Introduction to Ratemaking

Multivariate Methods

March 15, 2010

Jonathan Fox, FCAS,
GuideOne Mutual Insurance Group

Content Preview

1. Theoretical Issues
2. One-way Analysis Shortfalls
3. Multivariate Methods

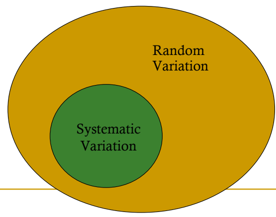
There are several theoretical stumbling blocks to overcome to develop rating relativities

- Separating the Signal from the Noise
- Not double counting Correlated Exposures
- Addressing Variable Interactions

1. Insurance is inherently a stochastic, or random, process

Any set of data you examine will contain:

- *systematic* variation - signal, true relationships
- *random* variation - noise



Note: the presence of noise along with our signal is the basic reason credibility was conceived. Due to the presence of noise, we don't fully believe our point estimate.

Good multivariate methods provide credibility statistics for estimates.

Filtering the Signal from the Noise requires credibility estimates based on measuring Process Variance

For example:
A coin is tossed 10 times and only lands on heads 4 times (40%).

Is the coin biased?

Over last three years territory 101 has a fire frequency of 0.005 while territory 102 has a fire frequency of 0.007.

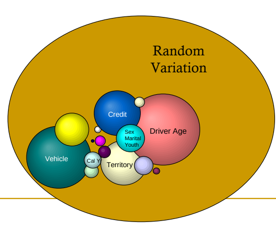
Can we say that territory 102 is truly more risky than 101?

To answer these questions we need a measure of the variability in each processes

2. The 'signal' – once detected – is usually made up of inter-related effects

Correlations

Often, distributional biases in *exposures* exist in the data and cause results to be linked

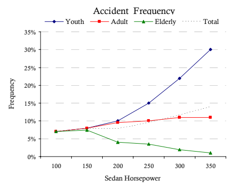


Possible Example:
If all Youth live the City and all Adults live in Rural areas, then Age and Territory effects will be correlated

3. The ‘signal’ is usually made up of inter-related effects

Interactions

This occurs when two variable’s *indicated factors* are correlated; hence the outcome of one depends on the level of the other



Possible Example:
Youngsters with muscle cars are more likely to drive them often and fast while senior citizens with muscle cars are more likely to leave them in the garage and polish them: Horsepower and Driver Age interact

Summary on Theoretical issues

Processes Variability measures allow us to gauge the strength of the ‘signal’ or indicated factor estimates.

Correlations between two variables’ exposure distributions cause the indications to be linked. This is NOT an interaction; it is an important effect and multivariate techniques can resolve this problem.

Interactions are correlations between two variables’ indicated factors: the indicated factors behave differently across levels of a secondary variable.

It is perfectly possible for two variables to be correlated but have no interaction, or for two variables to have an interaction but *not* be correlated.

One-way Analysis Techniques and their Shortfalls

Pure Premium Example

Youthful		Mature	
Losses:	$266,667	Losses:	$133,333
Exposures:	2,000	Exposures:	2,000
Pure Premium:	$133	Pure Premium:	$67
Relativity:	2.00	Relativity:	1.00

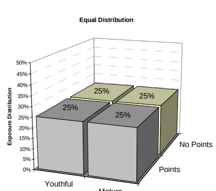
Problem: One-way Pure Premium analysis is blind to the rest of the class plan

No Points	
Losses:	$150,000
Exposures:	2,000
Pure Premium:	$75
Relativity:	1.00
Driver With Points	
Losses:	$250,000
Exposures:	2,000
Pure Premium:	$125
Relativity:	1.67

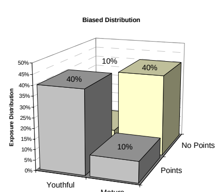
Should a Young Driver with Points be charged 3.33 times the rate of a clean Adult (2.00 * 1.67) ?

No Points	
Losses:	$150,000
Exposures:	2,000
Pure Premium:	$75
Relativity:	1.00
Driver With Points	
Losses:	$250,000
Exposures:	2,000
Pure Premium:	$125
Relativity:	1.67

Should a Young Driver with Points be charged 3.33 times the rate of a clean Adult (2.00 * 1.67) ?



