

# CREDIT SCORING AND CREDIT CONTROL X EDINBURGH 2007

## Analytical Data Profiling

### Abstract

Fundamental to the production of accurate and robust scorecards is the availability of appropriate development data. The sample must be large enough to ensure accurate statistical inference but it must also be appropriate to the situation. Where the available development data is too dissimilar to the current population, or is too small to exhibit robust trends, statistical techniques are unlikely to produce satisfactory models.

In this paper, the term *Data Profiling* is used to describe the adjustment of one data sample to resemble more closely a different, primary data sample. We present a methodology that can be used to adjust or augment the existing development data, thereby addressing its deficiencies.

Using case studies and applications from recent projects, the paper demonstrates how Data Profiling can strengthen existing analytics, or open up entirely new development possibilities. For example:

A small sample can be supplemented with additional data from a secondary source adjusted to match the profile of the original sample, adding robustness to any statistical inferences  
A scorecard development sample can be tuned to resemble more closely the current application population, ensuring that the resultant scoring models are as relevant as possible to the current business need

Data from an established portfolio can be profiled to match applicants for a new product, allowing 'custom' models to be built even in the absence of performance data

In summary, the Data Profiling methodology enables solutions to be developed that are more sophisticated and more appropriate to a business in situations where there is limited data available.

**Daniel Weaver**  
**Business Consultant**  
Experian Decision Analytics  
Landmark House  
Experian Way  
Nottingham  
NG80 1ZZ

daniel.weaver@uk.experian.com  
Tel : +44 (0) 115 82 86568  
Fax : +44 (0) 115 82 86603