Mining Platinum and Gold: *Using Statistics for Credit Decisions*

Tim K. Keyes, Ph.D.
VP. GE Capital Commercial Finance

Abstract: We discuss some statistical mysteries behind credit decisions.

1. Introduction

If you have a credit card, you may be aware that information about you is used at the point of sale by someone who wants to loan you money. If you are in good credit standing (whatever that means) and can afford the new item, you get "approved;" otherwise, you get "denied" (too bad, so sad). Sometimes, and often for no apparent reason, the clerk may ask you to surrender your card and then stare at you as you walk away nonplussed. Luckily, more often you'll receive a bonus for having successfully managed the card, e.g., a discount on your desired purchase, or a coupon for other purchases.

What, exactly, is going on here, and who's in charge, anyway? What personal information is being used? How?

You've probably guessed that a statistician or data miner is working somewhere behind the scenes, making mathematical guesses about your future payment behavior. You're right! The talents embodied by these people are a necessary ingredient to the effective management the *millions* of credit applications, and *billions* of purchase requests annually in the United States and abroad. From a statistician's point of view, this application is extremely exciting and gratifying: *billions* of dollars worth of credit decisions are made each year from the integration of statistical models into practical business solutions. The Federal Reserve Board estimated that credit cards in 1998 carried $550 billion of debt in the U.S. alone. In few other ways can a statistician or data miner have so much impact! See Berry and Linoff (1997) for more on other interesting applications.

2. Consumer Credit

Let it be stated here that you (the credit holder) are in charge (forgive the pun), so long as you manage your spending patterns. It helps to know what information about you is collected, managed, and used. A good website for general information is http://www.thecreditreportsite.com/index.html .

People who want to loan you money want to know as much about your financial habits as possible. Fortunately, it's impossible to quantify this exactly (wouldn't statistics be boring otherwise?). Nevertheless, credit grantors collect a wealth of information about you, and people like you, from various sources. This information is used only to manage their risk of extending credit.

When you apply for credit, for example at a bank for a car, at a retailer for a credit card, etc., some basic personal information is requested. Your name, address, and income, for example, are required entries on a typical credit application at your local CompuGlobalHyperMegaNet store that sells software. The clerk may issue you a temporary "credit line," i.e., the ability to charge up to some nominal amount, say $500. How'd they arrive at that number? Why not $425 or $750?

Based on the store type and location, and your application information, credit grantors will "score" you with a prediction of payment behavior, i.e., they associate these characteristics with a likelihood that you'll fail to make a payment. How? Many thousands of individuals like you have applied for credit in the past at similar stores, and have demonstrated the ability, or inability, to pay off their accumulated purchases. This historical database represents a rich "model building" data set for credit grantors to use in the construction of statistical models (for the interested reader, they're chiefly logistic regression models), which predict your likelihood of failure to pay. For good or bad, you get a predicted response based on the behavior of all the people who are just like you - in terms of the few variables collected on the credit application (and to whom the retailer has historically extended credit). You can thank Sir Francis Galton for all this - he invented the idea of regression. See Lewis (1992) for a good treatise on 'scoring' methods.

The prediction will say something like, "applicant has a 2% chance of defaulting on one of first 6 months' payments." Below some threshold risk, say 5%, you'll be extended an initial credit line that is balanced against the retailer's (and card issuer's) appetite for losing money. Above the threshold…too bad, so sad…you're 'denied'. Try again later.

After you establish your credit history, other decisions such as credit line extensions, interest rate decreases, bonuses, and the like, are based on similar statistical models, built from your continuing financial and purchasing characteristics and those of all creditees who came before you.

Why are statistical models required? Primarily because they can be programmed in a computer, which aids in the efficient disposition of large volumes of credit applications. Secondarily, statistical models are not given to the inherent inconsistencies we humans are when we process large volumes of information, day-in, day-out. So long as the data are gathered with as much accuracy as practical, and the statistician uses good practices when building the models, it is reasonable to expect that appropriate credit decisions will be made in a routine way. It is always necessary, of course, to review the models for degradation. Statistics definitely has its benefits, but it has its limits, too!

3. Commercial Credit

When commercial businesses apply for credit, for example when CompuGlobalHyperMegaNet, Inc. applies for a loan to acquire other stores, they go through essentially the same process that consumers do - as outlined above. However, the loan amount sought is typically much larger - on the order of millions (USD) rather than hundreds or thousands, as in consumer credit. Another key difference is that businesses seeking large loans may be required to offer collateral security (i.e., something of value in case of default). Loans may be secured by property, such as buildings or land, or by inventory, such as CompuGlobalHyperMegaNet's PC inventory. Statistical models can also be built to estimate the value of collateral offered, which further aids the credit granting decision process when large numbers of businesses are under consideration.

4. Summary

The credit industry is an enormously exciting and gratifying place for a statistician / data miner to work. It's exciting principally owing to the fact that billions (US dollars) are at stake, and it's gratifying chiefly because statisticians and their models are playing a huge role in the disposition of many, if not most, credit decisions. In short, we get to "see" the results of our endeavors measured against key corporate initiatives and revenue targets. There is always the possibility of downside risk…but, a statistician's role (and responsibility) calls for the appropriate use / mix of traditional decision making, along with statistical models, to make sure that the right credit decisions are made at the right time with the right effect. All this gets more even exciting as data 'minable' information becomes available via the Internet. It's a great time to be a statistician!

References

1. Berry, M. J. A., and Linoff, G., <u>Data Mining For Marketing, Sales, and Customer Support</u>, John Wiley & Sons, Inc., New York, 1997.

2. Lewis, E. M., <u>An Introduction to Credit Scoring</u>, Fair, Isaac, and Co., Inc, 1992.