Experience Studies and Assumption Development

Southeastern Actuaries Conference
November 20, 2008
Atlanta, Georgia

David Andreae, FSA, MAAA
Unum

General Overview

- Data issues
- Study methodology
- Assumption development
  - Visual approach
  - Linear regression
  - Logistic regression

Data Issues

- Count or amount basis
- Identify data items to collect
- Determine expected/valid values
- Strategy for unexpected/missing values
  - Ignore this observation/policy
  - Change to default value
- Dealing with values that change over time
  - Use initial value
  - Use current value
  - Use dynamic value

Data Issues

- Validate to source believed to be accurate
  - Policy count
  - In-Force premium
  - Claim count
  - Paid claims

- Validate against model/projection system

Study Methodology

- Determine time period for study
  - Advantages of longer study period
    - Broader range of economic/environmental influences
    - More months/years of exposure add credibility
  - Reasons to restrict study period
    - Changes in products
    - Changes in underwriting/marketing
    - Environment has changed non-cyclically
    - Reporting lags
- Selected expected basis or start from scratch?
- Define exposure

Exposure Calculation Example

Policies In Force	Number of Deaths
100	20

Deaths occur on average at mid-month

==========

Method 1 Exposure = 100 - .5 *20
= 90

Death rate = 20 / 90 = 22.2%

Projection model should apply mortality rate to:
(# in-force policies at start * (1 - .5 * Death rate))

==========

Method 2 Exposure = 100

Death rate = 20 / 100 = 20%

Projection model should apply mortality rate to:
(# of in-force policies at start)

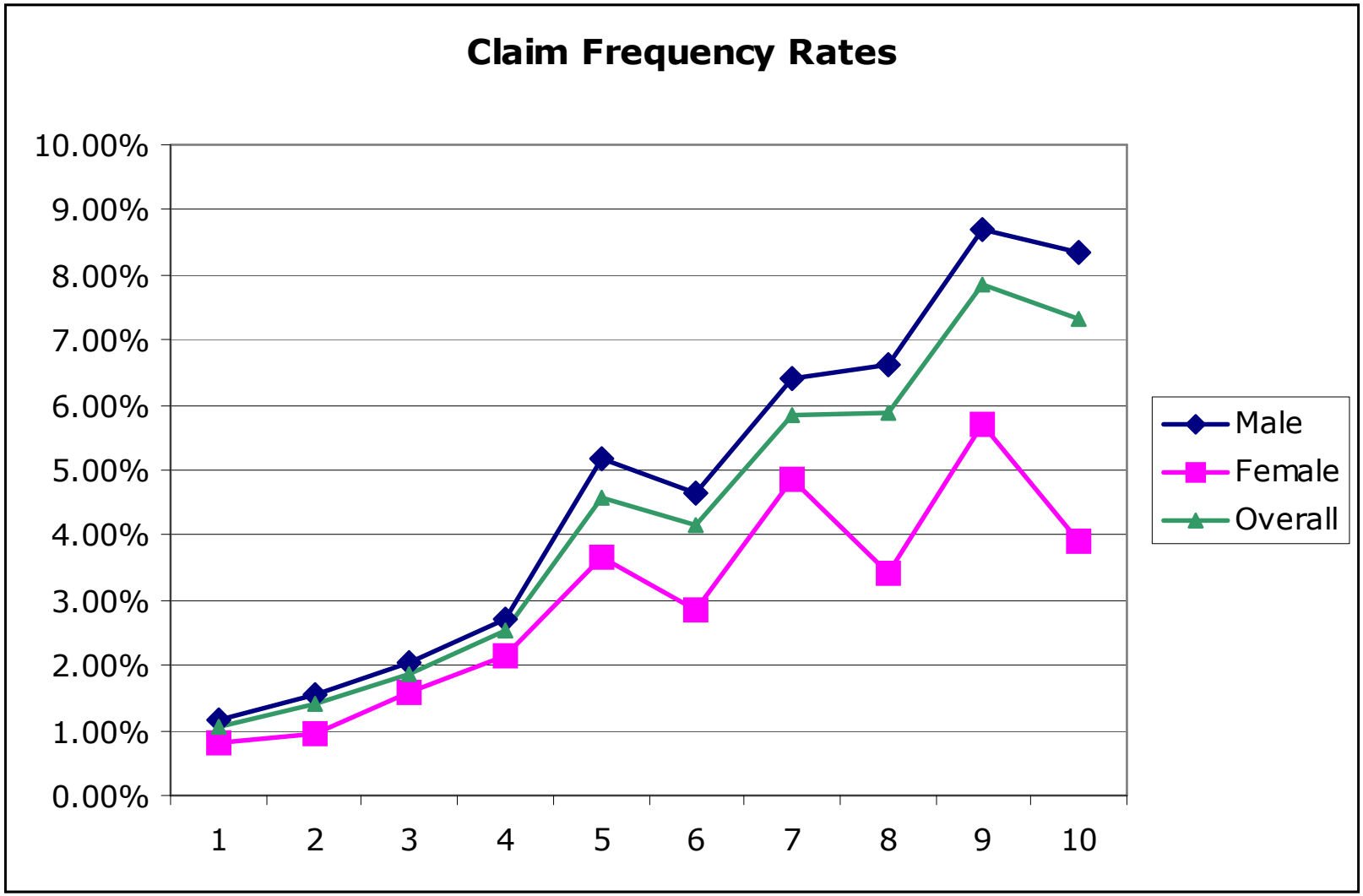
Assumption Development – Visual Approach

