PREDICTIVE MODELING

Central Ohio Insurance Education Day – April 15, 2015

Tim Crespin, Ph.D.
AVP-Director, Predictive Modeling
State Auto Insurance Companies
Organizing Questions

• **What** is Predictive Modeling (PM)? – *Part 1*
• **What** is Predictive Modeling (PM)? – *Part 2*
• **Why** is PM of interest?
• **When** did PM come on scene?
• **Where** is PM?
• **Who** works with PM?
• **How** is PM done?
• **What** is Predictive Modeling (PM)? – *Part 3*
• **Q & A**
WHAT IS PM?
Part 1: Logic
Formal symbolic logic

GLM with Tweedie distribution in loss cost models

\[ f(n, y; \varphi/w, \rho) = a(n, y; \varphi/w, p) \exp\left\{ \frac{\omega}{\varphi} t(y, \mu, p) \right\} \]

Where

\[ a(n, y; \varphi/w, p) = \left\{ \frac{(w/\varphi)^{\alpha+1}}{(p-1)^\alpha} \right\}^n \frac{1}{n! \Gamma(n\alpha)y} \]

\[ \alpha = \frac{2-p}{p-1} \]

\[ t(y, \mu, p) = y \frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p} \]

\[ \Gamma(n\alpha) = \int_0^\infty t^{n\alpha-1} e^{-t} dt \]
“Beware of geeks bearing formulas”
- Warren Buffett

“Because the math is really complicated people assume it must be right”
- Nigel Goldenfeld

"All models are wrong but some are useful"
- George E.P. Box
Applied Multivariate Statistical Analysis

- Mathematical constructs put into symbolic form
- Allow us to understand and develop the logic of PM
- Advanced, specialized education is required
- Most predictive modelers have advanced degrees (Ph.D.’s or Master’s)
  - Statistics
  - Biostatistics
  - Decision sciences
  - Econometrics
  - Social/behavioral sciences
  - Emerging field: Data Science
WHAT IS PM?

Part 2: Real life represented in equations
Intelligent Simplification of Reality

Premiums Are Higher in States with Higher Awards Per Doctor

- Med-Mal Insurance Premiums ($)
- Awards Per Doctor

\[\text{Premiums} = a + b \times \text{Awards Per Doctor}\]

- \(a\): minimum asymptote
- \(b\): steepness of the curve
- \(c\): inflection point
- \(d\): maximum asymptote

X-Y-Z Scatter Plot
- Weighted surface
- Weight plot

- SC05
- Ordinary Least Squares
- Least median squares
- Major Axis Analysis
- Bayesian
Fundamental Characteristics of PM’s

- Intelligent simplification of reality
- Built on specialized theory, technique, assumptions
- Dependent on data input, calculations, output
- Produces a prediction relevant to business context
- May greatly increase efficiency and effectiveness of decision
- Challenging to understand and communicate
Methodical, Measurable

- Conceptual and design soundness
- Rigorous technical testing and validation
- Interpretability of model content and output
- Actionable insight to aid business decision making
- Quantifiable improvements over existing operational processes and business outcomes
WHY IS PM OF INTEREST?
Necessary and impactful

- Accuracy and efficiency in decision making
- Necessary to compete in today’s world
- Changes the nature of work
- Impact on finances, customers, reputation
Endless possibilities for applications

Types of Target Variables
- Dichotomous
- Categorical
- Ordinal
- Interval/ratio
- Change, trend

Types of Questions
- Level, amount
- Odds, probability
- Frequency
- Variability
- Duration, latency
- Group membership

Types of Applications
- Predict
- Describe, Explain
- Estimate
- Prioritize
- Optimize, economize
- Alert
Developing quantitative vision and insight

