SOCIETY OF ACTUARIES®

SOA Big Data Seminar

13 Nov. 2018 | Jakarta, Indonesia

# Session 2

# Motor Insurance Pricing

George Kau, FSA
Victor Khong

# SOA Big Data Seminar
# Motor Insurance Pricing

George Kau FSA, FASM
KPMG PLT

Victor Khong
Nicholas Actuarial Solutions

13 November 2018

SOCIETY OF ACTUARIES

## Brief Introduction of Motor Insurance Rating in Malaysia

SOCIETY OF ACTUARIES

# Motor Insurance
## – Basic Cover

Motor Insurance in Malaysia is renewed yearly
Premiums are paid before insurance coverage starts

Death or injury to other parties (TPBI)

Damage to other parties' property (TPPD)

Third Party Cover

Own loss due to theft or fire

Third Party Fire and Theft Cover

Own damage to vehicle due to accident (OD)

Comprehensive Cover



# Motor Insurance
## – Extension Cover

Additional perils can be added to the policy with additional premiums

- Flood, earthquake, hurricane, landslide
- Strike, riot and civil commotion
- Passenger liability
- Additional business use
- Breakage of glass in windscreen or windows
- Additional named driver
- Liability of passengers for acts of negligence
- Tuition and testing purposes

# Motor Tariff
## - Rating Factors

Premium rates charged by insurance companies were ranging within the allowable loading limit of Motor Tariff.

Rating factors set out in the motor tariff

- Sum insured
- Engine capacity of vehicle
- Region such as West and East Malaysia
- Loadings for age of driver, age of vehicle and past claims history



# Liberalization of Motor Tariff
## - Additional Rating Factors

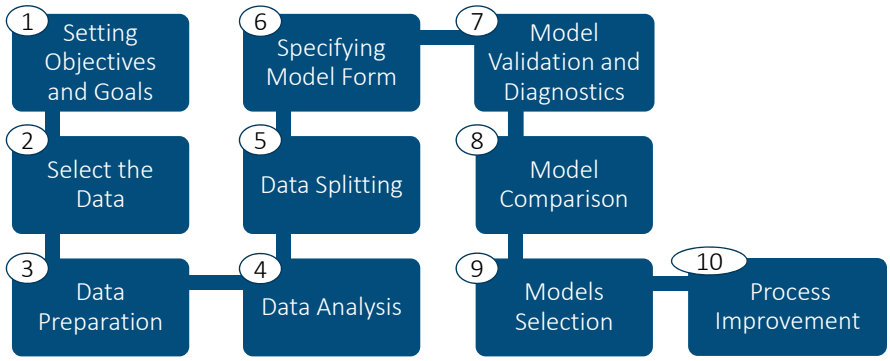
General insurance companies began to use **Generalized Linear Model (GLM)** in self motor insurance rating

