be used to determine the optimal choice of putter in order to maximize one's putting success. If the length of each putt was recorded, ODA could determine the optimal choice of putter for maximizing the number of long- or short-range successful putts made. Other attributes, such as the type of grip, stance, or aiming protocol could also be studied. The findings could be generalized across different samples such as different courses or types (e.g., flat or curvy) of greens, and samples for which consistent findings emerged could be identified.

History

Imagine that one is interested in determining whether the relative frequency of reports concerning different types of aberrant interpersonal relationships published in the local newspaper fifty years ago is consistent with reports published in that newspaper today. To address this issue one might conduct a study in which the local paper is obtained for the entire years of both 1942 and 1992. Every article concerning negative (by Western standards) interpersonal events (divorce, adultery, beatings, robberies, murders, rapes) is pulled. With these data, ODA could determine whether the relative frequencies of such events within time were consistent across time, thus reflecting historical homogeneity. ODA might also discover events for which the 1942 and 1992 relative frequencies are inconsistent, reflecting historical heterogeneity. Were data collected from multiple samples, such as different newspapers or books, ODA could determine the extent to which the findings generalized, and could identify samples for which consistent findings emerged. It should be noted that one can use this technique to discover distinguishing features, if they exist, between (time-lapsed or contemporary) competing entities on the basis of the comparability of the corresponding relative frequency of the different aspects (features) that define their constitution. ODA can then be used to optimally contrast the relative frequency of the content (aspects or features) offered by, for example, different television shows, colleges, resorts, or mountain ranges.

Hostage Negotiation

Imagine that one is interested in evaluating the relative efficacy of different strategies for negotiating for the safe return of hostages. To address this issue, one might conduct a study in which data are collected for all hostage situations occurring in the United States during 1991. The class variable is whether any hostages were killed, and the attribute is whether the authorities used force in an attempt to free the hostages. With these data, ODA could determine the optimal ability of the use of force to predict whether any hostage lives would be lost. Had the number of hostage deaths been recorded, ODA could predict with maximum accuracy the number of hostage deaths resulting from the use of force. Other class variables (e.g., presence and/or amount of property damage or of kidnapper casualties; whether the situation involved kidnapper relatives) and attributes (e.g., political affiliation, number, gender, age, and cross-cultural communication and knowledge status of the kidnappers; type of demands; type and/or quantity of weapons; nature of the area where hostages are being held) could also be investigated. ODA could also assess the generalizability of the findings across multiple samples (e.g., different countries or different time periods), and, if findings were not consistent over all samples, ODA could identify samples for which consistent findings emerged.

Hurricane Forecasting

Imagine that one is interested in determining factors that predict with maximum PAC whether a hurricane will falter at sea or come crashing ashore. To address this issue one might conduct a study in which a sample of hurricanes is tracked from their inception (note that the development of full-fledged hurricanes from tropical storms constitutes an interesting class variable). Assessed at both the eye and periphery, attributes include hourly measurements of the storm speed, height, surface area, and cubic volume; direction and sustained velocity of the storm winds and of the steering winds and currents; strength, direction, and temperature of opposing frontal weather systems; water temperature and depth (and change in water temperature and depth); direction and strength of tide and saline thermoclines; location and phase of the moon; and time of day. ODA can determine the optimal ability of each attribute to predict whether a storm will come ashore. If the category of the storm (Category I [least severe] to Category V [most severe]) were recorded, ODA could optimally predict the strength of the storms that came ashore. Were data concerning the cost of the property and other resources destroyed by the hurricanes recorded, ODA could optimally predict the cost of the storms that came ashore. Of course, ODA could evaluate the generalizability of the findings across, for example, different oceans or hemispheres. Conceptually related research might also focus on other natural disasters, such as avalanches, blizzards, droughts, earthquakes, fires, floods, tornadoes, tsunamis, or volcanic eruptions. Although less exotic, such methods may facilitate optimal forecasting of precipitation (rain, sleet, snow), or of changes in weather phenomena, such as temperature, pressure, winds, or humidity. Finally, similar methods might be fruitfully used in forecasting the emergence (etiology) and spread (ontogenesis) of biological disasters, be they from land (army ant, locust), air (killer bee, mosquito), or sea (lamprey, zebra mussel).

