

# Insurance Analytics and Pricing using R

## Detailed Course Outline

### Day 1: The basics

#### *Introduction to the course.*

- What it is and isn't. How we will work: format, exercises, and homework.
- Tools: setting up R with the required and recommended packages.

#### *Introduction to pricing*

- The business problem. Terminology.
- The data sets we will be using.
- Exercise: load and visualize data.

#### *Generalized Linear Models – the basics*

- Assumptions: (i) Policy independence, (ii) Time independence, and (iii) Homogeneity.
- Multiplicative models.
- Modelling claim frequency
- Modelling claim severity
- Modelling premium
- Exercise: Motorcycle insurance

#### *Homework*

- Motorcycle insurance

### Day 2: Model building

#### *Recap and answers*

- Recap of motorcycle insurance exercise
- Answers to selected homework exercises

#### *GLM model building*

- Confidence intervals
- Variable selection
- Residuals
- Model tests
- Cross-validation and bootstrap
- Dealing with: Overdispersion, interactions, offsets, outliers
- Dealing with deductibles / excess

#### *Exercise and homework*

- Model selection in motorcycle insurance

### Day 3: GLM extensions

#### *GLM multilevel factors and credibility theory*

- With example from bus insurance

***GAM: Continuous rating factors***

- Introduction to Generalized Additive Models
- Overview of the theory
- One rating variable
- Many rating variables
- Example

***GLMM: Longitudinal claims***

- Random and mixed effects
- Introduction to GLMM
- Uses of GLMM in insurance
- Example

***Exercise and homework***

- Extending the GLM from the previous days

**Day 4: Ensemble models*****The basics***

- Bagging and boosting
- Penalizing
- Examples

***Specific implementations and tuning***

- Random Forest
- Package gbm
- Model tuning with the caret package

***Examples and implementation***

- Example on motorcycle data set
- Using caret for ensemble modelling

- Some recommendations and learnings from practical implementations

***Exercise and homework***

- Build models (at least Random Forest and gbm) on data set provided

**Day 5: Understanding, business change, and results*****Answers to exercise and homework***

- Comparing GLM, RF, gbm, and RF+gbm for maximum predictive power

***Understanding ensemble models***

- Variable importance
- Visualizing effects of variables
- Model reduction
- Creating derived variables for GLM
- Examples

***Business change***

- Implementing change in the business
- Model validation: restricting the validity domain or using derived variables
- Regulatory compliance

***Conclusions***

- Summary
- Questions
- Additional topics requested