Premiums determined after liberalization of motor tariff

Additional rating factors

- Safety features
- Vehicle make
- Gender of driver
- Experience of driver

## Process of Building a Generalized Linear Model



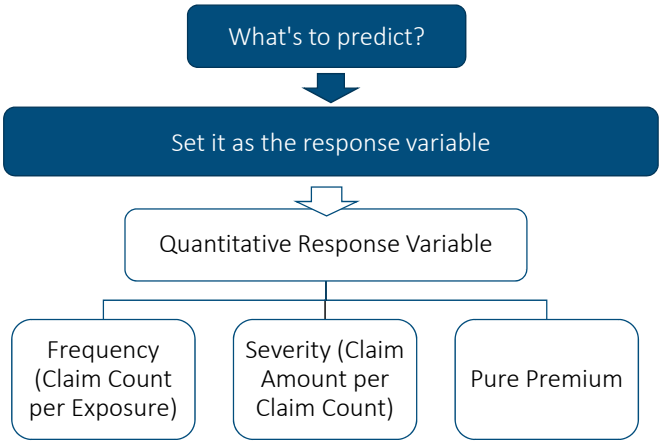



# GLM – Data Preparation

Step 1 - 5

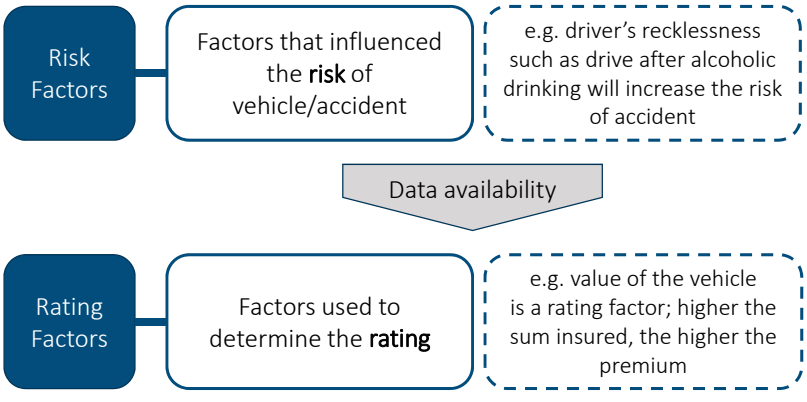
SOCIETY OF ACTUARIES

# Step 1. Setting Objectives and Goals
## – Purpose of Modelling






# Step 2. Select the Data
## – Risk Factor Vs. Rating Factor



## Step 2. Select the Data (cont'd)
– Driver Factor Category

| Rating Factor | Description | Data Structure |
|---|---|---|
| Age of Driver | Age of **vehicle owner**, or age of policyholder | Integer |
| Driving Experience | **Length** of **driving period** or Experience | Integer |
| Driving Record | **Number** of **traffic offences** or **bad record** | Integer |
| Gender | Male or Female | Categorical |
| Marital Status | Single or Married | Categorical |
| Number of Driver | **List of drivers** in the policy | Integer |



## Step 2. Select the Data (cont'd)
– Vehicle Factor Category

| Rating Factor | Description | Data Structure |
|---|---|---|
| Cubic Capacity | **Dimension of vehicle engine** | Integer |
| Manufactured Year | Number of **years** since the vehicle is manufactured | Integer |
| Safety Features | Number of **safety installations** | Integer |
| Odometer | **Distance travelled** by the vehicle | Numerical |
| Vehicle Type | Sports or Normal vehicle | Categorical |

## Step 2. Select the Data (cont'd)
– Location Factor Category

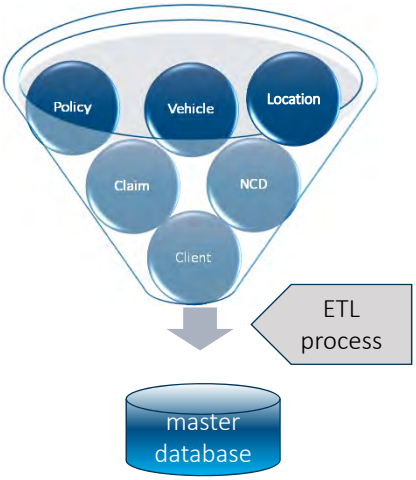
| Rating Factor | Description | Data Structure |
|---|---|---|
| Region | **East** or **West Malaysia** | Categorical |
| Address Location | **Postcode** | Categorical |
| Urbanization Level | **City, rural and suburban** | Categorical |



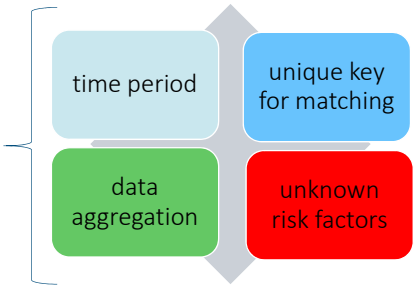
## Step 2. Select the Data (cont'd)
– Policy Factor Category

| Rating Factor | Description | Data Structure |
|---|---|---|
| Sum Insured | Market value or agreed value of the vehicle | Numerical |
| Policy Coverage | **Type** of coverages | Categorical |
| Renewal Indicator | **New** business or **renewal** Business | Categorical |
| Claim Count Experience | **Number** of claim incurred in the past | Integer |
| Claim Amount Experience | **Amount** of claim incurred in the past | Numerical |
| No Claim Discount (NCD) | Discount offered for good driving record | Numerical |