Life Insurance

Imagine that one is interested in determining the most desirable clients (observations) to whom to attempt to sell life insurance. To address this issue, one might conduct a study in which data are collected from all people who were interviewed—in the course of regular business—by a specific insurance agent during the past six months. The class variable is whether the observation bought life insurance. Attributes that might predict the class variable included, for example, the observations' gender, parental status, age, socioeconomic status, education, occupation, perceived health status, recent experience with death, or state, trait, or somatic anxiety. Using ODA, one may determine the optimal ability of each attribute to predict whether an observation will purchase insurance. With data concerning an observation's corresponding net return (i.e., payments – costs), ODA could determine each attribute's optimal ability to predict the overall return of the insurance sold. A conceptually related problem would involve maximizing the prediction (or minimizing the occurrence) of delinquent accounts or catastrophic diseases/accidents.

Missionary Work

Imagine that one is interested in determining ways to improve the efficiency of missionary work in a specific developing country. To address this issue, one might conduct a study in which data are collected over a one-year time span for all of the Catholic missions in that country that have been active for at least ten years. At the end of the year it is determined whether each mission (a) showed an increase in the total number of registered members; (b) had any registered members enter into professional religion (class variables); or (c) had any charitable donations. Attributes that might be investigated include the length and/or orientation (pro-goodness or anti-sin) of the stereotypic sermon; psychoethnographic characteristics of the local constituency; type and quantity (measured in terms of either hours worked per week or percentage of theoretical need of one's constituency) of community service, educational programs, or extracurricular activities; availability of medical expertise, food, water, clothing, and temporary shelter; use of a direct versus an indirect recruitment approach; or the reverence of the community for traditional (indigenous) values and customs. ODA could be used to determine the optimal extent to which each attribute predicts whether or not a mission's membership increases, any of its members enter the religious service, or a charitable donation is received. Were data available concerning the actual change in the number of registered members, the actual number of registered members entering religious service, or the dollar (or other monetary index) amount of the charitable donation(s), ODA could determine the optimal ability of each attribute to predict the number of new registered members, the number of members entering religious service, or the monetary amount of charitable donations. Finally, were like data collected for missions representing other religious orientations, ODA could determine if findings generalized across religion, and, if findings did not generalize, ODA could determine religions for which consistent findings emerged.

Personnel Selection

Imagine that one is interested in using a questionnaire measure of motivation to predict whether applicants for a sales job would, if hired, be rated as being desirable or undesirable employees by their supervisors (class variable). To address this issue, one might conduct a study in which data are collected from 100 observations randomly selected from the sample of all applicants hired for a sales job under a one-year contract. At the end of the year, supervisors rate the desirability (pro or con) of the observations. With these data, ODA could determine the value of the score on the motivation test that resulted in maximum PAC when used to predict desirability ratings and could determine the value of the score that optimally predicted desirable (or undesirable) employees. Were ratings of observations made on more sensitive scales, one could use the actual ratings as weights and ODA would find the value of the score that maximized prediction of overall desirability ratings. Were objective measures of the profitability of the observations available, ODA could determine the value of the score that facilitated optimal prediction of the return of the hiring decisions. Scores on other questionnaires that assess additional factors that are theoretically relevant to job performance could also be investigated. Had data been collected for multiple samples, such as reflected by applicants for other types of jobs, ODA could determine the consistency of the findings across samples, and could identify samples for which findings were consistent. A particularly compelling example of the need for simultaneous multiple sample analysis is reflected in the legislation that regulates personnel selection practices. That is, the Equal Employment Opportunity Commission guidelines require that the relationship between scores on the test and supervisor-rated job performance should be consistent across samples created by crossing the factors of the observation's gender, age, and race. Using ODA, this would simply entail determining whether the model relating the score on the test to the rated desirability of the employee was consistent across the samples created by crossing gender, age, and race.