Yes – If Pointed drivers are evenly *spread through Youth and Matures. Then no correlation exists between Pointed vs. Youthful exposures*



No – If Young drivers are more (or less) likely to have Points. Then a correlation will exist between the exposures, and one-way analysis will be distorted by the bias.

Example Detail

Exposures

	Youth	Mature	Total
No Points	400	1,600	2,000
Points	1,600	400	2,000
Total	2,000	2,000	4,000

Exposures

	Youth	Mature
No Points	20%	80%
Points	80%	20%
Total	100%	100%

Losses

	Youth	Mature	Total
No Points	66,667	83,333	150,000
Points	200,000	50,000	250,000
Total	266,667	133,333	400,000

Losses

	Youth	Mature
No Points	25%	62%
Points	75%	38%
Total	100%	100%

Pure Premium

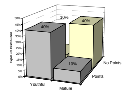
	Youth	Mature	Total
No Points	167	52	75
Points	125	125	125
Total	133	67	100

Pure Premium

	Youth	Mature
No Points	1.25	0.78
Points	0.94	1.88
Total	1.00	1.00

One-Way Relativity

	Youth	Mature	Points One-way
No Points	2.00	1.00	1.00
Points	3.33	1.67	1.67
Age One-way	2.00	1.00	



Two-Way Relativity

	Youth	Mature	Points One-way
No Points	3.20	1.00	1.00
Points	2.40	2.40	1.67
Age One-way	2.00	1.00	

There are several problems with one-way analysis

- Usually does not provide a measure of significance
- Can overlook Exposure Correlations
- Not sensitive to Factor Interactions

Multivariate Techniques can overcome these shortfalls

- Multi-way Pure Premium
- Loss Ratio
- Minimum Bias
- Multi Linear Regression
- Generalized Linear Regression

One-way Loss Ratios are inherently Multivariate: the premium takes into account the rest of the class plan

$$\text{Youth LR} = 73\% = \frac{\sum \text{Loss}}{\sum \text{Premium}} = \frac{\sum \text{Loss}}{\sum (\text{Base Rate}) \cdot \text{Rel}_{\text{Age}} \cdot \text{Rel}_{\text{Points}} \cdot \text{Rel}_{\text{Terr}} \ldots}$$

For example, if you look at the relative *loss ratios* between Youthful and Adult drivers, the premium within that loss ratio will reflect the current factors for Points.

Because Youthfuls have a higher percentage of Points, their average premium will be higher due to the higher Points factors. This will lower the loss ratio. In this way we don't "double count" the effect of Points and Age.

Side note…what if Points didn't exist? *Appropriate Age factors would change.*

One-way Loss Ratio analysis has a significant shortcoming

It assumes the rating plan's <u>other factors</u> are accurate.

$$= \frac{\sum \text{Loss}}{\sum (\text{Base Rate}) \cdot \text{Rel}_{\text{Age}} \cdot \text{Rel}_{\text{Points}} \cdot \text{Rel}_{\text{Terr}} \ldots}$$

This assumption is often not appropriate, as is the case when there are *multiple* changes which need to be made.

e.g. suppose you want to examine the adequacy of both your Age and Points curves. When you look at loss ratios by Age, you are assuming your current Points factors are good and vice versa for when you look at loss ratios by Points.

Minimum Bias Techniques overcome several common short comings

Min Bias is an iterative approach for reducing the error between observed and indicated relativities

- Optimizes the relativities for multiple changes
- Can use either Pure Premiums or Loss Ratios
- Calculates relativities which minimize the error with observed relativities based on a selected error (Bias) function
- The Bias Function determines how the error which can't be removed is then distributed
- Iterative technique - when the equations converge you have the optimal solution

Example 1: Use the Minimum Bias procedure to find the optimal relativities for Age and Points

Observed Pure Premiums

	Youth	Adult	Total	Rel
Points	333	167	500	**1.67**
No Points	200	100	300	**1.00**
Total	533	267		
Rel	**2.00**	**1.00**		

Indicated Relativity

	Youthful	Adult
Points	2.00 x 1.67	1.00 x 1.67
No Points	2.00 x 1.00	1.00 x 1.00

	Youthful	Adult
Points	3.33	1.67
No Points	2.00	1.00

Indicated Pure Premium

	Youthful	Adult
Points	333	167
No Points	200	100

Error (Indication vs. Actual)