# Step 3. Data Preparation
## – Merging and Consideration

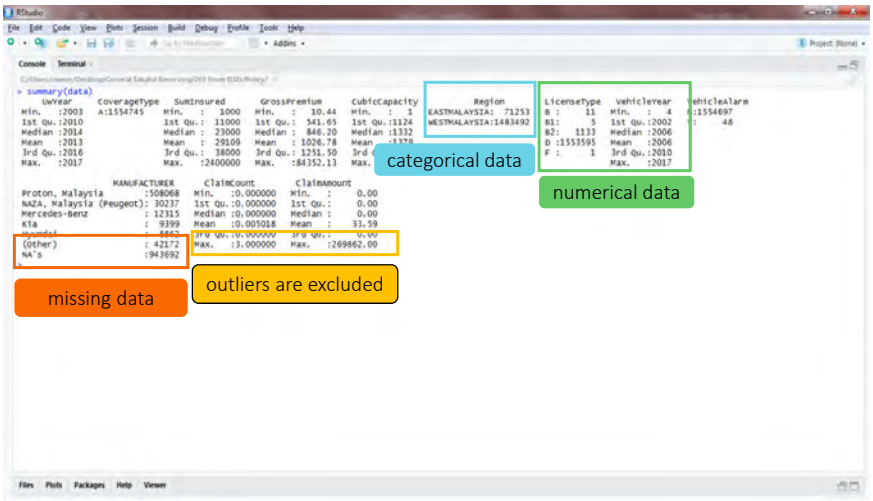



Consideration before merging

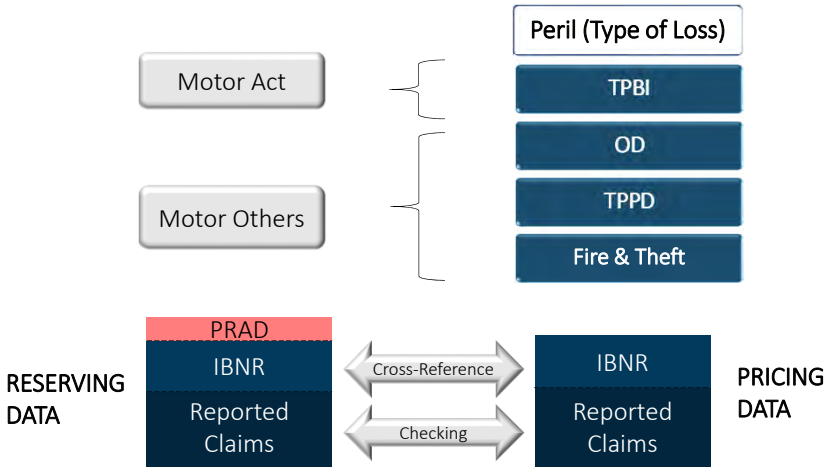





# Step 3. Data Preparation (cont'd)
## – Merging and Consideration



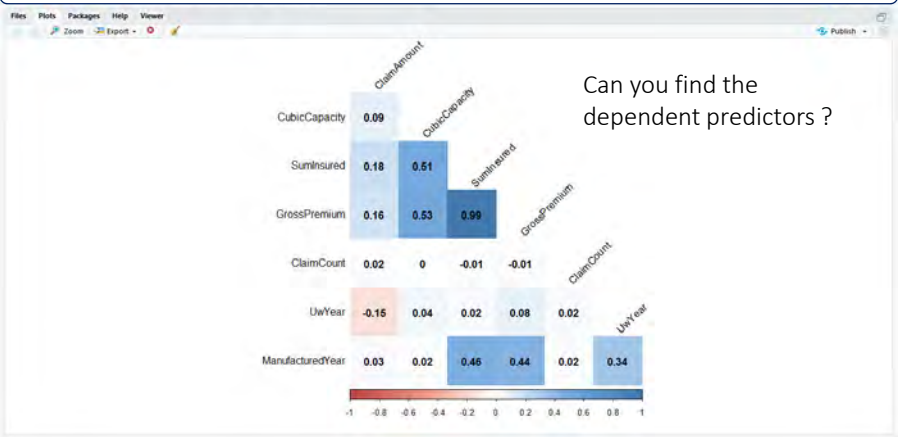
# Step 4. Data Analysis
## – Reserving vs Rating






