GIRO Conference and Exhibition 2012
Juggling uncertainty the actuary’s part to play
Analytics: A transatlantic comparison

James Rakow, James Guszczza & Matthew Wilson
Deloitte
Analytics: A Model Definition
Data

- Historically, use of data driven by legislation and customer attitudes
  - Regulatory requirements
  - Route to market
  - What is acceptable to the customer

- More recently, an explosion in global data availability has led to insurers looking elsewhere
  - 3rd party data
  - Non-traditional data
  - Open data
Models: GLM and Beyond

- Two levels of Discussion:
  - Tools and Methodology
  - Application

- At the level of methodology:
  - Regression and its Relations
  - Statistics vs machine learning
  - Traditional vs “modern” approaches to data analysis
  - Hallmarks of “Data Science”
  - “Actuarial Science as Data Science”
Data Science: An Implicit Definition

Image borrowed from Drew Conway’s blog
Putting GLM in Statistical Perspective
(Or: Why Programming with Data is Important)

• Linear Model Assumptions:
  – Courtesy of Andrew Gelman and Jennifer Hill.

1. Valid model design and variable selection
2. Additivity and Linearity
3. Independence of errors
4. Equal variance of errors (homoscedasticity)
5. Normality of errors

\[ Y = \beta \cdot X + \varepsilon \]
\[ \varepsilon \sim_{iid} N(0, \sigma^2) \]

(Paraphrase from Data Analysis Using Regression and Multilevel/Hierarchical Models by Andrew Gelman and Jennifer Hill)
GLM and Beyond: Tools and Techniques

- **Supervised Learning:**
  - Regression → GLM
  - Multilevel / Hierarchical and Bayesian Models
  - Generalized Additive Models
  - CART, MARS, Random Forests
  - Neural networks, Support Vector Machines

- **Unsupervised Learning**
  - Clustering
  - Factor / Principal components Analysis
  - Association Rules
  - …
Beyond GLM: Methodology and Mindset

• Predictive variables are the “fuel” that “power” scoring engines.

• This means that the most important part of the modeling process occurs before the modeling begins
  – EDA: Exploratory Data Analysis
  – KDD: Knowledge Discovery in Databases

• Focus on:
  – Model design
  – Programming with data
  – Variable creation and selection
    – Remember Gelman’s #1 assumption
    – Early example: credit scoring has been a major innovation for US actuarial work
  – Out-of-sample model validation
Lift Curves and Out-of-Sample Validation

- The availability of a multitude of variables and powerful machine learning tools heightens the risk of overfitting.

- Out-of-sample validation became a “gold standard” in US actuarial practice around the same time the “data science” approach took hold.

- Key tool: the lift curve.
Training

What do actuaries (and others) learn about Analytics?

Image borrowed from Drew Conway’s blog
When in your education did you meet these distributions?

- Binomial Distribution
- Beta Distribution
- Normal Distribution
- Multinomial Distribution
- Dirichlet Distribution
- Multivariate Normal
Analytics – the Full Picture?

Data → Model

Training
Analytics – the Full Picture

**Strategy** → **Data** → **Model** → **Decisions**

**Training**
It All Starts with a Strategy...

Moneyball – A great book/film about
- Business analytics
- Innovation through data science
- Behavioral economics
- Human psychology
- Culture change

- (... and, oh yes, baseball)
... And ends with decisions

Decision support – Where Analytics adds/destroys value

• Divergent outcomes
  – Human decision-makers remaining central, and models supplementing and anchoring human judgment; not replacing it.
  – Use of analytical techniques resulting in people being removed from the decision-making process
Closing Remarks and discussion

Strategy → Data → Model → Decisions

Training

© 2011 The Actuarial Profession • www.actuaries.org.uk
Questions or comments?

Expressions of individual views by members of The Actuarial Profession and its staff are encouraged.

The views expressed in this presentation are those of the presenters.