- Pivot tables and charts (personal preference)
- Iterative, trial and error approach
- Smoothing by hand/eye
- Start with base assumption reflecting one or two variables
- Add modifiers one or two variables at a time
- Continuously check A/E results against variables already in the mix
- Significant judgment required

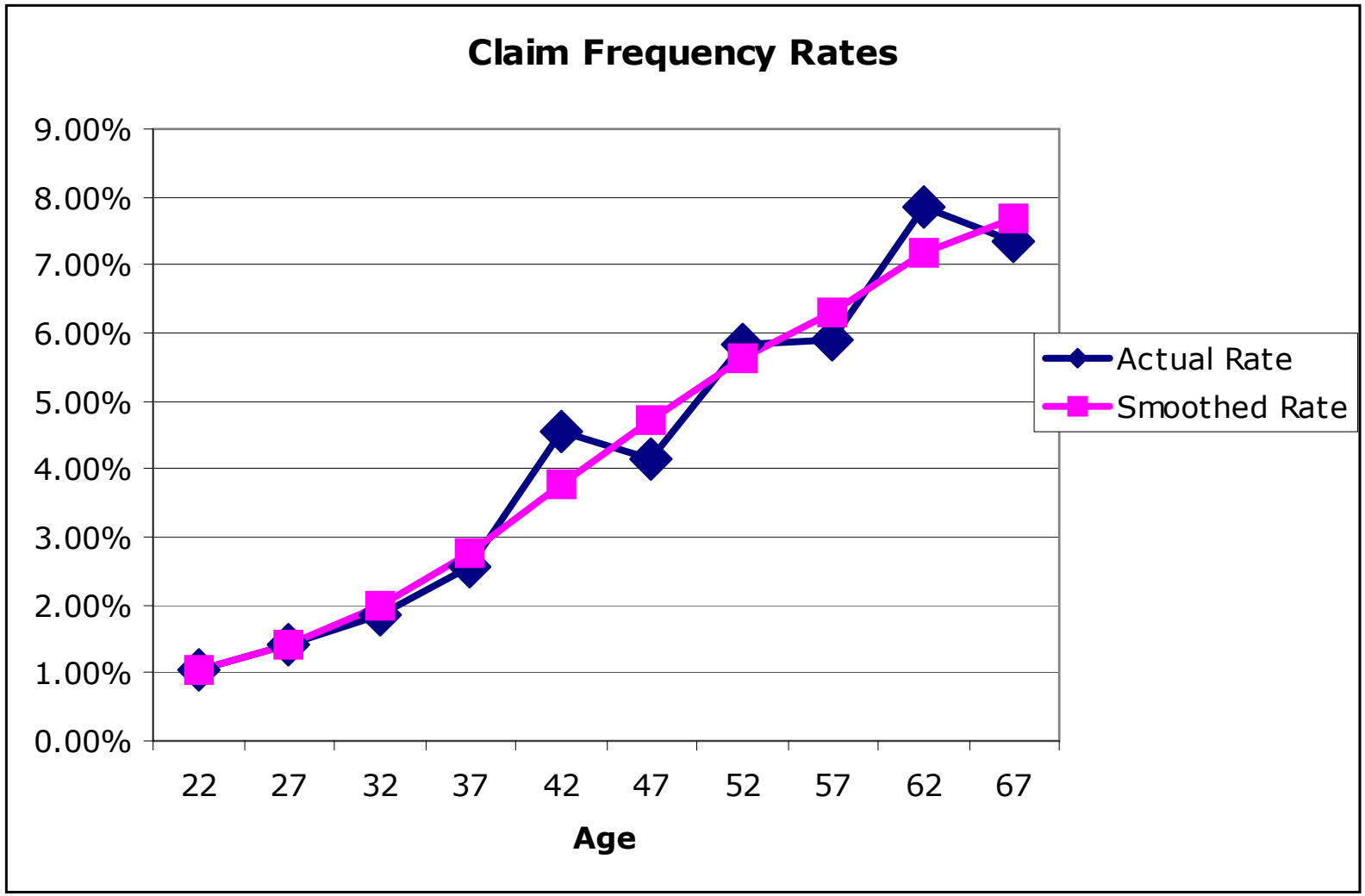
Visual Approach – Example

Age	Male Exposure	Female Exposure	Total Exposure	Male Claims	Female Claims	Total Claims
22	33,329	16,671	50,000	391	134	525
27	77,588	22,412	100,000	1,189	211	1,400
32	74,512	45,488	120,000	1,522	722	2,244
37	109,824	40,176	150,000	2,963	862	3,825
42	120,520	79,480	200,000	6,219	2,901	9,120
47	173,045	66,955	240,000	8,035	1,901	9,936
52	126,812	73,188	200,000	8,110	3,554	11,664
57	115,525	34,475	150,000	7,663	1,172	8,835
62	71,572	28,428	100,000	6,218	1,622	7,840
67	23,228	6,772	30,000	1,935	265	2,200

Visual Approach - Example



Visual Approach - Example



Visual Approach - Example

Age	exposure	claims	Actual Rate	Smoothed Rate	Expected Claims
22	50,000	525	1.05%	1.05%	525
27	100,000	1,400	1.40%	1.40%	1,400
32	120,000	2,244	1.87%	2.00%	2,400
37	150,000	3,825	2.55%	2.77%	4,155
42	200,000	9,120	4.56%	3.78%	7,560
47	240,000	9,936	4.14%	4.72%	11,328
52	200,000	11,664	5.83%	5.64%	11,280
57	150,000	8,835	5.89%	6.30%	9,450
62	100,000	7,840	7.84%	7.17%	7,170
67	30,000	2,200	7.33%	7.69%	2,307
		57,589			57,575

Visual Approach - Example

Age	Male Actual Rate	Female Actual Rate	Expected Rate	Male A/E	Female A/E
22	1.17%	0.80%	1.05%	111.7%	76.6%