# Step 4. Data Analysis (cont'd)
## – Correlation Plot

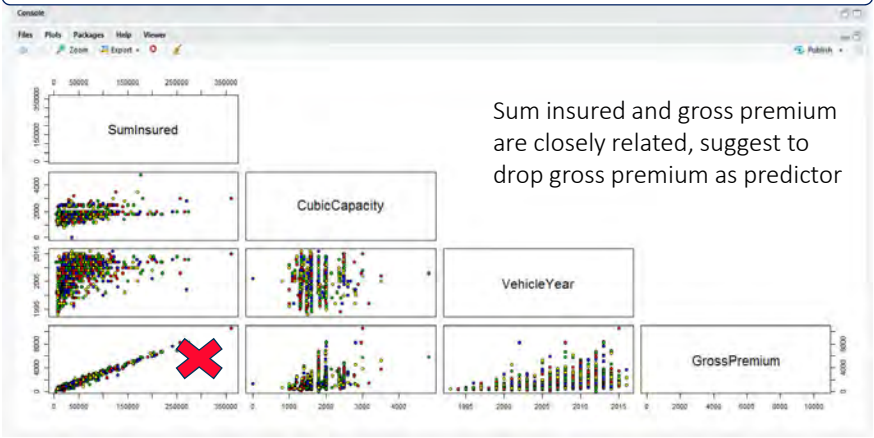
Correlation Plot – Pearson Coefficient Correlation Method




Can you find the dependent predictors ?

# Step 4. Data Analysis (cont'd)
## – Relationship Pattern Plot

Relationship Pattern Plot




Sum insured and gross premium are closely related, suggest to drop gross premium as predictor



# Step 5. Data Splitting
## – Training and Validation Sets

| | |
|---|---|
| Training Set | (*70%*) to **BUILD** the GLM model using rating factors |
| Validation Set | (*30%*) to **REFINE** the GLM model |

# GLM - Modelling
## Step 6 - 9

SOCIETY OF ACTUARIES



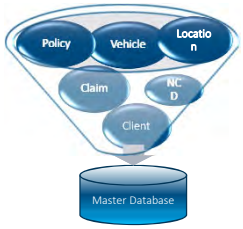
# Generalized Linear Model
## - Response variable

Regression analysis is a form predictive modeling technique which investigates the **relationship** between a **response variable** $Y$ and the **predictors** $X_i$

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

Response variable

Specifies the explanatory variables $(X_1, X_2, \ldots X_N)$ in the model



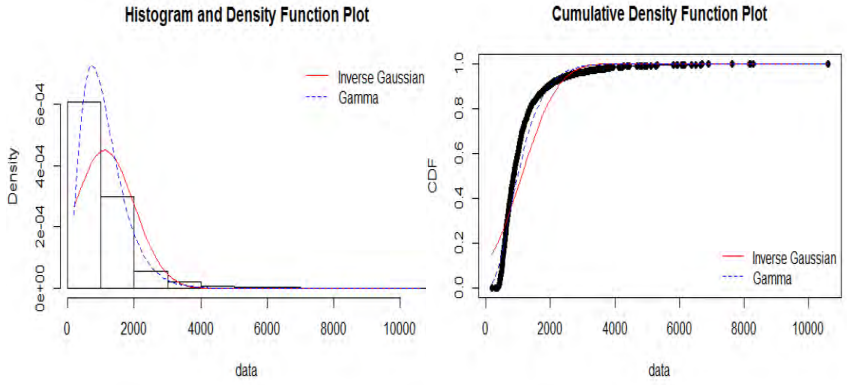
# Generalized Linear Model (cont'd)
- Response variable

| | |
|---|---|
| **Continuous** Response Variables e.g. severity, net premium | Inverse Gaussian / Gamma Regression |
| **Count** Response Variables e.g. claim count | Poisson / Negative Binomial Regression |
| **Categorical** Response Variables e.g. fraud, lapse (yes or no) | Binomial/Logistic Regression |