	Youthful	Adult
Points	0	0
No Points	0	0

Perfect Match – no need to use Minimum Bias Approach

Example 2: Use the Minimum Bias procedure to find the optimal relativities for Age and Points

Observed Pure Premiums

	Youthful	Adult	Total	Rel	
Points	400	200	600	**1.50**	P1
No Points	300	100	400	**1.00**	P2
Total	700	300			
Rel	**2.33**	**1.00**			
	A1	A2			

Indicated Relativity

	Youthful	Adult
Points	2.33 x 1.50	1.00 x 1.50
No Points	2.33 x 1.00	1.00 x 1.00

Points	3.50	1.50
No Points	2.33	1.00

Indicated Pure Premium

	Youthful	Adult
Points	350	150
No Points	233	100

Error (Indication vs. Actual)

	Youthful	Adult
Points	-50	-50
No Points	-67	0

Imperfect Match - use Minimum Bias Approach

Which Bias Function to Chose?

Balance Principle
Ensures that the total error *in each class* is zero

Least Squares
Minimizes the *total* error, using the Sum of Squared errors (nominal values)

Chi Squared
Minimizes the *total* error, using the Sum of Squared errors (as a Percent of expected)

Example 2: ‘Balance Principle’ Bias Function (1st Iteration)

Balance Principle ensures the total error in each class dimension is zero

Observed Pure Premiums

	Youthful	Adult	Total	Rel
Points	400	200	600	P1
No Points	300	100	400	P2
Total	700	300		
Rel	A1	A2		

Assumed Initial Relativity for Age:

	Symbol	Initial Value
Youthful	A1	2.333
Adult	A2	1.000

Formula representation of Indicated Pure Prem

	Youthful	Adult	Total	Formula Equivalent
Points	100 x A1 x P1	100 x A2 x P1	600	= (100 x A1 x P1) + (100 x A2 x P1)
No Points	100 x A1 x P2	100 x A2 x P2	400	= (100 x A1 x P2) + (100 x A2 x P2)

600 = (100 x 2.33 x P1) + (100 x 1.00 x P1)
400 = (100 x 2.33 x P2) + (100 x 1.00 x P2)

1st Iteration Solution for Points Variables

	Symbol	Solved Value	Rebased
Points	P1	1.8000	1.5000
No Points	P2	1.2000	1.0000

Next we substitute in the solved values for ‘Points’ to get new values for ‘Age’

Example 2: ‘Balance Principle’ Bias Function (1st Iteration)

Balance Principle ensures the total error in each class dimension is zero

Observed Pure Premiums

	Youthful	Adult	Total	Rel
Points	400	200	600	P1
No Points	300	100	400	P2
Total	700	300		
Rel	A1	A2		

Relativity from Prior Iteration

	Symbol	Value
Points	P1	1.500
No Points	P2	1.000

	Youthful	Adult
Points	100 x A1 x P1	100 x A2 x P1
No Points	100 x A1 x P2	100 x A2 x P2
	700	300

700 = (100 x A1 x P1) + (100 x A1 x P2)
300 = (100 x A2 x P1) + (100 x A2 x P2)

700 = (100 x A1 x 1.50) + (100 x A1 x 1.00)
300 = (100 x A2 x 1.50) + (100 x A2 x 1.00)

1st Iteration Solution for Age Variables

	Symbol	Solved Value	Rebased
Youth	A1	2.8000	2.3333
Adult	A2	1.2000	1.0000

	1st Iter.	2nd Iter.
Points	1.5000	1.5000
No Points	1.0000	1.0000
Youthful	2.3333	2.3333
Adult	1.0000	1.0000

Next we substitute in the solved values for ‘Points’ to get new values for ‘Age’

Convergence

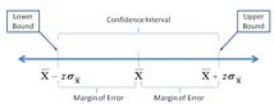
Example 2 Notes:

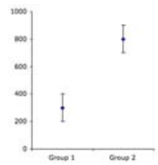
- Immediate convergence not always the case
- Method can be extended to many dimensions
- Possible to code the calculation directly into a spreadsheet (‘short cut’ formulas exist)
- Can use with multiplicative or additive pricing models
- Could use Least Squares, Chi Squared, or Maximum likelihood approaches instead
- For more information see CAS publication:
  The Minimum Bias Procedure, A Practioner's Guide, by Feldblum and Brosius

Minimum Bias techniques still have limitations

These techniques only give Point Estimates, yet we know all data contains both signal and noise.