27	1.53%	0.94%	1.40%	109.5%	67.2%
32	2.04%	1.59%	2.00%	102.1%	79.4%
37	2.70%	2.15%	2.77%	97.4%	77.5%
42	5.16%	3.65%	3.78%	136.5%	96.6%
47	4.64%	2.84%	4.72%	98.4%	60.2%
52	6.40%	4.86%	5.64%	113.4%	86.1%
57	6.63%	3.40%	6.30%	105.3%	54.0%
62	8.69%	5.71%	7.17%	121.2%	79.6%
67	8.33%	3.91%	7.69%	108.3%	50.9%
	4.78%	3.22%	4.30%	111.2%	75.0%

Linear Regression

- Sometimes used in combination with visual approach
- Determines the "equation" for a dependent variable (P) in terms of a set of dependent variables (X's)
- X terms can be transformations of other variables, e.g., age-squared, or combinations of other variables
- Form of the assumption equation:

$$P = \alpha + \beta_1 X_1 + \beta_2 X_{2+} + \ldots + \beta_i X_i$$

Linear Regression

- Excel allows you to do this for a single X at a time
- Sometimes results in predicted probability less than 0 or greater than 1
- In the example :

  death rate = -0.01359+.00137*Age
  -.01745*gender

  where male = 0 and female = 1
- For age 22 females this yields -0.09%

Logistic Regression

- Dependent (predicted) variable is the natural log of the odds of the event
- Varies from $-\infty$ to $+\infty$ instead of from 0 to 1
- Uses maximum likelihood estimation – an iterative approach - instead of least squares.
- Can test a large number of variables; SAS or SPSS will bring them in to the model one step at a time in order of predictive capability
- Several different "goodness of fit" tests
- Many good references available online

Comparison of Approaches