Prospecting

Is the presence of conduits from source to reservoir rock positively related to the presence of oil? To answer this question one might conduct a study in which an oil firm drilled 10 wells in scattered sites that all had conduits (cracks), and drilled 10 wells in close proximity to these but in areas without cracks. The presence versus absence of oil is recorded for each well. With these data, ODA can determine whether the presence of cracks is a reliable indicator of the presence of oil. If the number and/or size of the cracks were recorded, ODA could be used to determine the number and/or size of cracks that best predicts the presence versus absence of oil. Were the amount of oil recorded, ODA could determine the number and/or size of cracks that best predicts the amount of oil located. Were the return (i.e., the price received for oil minus price of getting the oil to market) recorded, ODA could determine the number and/or size of cracks that best predicts the return from one's oil prospecting. Other variables, such as the presence (versus absence), size, and number of reservoirs (sandstone, broken limestone), source rock (shale), or shale maturity (thermal alteration) could also be studied as possible predictors. Multisample analysis could be used to determine if the

ODA model for the prediction of oil generalizes to the prediction of natural gas or coal. Conceptually similar procedures could be used in other forms of prospecting, whether for commodities (water, mushrooms, medicinal plants, pearls), gems (rubies, diamonds, sapphires), metals (gold, silver, platinum, copper, uranium), or treasure (sunken or buried).

Selling Shoes

Imagine that one is interested in determining whether interpersonal style influences success at selling shoes. To answer this question, one might conduct a study involving all customers (observations) attended by a single salesperson during one month. Observations are randomly assigned to one of two conditions: the interpersonal style of the salesperson is either abrupt and directive, or cordial and receptive (attribute). If a record was kept of which observations bought shoes and which observations did not (class variable), ODA could determine the interpersonal style that best predicted the number of shoes sold. Were records kept of the price of (or profit from) the shoes sold, ODA could determine the interpersonal style that best predicted the value (profitability) of the shoes sold. In addition to the interpersonal styles described above, other perhaps more appropriate or refined interpersonal styles could be added to the design, and additional attributes (e.g., amount of time spent with the observation; the gender, age, or mood of the observation; the number of different shoe styles from which to choose) could also be evaluated. ODA could also be used to evaluate the generalizability of these findings across different salespeople.

Speeding

Imagine that a motorist is interested in discovering factors that predict whether one receives a speeding ticket on the local interstate (class variable). To address this issue, one might conduct a study in which data are collected for all the motorist's journeys on the interstate over the past year. The attribute of primary importance is velocity (miles per hour). With these data, ODA can determine the velocity that provides optimal predictability concerning whether or not one will receive a ticket, or can determine the maximum velocity that minimizes the number of tickets received. Were information recorded concerning the dollar amount of the fine for each ticket, ODA could determine the velocity that optimally predicts the amount of fines that one receives, or that minimizes the amount of the fines that one receives. Other attributes that might be investigated include, for example, the time of day; weather conditions (wind, temperature, precipitation); the presence versus absence of salient social events (holiday, big game, town meeting, riots); or the terrain (straight or curvy; flat or hilly; open or forested). Were data collected for multiple samples, such as different types or models of automobiles, or different interstates, ODA could determine the extent to which the findings generalized across samples and could identify samples for which consistent findings emerged. It seems worthwhile to note that the highway police probably know about this by now, and may be using the same analyses to predict how best to nab speeders! However, unlike individual motorists who are probably most interested in minimizing their personal speeding fines, the police may weight the ODA analyses by the number of accidents, injuries, or fatalities incurred, and thus attend to different critical velocities, locations, terrains, and so forth in their decision making.