Minimum bias techniques provide no method for quantifying the extent and impact of the noise.

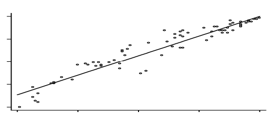




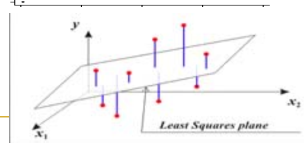
Classical Statistics Techniques provide a method for quantifying noise vs. signal

General Form: Dependent Variable = Signal + Noise

Simple Linear Regression:



Multiple Linear Regression:



With Insurance applications we use the rating factors as the Dimensions in the regression

- Observed Pure Premiums or Loss Ratios are used to determine the parameter values, or 'fit' the model
- Usually Categorical variables are used instead of Quantitative variables
- Specifying a Categorical model differs from a Quantitative model:
  - Categorical: each *Level* of a *Variable* has it's own parameter
  - Quantitative: each *Variable* has a 'slope' for all levels within it

We specify a Categorical model so each 'cell' is uniquely represented using 'dummy variables'

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

This two-dimensional example is formulated as…

$$y = \beta_{Youth} x_1 + \beta_{Adult} x_2 + \beta_{Clean} x_3 + \varepsilon$$

The x's take on values of 0 or 1

The default is a 'Pointed' driver

$$4{,}500 = \beta_{Youth}\cdot 1 + \beta_{Adult}\cdot 0 + \beta_{Clean}\cdot 0 + \varepsilon$$
$$1{,}500 = \beta_{Youth}\cdot 1 + \beta_{Adult}\cdot 0 + \beta_{Clean}\cdot 1 + \varepsilon$$
$$7{,}500 = \beta_{Youth}\cdot 0 + \beta_{Adult}\cdot 1 + \beta_{Clean}\cdot 0 + \varepsilon$$
$$5{,}000 = \beta_{Youth}\cdot 0 + \beta_{Adult}\cdot 1 + \beta_{Clean}\cdot 1 + \varepsilon$$

Solve the Parameters (β) by substituting in the observed Pure Premiums

$$4{,}500 = \beta_1\cdot 1 + \beta_2\cdot 0 + \beta_3\cdot 0 + \varepsilon$$
$$1{,}500 = \beta_1\cdot 1 + \beta_2\cdot 0 + \beta_3\cdot 1 + \varepsilon$$
$$7{,}500 = \beta_1\cdot 0 + \beta_2\cdot 1 + \beta_3\cdot 0 + \varepsilon$$
$$5{,}000 = \beta_1\cdot 0 + \beta_2\cdot 1 + \beta_3\cdot 1 + \varepsilon$$

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

Usually the system of equations will not have a fit that perfectly explains all variation. W*hat fit will be best at minimizing the error ?*

To find an answer, we need a criterion for what is the "best" answer.

A typical approach is to minimize the sum of the squared errors (SSE).

1. $SSE = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2 + \cdots$ where ε = (observed – expected)
2. Minimize by taking the derivative with respect to beta: $\delta SSE/\delta\beta_i$
3. Set the derivative equal to zero and solve for β_i: $\delta SSE/\delta\beta_i = 0$

Solve the Parameters (β) by substituting in the observed Pure Premiums and Minimizing the SSE

$$4{,}500 = \beta_1\cdot 1 + \beta_2\cdot 0 + \beta_3\cdot 0 + \varepsilon$$
$$1{,}500 = \beta_1\cdot 1 + \beta_2\cdot 0 + \beta_3\cdot 1 + \varepsilon$$
$$7{,}500 = \beta_1\cdot 0 + \beta_2\cdot 1 + \beta_3\cdot 0 + \varepsilon$$
$$5{,}000 = \beta_1\cdot 0 + \beta_2\cdot 1 + \beta_3\cdot 1 + \varepsilon$$