# Generalized Linear Model (cont'd)
- Response variable

**Gamma distribution** v.s. **Inverse Gaussian distribution** for **Severity** Model



# Generalized Linear Model (cont'd)
## - Response variable

| Distribution | Typical Uses | Support of Distribution |
|---|---|---|
| Gaussian (Normal) | Linear response data, constant increments or decrements | Real: $(-\infty, +\infty)$ |
| Inverse Gaussian | Positively skewed data with distribution's tail decreases slowly | Real: $(0, +\infty)$ |
| Gamma | Exponential response data, increase or decrease with constant ratio | Real: $(0, +\infty)$ |

| Distribution | Typical Uses | Support of Distribution |
|---|---|---|
| Binomial | Single outcome from N occurrences | Integer: 0,1,2 … , N |
| Poisson | Count data | Integer: 0,1,2 … |



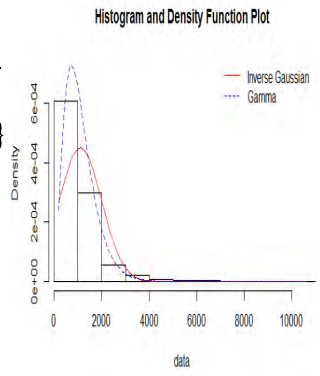
# Generalized Linear Model (cont'd)
## – Link Function

The **relationship** between the mean of the **response variable** distribution function and a linear combination set of **predictors**

Numerical example for a Gamma Log Link Model

$$\ln[E(Severity_i)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$
$$E(Severity_i) = Exp\{\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\}$$

$$\mu_i = E[Y_i] = E(Severity_i) = 3{,}000$$
$$\ln[E(Severity_i)] = \ln[3{,}000] = 8.01$$
$$E(Severity_i) = Exp\{8.01\} = 3{,}000$$



# Generalized Linear Model (cont'd)
## – Link Function

| Distribution | Link Name | Link Function, $X\beta = g(\mu)$ | Mean Function |
|---|---|---|---|
| Normal | Identity | $X\beta = \mu$ | $\mu = X\beta$ |
| Inverse Gaussian | Inverse Squared | $X\beta = \frac{1}{\mu^2}$ | $\mu = (X\beta)^{-\frac{1}{2}}$ |
| | Log | $X\beta = \ln(\mu)$ | |
| Gamma | Inverse | $X\beta = \frac{1}{\mu}$ | $\mu = -(X\beta)^{-1}$ |
| | Log | $X\beta = \ln(\mu)$ | |
| Binomial | Logit | $X\beta = \ln(\frac{\mu}{1-\mu})$ | $\mu = \frac{\exp(X\beta)}{1+\exp(X\beta)}$ |
| Poisson | Log | $X\beta = \ln(\mu)$ | $\mu = \exp(X\beta)$ |

Exponential Family



# Step 6. Specifying Model Form
## – Severity Model Example

| | |
|---|---|
| Objective | Predict the Expected Severity of Motor Insurance |
| Response Variable | Severity = Claim Amount / Claim Count |
| Predictors | Sum Insured, Underwriting Year, Cubic Capacity of Vehicle, Manufacturer of Vehicle, Manufactured Year, Region |
| Weights | Claim Count |
| Models | Inverse Gaussian Distribution or Gamma Distribution |
| Link Function | Log Link Inverse Gaussian and Log Link Gamma |