Based on: Competing on Analytics, Davenport and Harris, 2007

Degree of Complexity

<table>
<thead>
<tr>
<th>Competitive Advantage</th>
<th>Advanced Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic Optimization</td>
<td>Prescriptive and Predictive</td>
</tr>
<tr>
<td>Optimization</td>
<td>Support new business models and opportunities</td>
</tr>
<tr>
<td>Predictive modeling</td>
<td></td>
</tr>
<tr>
<td>Forecasting</td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td></td>
</tr>
<tr>
<td>Alerts</td>
<td>Descriptive Analytics</td>
</tr>
<tr>
<td>Query/drill down</td>
<td>Support ongoing business operations</td>
</tr>
<tr>
<td>Ad hoc reporting</td>
<td>Meet compliance requirements</td>
</tr>
<tr>
<td>Standard Reporting</td>
<td></td>
</tr>
</tbody>
</table>

What exactly is the problem?
What will happen next if …?
What if these trends continue?
What could happen…?
What actions are needed?
What exactly is the problem?
How many, how often, where?
What happened?

Based on: Competing on Analytics, Davenport and Harris, 2007
WHERE IS PM?

We touch predictive models and they touch us.
Woven Into the Fabric of Commerce

• Retailers: Product recommendation
• Mobile telecom: Cross sell, deep sell, churn, “savability”.
• Search engines: Target advertising and marketing
• Banks: Credit risk management
• Healthcare: Readmissions case-management and patient care improvement
• Weather service: Weather forecasting
Insurance: Many ways, many uses

Gain insight by seeing graphical patterns in data and modeling results.

Predictive Analytics
- Forecasting, data-mining, econometrics, statistical modeling

Simulation
- Replicate a system, process, behavior, or business problem.

Optimization
- Mathematically find best solution among many options and constraints.

Visualization

Improve business processes, decisions.

Predictive modeling creates/improves business rules. Also, business rules govern the output and use of predictive models in production.

Predictive modeling is critically dependent upon good data sources. Predictive modelers discover errors and new opportunities for data.

Uses:
- Underwriting
- Pricing
- Sales & Distribution
- Customer
- Marketing
- Contact Center
- Claims
- Billing
- Profitability
- Risk & Fraud
Every Phase of the Insurance Policy Lifecycle

- Marketing and customer insight
- Product development and innovation
- Risk selection and pricing
- Expense management and profitability
- Business process optimization
- Agency relations / distribution management
Example 1: Knowing our customers

Acquisition and relationship management

- Target Marketing
  - Target campaigns for products/features to relevant segments of population.

- Response Models
  - Lower costs of conversion by targeting customers most likely to respond.

- Retention
  - Estimate renewal probability, longevity of ongoing customer relationship.

- Cross-Sell
  - Identify products most likely purchased/bundled by similar customers.

Example Applications
Example 2: Claims Fraud Analytics

- Address fraud and abuse throughout the claims lifecycle.

**Prevention**
- Identify fraud risks at policy submission

**Alerts**
- Flag & route fraudulent claims at intake

**Identification**
- Help Adjuster ID fraud during adjudication

**Discovery**
- Perpetually analyze loss data to ID fraud

**Investigation**
- Investigate, prosecute & recover fraud

**Monitoring**
- Report claim fraud outcomes and statistics

**Preventive Underwriting**
- Number of Claims within 30 days of Eff Date from this Agent

**Anomaly Detection**
- Claimant, police officer, and body shop seen 5 times previously

**Identity Resolution**
- Twitter post identified and reported to Adjuster

**Patterns/Clusters**
- New pattern of data discovered - applied to entire claims book

**Unstructured data**
- Productivity increase as more pattern detection is automated

**Geo-Spatial Trend Analysis**
- Geo-spatial analysis of ratio of BI claims in specific area
WHEN DID PM APPEAR?

Introduction and proliferation within insurance
Historical View of Predictive Modeling

- WWII
- 1960’s
- 1970’s
Insurance risk scores based on credit data

- Introduced early 1990’s
- Not the same as credit worthiness score in banking
- Related to personal financial management and condition
- Future insurance losses correlate with credit attributes.
Current issues and influences

- Big data: Volume, variety, velocity
- Integrated (aka “optimized”?) pricing:

```
Target Pricing Zone

Target Market

Reduce prices to gain profitable market share

Raise prices to take more profit

Elasticity

Margin

high

low

low

high
```

Reduce prices to gain profitable market share

Target Pricing Zone

Raise prices to take more profit

Elasticity

Margin

high

low

low

high
WHO WORKS WITH PM?

