RPM Workshop 3: Basic Ratemaking

Introduction to Credibility

C. Nick Otterman, FCAS, MAAA
State Farm Insurance

March 2009

General Concept

- Set rate levels so that rates are “adequate, reasonable, and not unfairly discriminatory”
  - Adequate: Not too low
  - Reasonable: Not too high
  - Not unfairly discriminatory: Allocation of overall rate to individuals is based on cost justification

- At various steps in the ratemaking process, the concept of credibility is introduced

- The credibility of data is commonly denoted by the letter “Z”

Definitions of Credibility

- Common vernacular (Webster):
  - “Credibility” = the state or quality of being credible
  - “Credible” = believable
  - So, credibility is “the quality of being believable”
  - Implies you are either credible or you are not

- In actuarial circles:
  - Credibility is “a measure of the credence that…should be attached to a particular body of experience”

    -- L.H. Longley-Cook

  - Refers to the degree of believability of the data under analysis

    — A relative concept

Why Do We Need Credibility?

- Property / casualty insurance costs (losses) are inherently stochastic
  - Losses are fortuitous events
    - — Any given insured may or may not have a claim in a given year
    - — The size of the claim can vary significantly
  - Data can be viewed as an observation of a result
    - — Only one estimate of the “true” probability of having a claim and the distribution of sizes of claims
- So how much can we believe our data?

History of Credibility in Ratemaking

- The CAS was founded in 1914, in part to help make rates for a new line of insurance – workers compensation – and credibility was born out the problem of how to blend new experience with initial pricing
- Early pioneers:
  - Mowbray (1914) -- how many trials/results need to be observed before I can believe my data?
  - Albert Whitney (1918) -- focus was on combining existing estimates and new data to derive new estimates:

New Rate = Credibility*Observed Data + (1-Credibility)*Old Rate

  - Perryman (1932) -- how credible is my data if I have less than required for full credibility?
- Bayesian views were resurrected in the 1940's through the 1960's

Methods of Incorporating Credibility

- Limited Fluctuation
  - Limit the effect that random fluctuations in the data can have on an estimate
    - — "Classical credibility"
- Greatest Accuracy
  - Make estimation errors as small as possible
    - — Least squares credibility
    - — Empirical Bayesian Credibility
    - — Bühlmann Credibility
    - — Bühlmann-Straub Credibility

Limited Fluctuation Credibility Description

- "A dependable [estimate] is one for which the probability is high, that it does not differ from the [truth] by more than an arbitrary limit."

  -- Mowbray (1916)

- Alternatively, the credibility Z, of an estimate T, is defined by the probability P, that it is within a tolerance k%, of the true value

Limited Fluctuation Credibility Derivation

- New estimate = (Credibility)*(Data) + (1-Credibility)*(Prior Estimate)

$$E2 = Z*T + (1-Z)*E1$$

Add and subtract Z*E[T] →

$$E2 = Z*T + Z*E[T] – Z*E[T] + (1-Z)*E1$$

regroup →

$$E2 = (1-Z)*E1 + Z*E[T] + Z*(T–E[T])$$

Stability | **Truth** | **Random Error**

Limited Fluctuation Credibility Formula for Z

- Probability that “Random Error” is “small” is P
  - For example, the probability {random error is less than 5%} is 90%

$$\textbf{Prob } \{Z^*(T-E[T]) < k^*E[T]\} = P$$

$$\textbf{Prob } \{T < E[T] + k^*E[T]/Z\} = P$$

Assuming T is approximately Normally distributed, then

$$E[T] + k^*E[T]/Z = E[T] + z_p \text{Var}[T]^{1/2}$$

$$k^*E[T]/Z = z_p \text{Var}[T]^{1/2}$$

$$Z = (k^*E[T]) / (z_p \text{Var}[T]^{1/2})$$

Limited Fluctuation Formula for Z (continued)

- Assuming the insurance frequency process has a Poisson distribution, and ignoring severity:
  - Then E[T] = number of claims (N) and E[T] = Var[T], so:

$Z = (k*E[T]) / (z_p Var[T]^{1/2})$ becomes

$Z = (k*E[T]) / (z_p E[T]^{1/2})$

$Z = (k*E[T]^{1/2}) / (z_p)$

$Z = (k*N^{1/2}) / (z_p)$

Solving for N = Number of claims for full credibility (Z=1)

$N = (z_p / k)^2$

Limited Fluctuation– Standards for Full Credibility

- Claim counts required for full credibility based on the previous derivation:

Number of Claims	k			
P	2.5%	5.0%	7.5%	10%
90%	4,326	1,082	481	291
95%	6,147	1,537	683	584
99%	10,623	2,656	1,180	664

Limited Fluctuation Formula for Z – Part 2

- Generalizing to apply to pure premium:
  - T = pure premium = frequency * severity = N * S
  - E[T] = E[N]*E[S] and Var[T] = E[N]*Var[S] + $E[S]^2$*Var[N]