Suicide

Imagine that one is interested in predicting with maximum PAC the nature of depressed people who attempt to commit suicide. To address this issue one might conduct a study in which data are collected for a five-year comprehensive consecutive sample of clinical inpatients (observations) in a psychiatric hospital. By order of the court, all observations are subject to an involuntary 21-day commitment for psychiatric observation, and all are diagnosed as depressed. Observations who attempt to commit suicide during their stay at the hospital are to be discriminated against observations that do not attempt suicide prior to being discharged. The attribute of primary interest is the score that observations achieve on an objective component of an intake interview, that reflects the degree to which an observation manifests depressive and suicidal ideation. With these data, ODA can determine a model—that is, a cutpoint value on the score—that optimally classifies observations who do versus do not attempt suicide. Also, ODA can determine a model that optimally classifies the observations that specifically attempt (or specifically do not attempt) suicide. The observations who succeeded in committing suicide might well be separated from those who attempted but failed, forming a third category. The optimal ability of other attributes—such as prior behavior, religiosity, drugs, parental status, or stressful life events—to predict suicide status could also be determined. Were data available, ODA could also be used to evaluate the generalizability of the findings across different clinics and to identify clinics with consistent findings.

Target Recognition

Imagine that one is interested in developing a methodology of screening for training as potential helicopter gunners individuals who will be best able to discriminate friendly from hostile tanks during close-encounter, congested, daylight desert warfare. In pursuit of this objective, one might conduct a study in which data are collected for a random sample of gunners with an outstanding record during close-encounter, congested, daylight desert warfare in the war against Iraq (i.e., who destroyed at least one enemy tank and no friendly tanks), and from a random sample of gunners with a horrendous record (i.e., who destroyed at least one friendly tank and no enemy tanks). Images of enemy and friendly tanks are individually presented to each observation in a helicopter warfare simulator, and the total number of seconds taken by each observation to maneuver gunnery radar to orient with an aggressive disposition toward all of the enemy (attribute 1) and friendly (attribute 2) tank stimuli is recorded. With these data, ODA could find optimal cutpoint values (threshold number of seconds) on each attribute for best discriminating good gunners versus bad gunners. Were the number of enemy (or friendly) tanks destroyed by observations during the war against Iraq recorded, ODA could provide a cutpoint (on both attributes) that resulted in optimal prediction of the number of enemy (or friendly) tanks destroyed. The number of enemy (friendly) tanks destroyed during the simulated battle could be used similarly, although the validity of actual warfare data may be greater. The ability of additional attributes, such as visual acuity, visual-motor coordination, or experience with video games to predict skill could also be investigated. Were data collected for multiple samples, such as for dawn, twilight, and night fighting, or for other types of targets (vehicles, dwellings, encampments) or aircraft (other helicopters, airplanes, jets), ODA could determine whether findings were consistent across samples and, if that were not the case, could identify samples for which consistent findings emerged. Were it desirable to minimize civilian carnage, similar methods could be used to train gunners to recognize and avoid firing on civilian vehicles, dwellings, or encampments. Conceptually similar methods may be used to understand and perhaps enhance the discriminatory accuracy of players, coaches, or referees in sports such as baseball (strike versus ball, safe versus out) or tennis (fair versus foul shot).

Teaching

Imagine that one is interested in improving the quality of elementary school education. To address this issue, one might conduct a study in which data are collected for all first grade students who are randomly assigned to one of 16 different elementary schools (matched on socioethnographic factors including intelligence; ethnicity; socioeconomic level, gender, and transportation time) in a single county. Of these schools, 8 emphasize traditional teaching methods, and 8 emphasize self-directed, self-paced teaching methods (class variable). At the end of the school year, students are tested on a standardized measure of general achievement (attribute). With these data, ODA could be used to determine the optimal extent to which the two teaching methods result in discriminable scores on the achievement test. Other attributes that might be predicted by teaching method and that could be investigated include, for example student grade-point averages; relative frequency of students graduating to the next grade level or graduating from college (longitudinal research); ratings on Likert-type scales of how much students enjoy school; the number of hours per day that students spend studying and/or watching television, and/or the type of television programming (including both educational and non-educational programs) watched; the amount of time spent listening to music, watching videos, playing video games or sports; the education level and/or occupation of parents, siblings, and other relatives and/or friends; or reinforcement contingencies at home. The socioethnographic variables, alone or in combination, constitute possible alternative class variables. Alternatively, the socioethnographic variables may be treated as multiple samples (as could data from grades 2–8, or from multiple counties or states): ODA could then determine whether findings were consistent across the samples, and, if results were inconsistent, could find samples for which consistent findings emerged.