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

$$SSE = \sum_i (\varepsilon)^2 = \sum_i (\hat{y}_i - y_i)^2$$
$$= (4{,}500 - \beta_1)^2 + (1{,}500 - \beta_1 - \beta_3)^2 + (7{,}500 - \beta_2)^2 + (5{,}000 - \beta_2 - \beta_3)^2$$

$$\delta SSE/\delta\beta_1 = 2\cdot(4{,}500 - \beta_1)\cdot(-1) + 2\cdot(1{,}500 - \beta_1 - \beta_3)\cdot(-1) + 0 + 0 \overset{set}{=} 0$$

$$\delta SSE/\delta\beta_2 = 0 + 0 + 2\cdot(7{,}500 - \beta_2)\cdot(-1) + 2\cdot(5{,}000 - \beta_2 - \beta_3)\cdot(-1) \overset{set}{=} 0$$

$$\delta SSE/\delta\beta_3 = 0 + 2\cdot(1{,}500 - \beta_1 - \beta_3)\cdot(-1) + 0 + 2\cdot(5{,}000 - \beta_2 - \beta_3)\cdot(-1) \overset{set}{=} 0$$

Solve the Parameters (β) by substituting in the observed Pure Premiums and Minimizing the SSE

$$4{,}500 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \varepsilon$$
$$1{,}500 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \varepsilon$$
$$7{,}500 = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \varepsilon$$
$$5{,}000 = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 1 + \varepsilon$$

Loss	*Clean*	*Pointed*
Younger	1,500	4,500
Older	5,000	7,500

$$2 \cdot \beta_1 + \beta_3 = 6{,}000$$
$$2 \cdot \beta_2 + \beta_3 = 12{,}500$$
$$\beta_1 + \beta_2 + 2 \cdot \beta_3 = 6{,}500$$

$$\beta_1 = 4{,}375$$
$$\beta_2 = 7{,}625$$
$$\beta_3 = -2{,}750$$

Finding the optimal answer for a multi-linear regression boils down to systems of equations

$$4{,}500 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \varepsilon$$
$$1{,}500 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \varepsilon$$
$$7{,}500 = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \varepsilon$$
$$5{,}000 = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 1 + \varepsilon$$

- Expressing a System of Equations is more conveniently done via Matrix Notation and Linear Algebra

  …Especially as the number of variables and observations get more numerous

Matrix notation allows systems of equations to be more elegantly represented

$$4{,}500 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + \varepsilon$$
$$1{,}500 = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 1 + \varepsilon$$
$$7{,}500 = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 0 + \varepsilon$$
$$5{,}000 = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 1 + \varepsilon$$

$$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon}$$ ← *Y = signal + noise*

$$\begin{bmatrix} 4500 \\ 1500 \\ 7500 \\ 5000 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \bullet \begin{bmatrix} \beta_{Youth} \\ \beta_{Adult} \\ \beta_{Clean} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

4 obs — Youth — Adult — Clean — solving for these

We think of the Linear Model as having Three Parts:

$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon}$

Random Component: The Observations being predicted

Systematic Component: The Predictor variables, also notated as η ('eta')

Link function: Defines the relationship between the predictors and the observations

Linear Modeling is subject to some assumptions that may not fit Insurance applications well

$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon}$

Random Component: Observations are independent and come from a normal distribution with a common variance.

Systematic Component: Predictor variables are related as a Linear sum, η

i.e. Predictors are related as a linear combination:
$\eta = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots$

Link function: The expected value of Y is equal to η

$E[Y] = \eta$

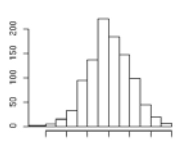
Linear Modeling is subject to some assumptions that may not fit Insurance applications well

$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon}$

Random Component: Observations are independent and come from a normal distribution with a common variance.

For each variable in our model, there is an expected mean and randomness about that mean. The average loss for "younger drivers" may be $100, but why should the distribution of individual observations be Normal about this?

In fact, normal distributions extend to negative numbers. What's a negative loss?



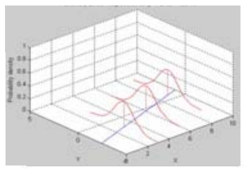
Linear Modeling is subject to some assumptions that may not fit Insurance applications well

$$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon}$$

Random Component: Observations are independent and come from a normal distribution with a common variance.