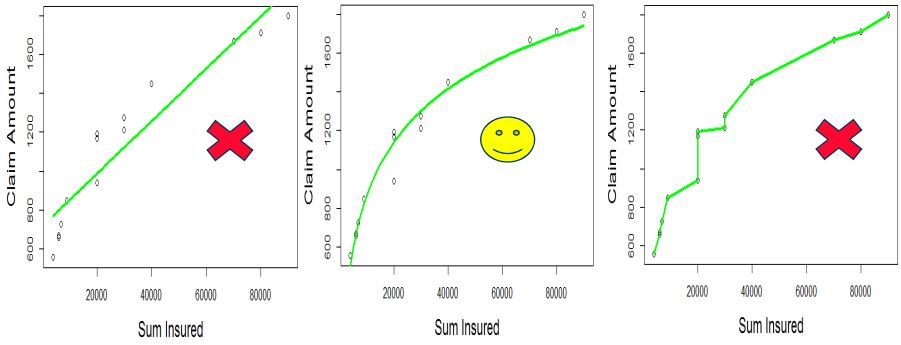
# Step 7. Model Validation and Diagnostics

Validation Set

Model Validation

Test for overfitting or underfitting using validation set

| Under fitting | Fitting | Over fitting |
|---|---|---|






kwh11
kwh12

# Step 8. Models Comparison
## – Goodness of Fit Test

Validation Set

| | |
|---|---|
| Coefficient of determination, $R^2/Adjusted\ R^2$ | $R^2 = 1 - \frac{Residual\ Sum\ of\ Squares}{Total\ Sum\ of\ Squares} = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2}$ <br> $R^2\ Adjusted = 1 - (1 - R^2)(\frac{n-1}{n-p-1})$ |
| Likelihood, $L(\theta)$ or Log-likelihood, $l(\theta)$ | $L(\theta) = \prod_i^n f(x_i\|\theta)$ or $l(\theta) = \sum_i^n \log[f(x_i\|\theta)]$ |
| Akaike Information Criterion | $AIC = -2(log - likelihood) + 2(n_{paramter})$ |
| Pearson Chi-Squared | $\chi^2 = [\frac{(Observed\ Value\ - Expected\ Value)^2}{Expected\ Value}]$ |

**Slide 30**

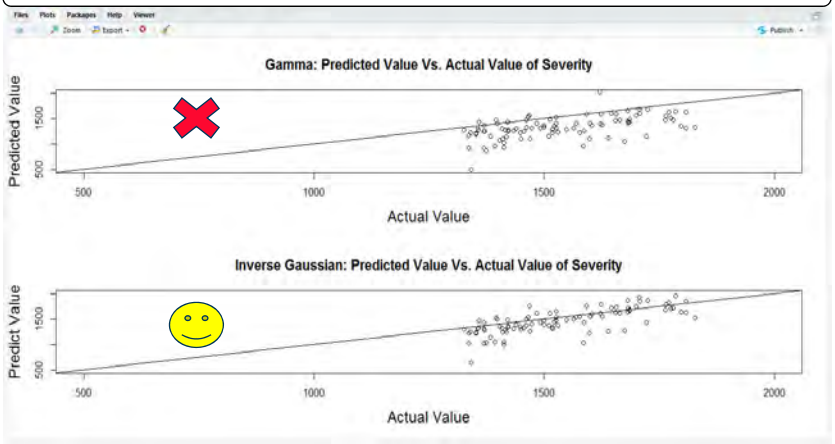
**kwh11** added a new variable (predicted) will increase the Total Sum of Square (SStotal) while the SSerror might not reducing or in fact increase but at the ratio of lower than the increase of SStotal
khong wei hung, 11/11/2018

**kwh12** So the R squared will increase. To avoid this circumstance, Adjusted R squared is introduced
khong wei hung, 11/11/2018