Hint: Just about everyone
Modeling lifecycle: interdisciplinary!

**Conceptualization**
- Business Strategy
- Modeling Objectives
- Dependencies

**Specification**
- Analytic Strategy
- Data Sourcing
- Model Building
- Validation

**Implementation**
- Business Process
- Design, control
- Production infrastructure

**Application**
- Adoption, Dissemination
- Interpretation, Reliance
- Monitoring, Control

**Versioning**
- Refresh Frequency
- System Updates
- Next Generation
HOW DO WE DO PM?

How predictive modeling is accomplished
Multivariate model building process

- Pare down large collection of predictive variables to a manageable set
- Examine correlations among the variables
- Weed out redundant, weak, poorly distributed variables
- Build candidate models using various techniques
  - Regression/GLM
  - Decision Trees/MARS
  - Neural Networks
- Analytically derive optimum weighting of predictors
- Validate/compare models
- Select final model
Business understanding is primary

- **Business Understanding**
  - Determine business objectives
  - Determine analytic strategy
  - Develop project plan

- **Data Understanding**
  - Collect initial data
  - Describe and explore data
  - Verify data quality

- **Data Preparation**
  - Select data
  - Clean and derive data
  - Integrate and format data

- **Analysis**
  - Select modeling technique(s)
  - Build model(s)
  - Assess model(s)

- **Evaluation**
  - Assess results against objectives
  - Review process
  - Determine next steps

- **Deployment**
  - Plan deployment
  - Produce final report
  - Review Project

Source: “CRISP-DM”
Model performance is measurable

- **Scoring engines**
  - Assign a “predictive score” to each unit…

- **Lift curves**
  - How much [riskier] than average are the policies with the [highest] scores?

- **Out-of-sample tests**
  - How well will the model work in the real world?
  - Un-bias the estimate of predictive power
Example of Model Evaluation: Lift Curve
Analytics of implementation

- **Perform model description analytics**
  - Necessary for comfort using the model

- **Calibrate Models**
  - Create user-friendly “scale” – client dictates

- **Implement models**
  - Programming, process design, system integration

- **Monitor performance**
  - Distribution of scores over time, predictiveness, usage of model
  - Plan model maintenance, versioning
Business considerations for rating models

So our predictive models show our pricing can be improved by adding a new rating component… What else do we consider before implementing?

- Impact to agents
- Competitor Pricing
- Regulatory approval
- IT & Data Costs
- Will we be better positioned for profitable growth?
WHAT IS PM?

Part 3: Metaphors help/hurt our understanding of new things
Most Understanding is Implicit
Silly Metaphor: “Mathe-magic”

Psychic Precognition May Exist, Cornell Study Finds
Helpful Metaphor: Rear View Mirror
Silly Metaphor: Rocket Science
Helpful Metaphor: Betting system

<table>
<thead>
<tr>
<th>Split</th>
<th>Stand</th>
<th>Hit</th>
<th>Double</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>A</td>
<td>Free e-Flash Cards! MODERN CARD COUNTING</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Silly Metaphor: Alien Crew Member
Helpful Metaphor: Factory Automation
Helpful Metaphor: Power tool
Silly Metaphor: “Numerati”
Helpful Metaphor: Uphill Earns Downhill

Profiling
- Context, scope
- Selection
- Definitions
- Acquisition
- Interpretation
- Validation

Preparation
- Primary Key
- Analysis
- Aggregation
- De-duplication
- Derived Data
- Mapping

Integration
- Merge/
- Concatenate
- Data Integration
- Testing
- Quality Testing
- Reconciliation

Reconditioning
- Normalizing
- Imputations
- Translations
- Cleaning
- Mapping

Modeling-Ready Datasets
Q & A
PREDICTIVE MODELING

Central Ohio Insurance Education Day – April 15, 2015

Tim Crespin, Ph.D.
AVP-Director, Predictive Modeling
State Auto Insurance Companies