Why should the distribution of losses for 25K limits have the same variance as the distribution of losses for 100K limits?

The 25K limits, with a low mean, would likely have *less* variance than 100K limits.



Linear Modeling is subject to some assumptions that may not fit Insurance applications well

$$E[Y] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots$$

Systematic Component: The Predictor variables are a Linear sum, η

$\eta = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots$

Link function: The expected value of Y is equal to η

This pair assumes that Y is predicted by the *additive* combination of the X variables.

However, most insurance effects tend to combine *multiplicatively*.

$E[Y] = (\beta_1 x_1) * (\beta_2 x_2) * (\beta_3 x_3) * \ldots$

Round up of Multi-variate approaches and their limitations

One-way Pure Premium:

Youthful	Mature

- No measure of Significance (Point estimate only)
- Can overlook Exposure Correlations
- Not sensitive to Factor Interactions

Loss Ratio:

$$= \frac{\sum \text{Loss}}{\sum (\text{Base Rate}) \cdot \text{Rel}_{\text{Age}} \cdot \text{Rel}_{\text{Points}} \cdot \text{Rel}_{\text{Terr}} \ldots}$$

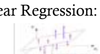
- Is Sensitive to correlations
- But assumes all other pricing factors accurate: can't use if changing multiple structures
- Provides Point estimates only

Minimum Bias:

	Youthful	Adult
Points	100 x A1 x P1	100 x A2 x P1
No Points	100 x A1 x P2	100 x A2 x P2

- Does accommodate multiple structure changes
- But only Provides Point estimates

Multi Linear Regression:
(MLR)



- Does accommodate multiple structure changes
- Provides Confidence Ranges & Significance tests
- But requires assumptions that don't fit insurance data

Generalized Linear Models use more lenient assumptions

$$\underline{Y} = g^{-1}(\mathbf{X}.\underline{\beta}) + \underline{\varepsilon}$$

Random Component: Observations are independent, but come from one of the family of Exponential Distributions (Normal, Poisson, Gamma,…)

now the variance can change with the mean and negative values can be prohibited

Systematic Component: Predictor variables are related as a Linear sum, η

No Change here: $\eta = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots$

Link function: The expected value of Y is equal to a transformation of η: $g(E[Y]) = \eta$ *or* $E[Y] = g^{-1}(\eta)$

A log link results in a multiplicative relationship between the X's

log-link: $g(x) = \ln(x)$ → $g(E[Y]) = \eta = \ln(E[Y])$ → $E[Y] = e^{(\eta)} = e^{(x_1\beta_1 + x_2\beta_2)} = e^{(x_1\beta_1)} e^{(x_2\beta_2)}$

Generalized Linear Modeling assumptions are better suited for Insurance applications

$$\underline{Y} = g^{-1}(\mathbf{X}.\underline{\beta}) + \underline{\varepsilon}$$

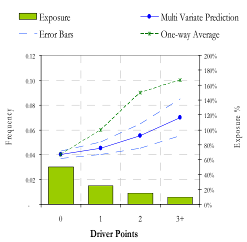
First Step is choosing a Link function.

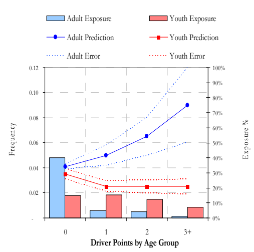
You must select a Distributional Family that mimics your data.

Then you need to decide on your design matrix (which variables to include in your model and how to combine them)?

This process is best done through an evaluative, trial and error process that combines both statistics and judgment.

GLM output can show Predictions vs. Error, Correlated effects, and Interactions





In summary of GLMs …

As a statistical model, GLMs allow us to have some measure of the *Noise* as well as the *Signal*.

GLMs *assumptions* are flexible enough to reasonably fit real-world insurance situations.

It turns out that many Minimum Bias techniques, all One-way, and all Linear Regression approaches are just *special forms* of GLMs.

GLMs are *multivariate* and automatically solve the "double counting" problem presented by *correlated* variables. They also allow for many model forms, including *interactions*.

GLM's are fairly standard in the industry but there are other, *non-linear* multivariate techniques as well

- Decision Trees (CART, C5, CHAID, etc.)
- Neural Networks
- Polynomial Networks
- Clustering
- Kernels
- Others…