# Step 8. Models Comparison (cont'd)
## – Goodness of Fit Test

Validation Set

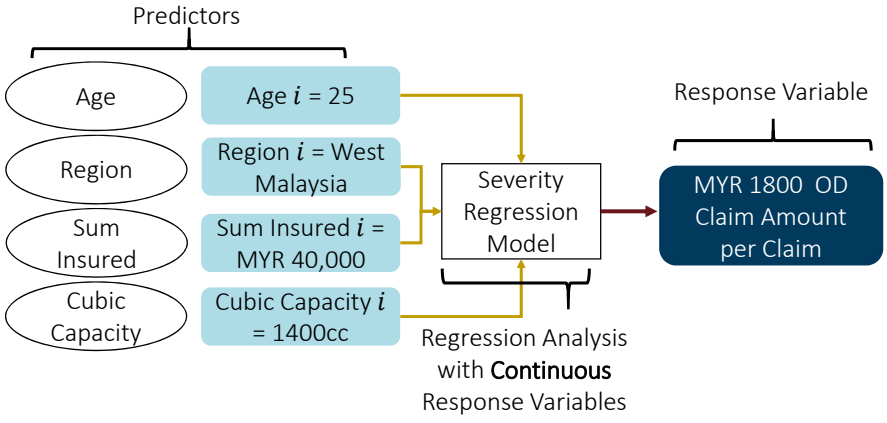
Assessing with plot of the Actual vs. Predicted Value to select a final model



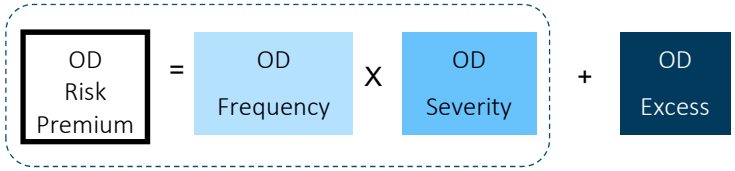



# Step 9. Model Selection
## – Final Model

Validation Set



## Step 9. Model Selection (cont'd)
– Final Model

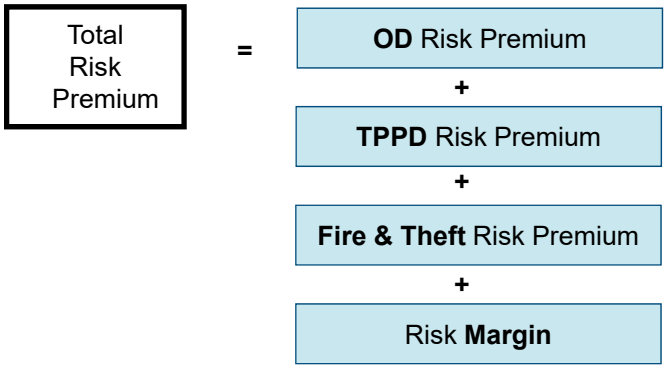



Any trending adjustments will take place at the frequency and severity model level (judgement required)

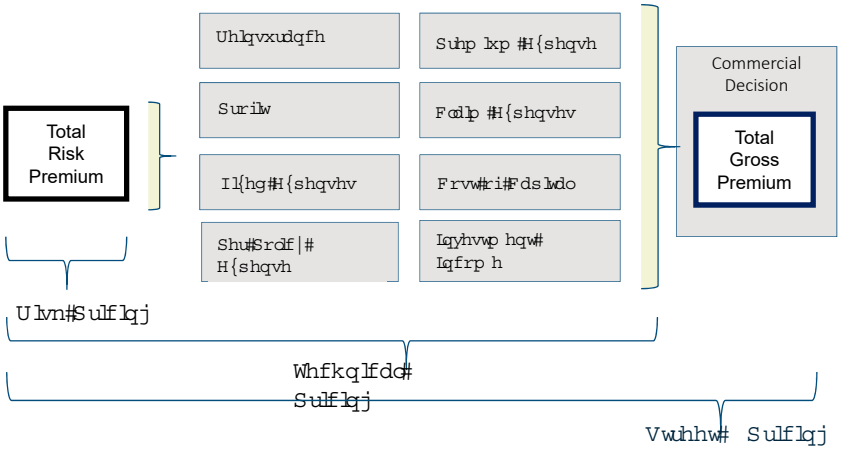
OD Excess is the estimated loading for the large losses excluded from the dataset (judgement required)