Vacationing

Imagine that one is interested in identifying attributes that optimally facilitate the personal a priori knowledge of whether a planned vacation will be enjoyable. In order to develop a personal ODA capable of providing this knowledge, one might conduct a study in which data are collected for one's last ten vacations (see the history example). On the first day after one's return home from each vacation, a rating indicating whether the vacation was enjoyable is recorded (class variable). Possible attributes include, for example, the presence and/or number of others and one's relationship with them; weather conditions; the location and expense of the vacation; frequency of opportunity and types of activities; and one's health status. With these data, ODA could determine the optimal ability of each attribute to predict with maximum PAC whether the vacation was rated as being enjoyable. If subjective weights reflecting a more precise rating of the enjoyment experienced on vacation were made (e.g., 10-point Likert-type ratings), ODA could predict with maximum PAC the overall enjoyability of the vacations. A variation on this involves a travel agent coaching clients in this methodology in an attempt to optimize the enjoyability of their vacations. Post-vacation interviews and records of return business would be useful in validational analysis.

Weight Loss

How does one identify variables that predict with maximum PAC whether a person attempting to lose weight is successful? To address this issue, one might conduct a study in which data are collected from a sample of people who enroll in a weight loss course with the objective of losing weight. On the first day of the course, observations are weighed and complete a measure of self-efficacy (belief that "I can"). At the end of the course, it is determined whether observations have lost weight relative to their weight at the beginning of the course (class variable). With these data, ODA could determine the value of the score on the self-efficacy scale that predicted with maximum PAC whether observations lost weight. Were information recorded concerning the absolute amount of weight lost or gained (this enforces that individuals with the greatest absolute weight changes are weighted most strongly by the model), ODA could determine the value of the score on self-efficacy that predicted with maximum PAC the amount of weight lost or gained. The ability of additional attributes, such as gender, age, marital status, amount of daily free time, desk-bound or mobile occupation, home location (transportation time from parks, beaches, bike paths, or other recreational resources), or car ownership status to predict frequency or amount of weight loss success could also be investigated. Were data collected from other samples (such as other weight-loss groups; groups of individuals attempting to terminate smoking, drinking, or other substance abuse behavior; or groups undergoing different clinical intervention methods), ODA could be used to assess whether findings were consistent across samples and to identify samples with consistent findings.

Zoology

What factors are related to the birth (class variable) of lion cubs in captivity? To address this question, one might conduct a study in which data from all U.S. zoos with at least one pair of sexually mature male and female lions are collected. For every non-sterile sexually mature female lion, it is determined whether at least one surviving cub was born (class variable) during 1992. The attribute is the length of time that the female had been in captivity (attribute). With these data, ODA could determine the length of captivity that optimally predicted whether the lioness would bear a cub. Were the number (or overall weight) of cubs recorded, ODA could determine the length of captivity that predicted with maximum PAC the number (weight) of cubs born. Other attributes that might predict the fertility of lions that could be investigated, for example, might include the number of human visitors; distance from visitors to lions; mean noise level; mean temperature, humidity, or barometric pressure; the age of the female or the male; or the length of time the female and male lions had cohabited. Using ODA, the generalizability of the findings could be assessed over multiple samples, such as data collected for other felines (tigers, pumas, bobcats), mammals (wolves, polar bears, elephants), or animals (reptiles, birds, fish). If findings were inconsistent across all samples, ODA could be used to identify the samples for which consistent findings emerged.

Now It Is Time to Start Analyzing Data

There is a new tool on the block. Applied research in a myriad of substantive areas has been conducted over the past decade using earlier versions of the software accompanying this book, and has met with stellar reception vis-à-vis publication in a host of scientific journals. As we hope the hypothetical examples illustrated, the ODA paradigm applies optimally to any quantitative data set and can be specifically tailored to evaluate any stated hypothesis—truly a "designer" statistical methodology. At this point in the discussion, we believe that many readers must be eager to witness the new tool in action. That is what we turn to next.