## Step 9. Model Selection (cont'd)
– Net Rating



# Step 9. Model Selection (cont'd)
## – Gross Rating

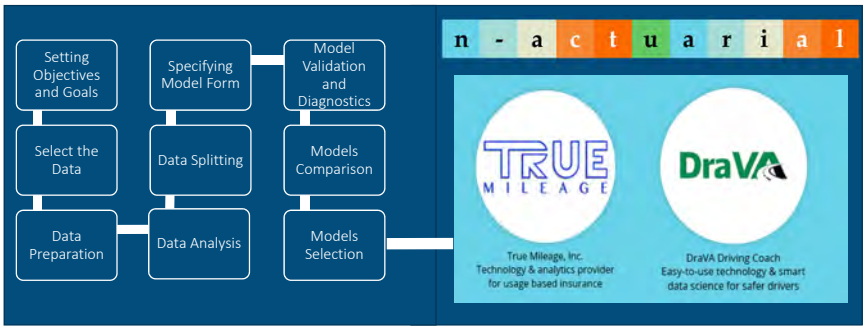





# GLM – Big Data
## Step 10

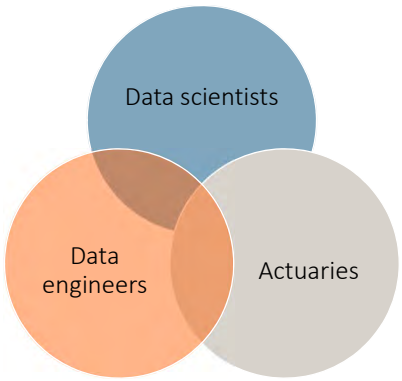
SOCIETY OF ACTUARIES

# Step 10. Process Improvement
- Upskilled actuaries






# Step 10. Process Improvement
- Upskilled actuaries



# Step 10. Process Improvement
## - Upskilled actuaries

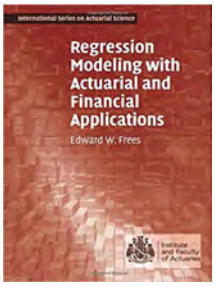
SOCIETY OF ACTUARIES › 2018 December Exam PA

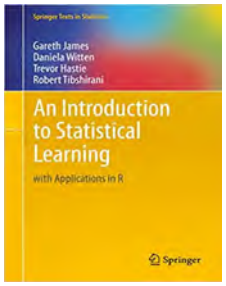
- ❑ Predictive Analytics Problems and Tools
- ❑ - (**R, RStudio**)
- ❑ Problem Definition
- ❑ **Data** Visualization
- ❑ **Data** Types and Exploration
- ❑ **Data** Issues and Resolutions
- ❑ Generalized Linear Models
- ❑ Decision Trees
- ❑ Cluster and Principal Component Analyses
- ❑ Communication

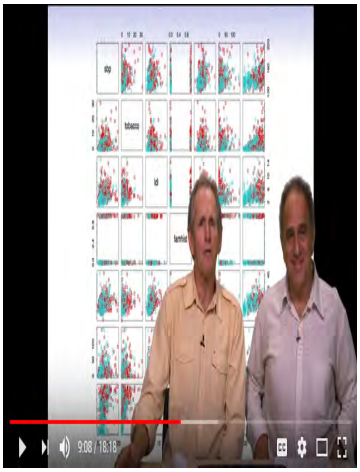
https://www.soa.org/Education/Exam-Req/edu-exam-pa-detail.aspx

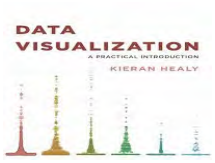


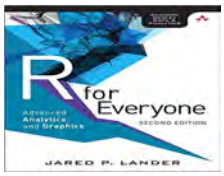
# Step 10. Process Improvement
## - Upskilled actuaries





## Actuaries in action

- Analyze, measure, convert and manage risk
- Use math, statistical skills, financial theory, business knowledge, and an understanding of human behavior


- Develop and validate financial models to guide decision making and turn risk into opportunity



## Questions?