$Z = (k*E[T]) / (z_p Var[T]^{1/2})$

Reduces to, when solving for N = Number of claims for full credibility (Z=1)

$N = (z_p / k)^2 * \{Var[N]/E[N] + Var[S]/E[S]^2\}$

Degree of confidence multiplier	Frequency distribution: tends to be close to 1 (equals 1 for Poisson)	Severity distribution: square of coefficient of variation (can be significant)

Limited Fluctuation – Partial Credibility

- Given a full credibility standard based on a number of claims N_{full}, what is the partial credibility of data based on a number of claims N that is less than N_{full}?
- $Z = (N / N_{full})^{1/2}$
  - Square root rule
  - Designed such that the partial credibility Z will be inversely proportional to the standard deviation of the partially credible data
- Exposures vs. Claims

Limited Fluctuation – Increasing Credibility

- Under the square root rule, credibility Z can be increased by
  - Getting more data (increasing N)
  - Accepting a greater margin of error (increasing k)
  - Conceding to smaller P = being less certain (decreasing z_p)

— Based on the formula

$$Z = (N/ N_{full})^{1/2}$$
$$Z = [N/(z_p/k)^2]^{1/2}$$
$$Z = k*N^{1/2}/z_p$$

Limited Fluctuation – Complement of Credibility

- Once the partial credibility Z has been determined, the complement (1-Z) must be applied to something else – the "complement of credibility"

If the data analyzed is…	***A good complement is...***
Pure premium for a class	**Pure premium for all classes**
Loss ratio for an individual risk	**Loss ratio for entire class**
Indicated rate change for a territory	**Indicated rate change for the entire state**
Indicated rate change for entire state	**Trend in loss ratio or the indication for the country**

Limited Fluctuation – Weaknesses

- The strength of limited fluctuation credibility is its simplicity
  - Thus its general acceptance and use
  - But it has its weaknesses
- Establishing a full credibility standard requires subjective selections regarding P and k
- Typical use of the formula based on the Poisson model is inappropriate for most applications
- Partial credibility formula – the square root rule – only holds for a normal approximation of the underlying distribution of the data.   Insurance data tends to be skewed.
- Treats credibility as an intrinsic property of the data.

Limited Fluctuation – Example

- Calculate the loss ratios, given that the expected loss ratio is 75%, and using the square root rule

	Loss Ratio	Claims	Credibility at: 1,082	Credibility at: 5,410	Weighted Loss Ratio	Indicated Rate Change
2002	67%	535				
2003	77%	616				
2004	79%	634				
2005	77%	615				
2006	86%	686				
3 year	81%	1,935	100%	60%	78.6%	4.8%
5 year	77%	3,086	100%	75%	76.5%	2.0%

Example: 78.6% = 81%(0.60) + 75%(1-0.60)

Example: 1.020 = 76.5%/75% -1

Limited Fluctuation – Example 2

- Determine what the indicated factor is for a territorial factor.

Year	Territory's Exposure	Territory's Claims	Territory's Loss Ratio	Statewide Loss Ratio
2004	**3,000**	**335**	**125%**	**78%**
2005	**3,020**	**416**	**153%**	**83%**
2006	**3,030**	**634**	**269%**	**85%**
2007	**3,020**	**215**	**122%**	**79%**
2008	**3,050**	**186**	**108%**	**72%**
3 year	**9,100**	**1,035**	**196%**	**78%**
5 year	**15,120**	**1,786**	**162%**	**80%**

Current rating factor for the territory under review: 1.08

Limited Fluctuation – Example 2 (cont.)

- Assuming Poisson frequency and a coefficient of variation of 1.5 for severity, we then need to select confidence levels for the following formula:

$$N = (z_p / k)^2 * \{Var[N]/E[N] + Var[S]/E[S]^2\}$$

- If we want to be within 5% of the true value 90% of the time, the value for $(z_p / k)^2$ is 1,082. Plugging into the formula:

$$N_{claims} = 1{,}082 * \{ 1 + 1.5^2 \} = 3{,}516.5$$

- Assuming the 5-year statewide frequency is 0.2:

$$N_{exposures} = 3{,}516.5 / 0.2 = 17{,}582.5$$

Limited Fluctuation – Example 2 (cont.)

- To show the impact of our selection of an exposure standard instead of a claims standard.

Year	Exposure	Claims	Exposure Credibility	Claims Credibility
2004	3,000	335	41.3%	30.9%
2005	3,020	416	41.4%	34.4%
2006	3,030	634	41.5%	42.5%
2007	3,020	215	41.4%	24.7%
2008	3,050	186	41.6%	23.0%
3 year	9,100	1,035	71.9%	54.3%
5 year	15,120	1,786	92.7%	71.3%

Using a claims standard of 3,516.5 and an exposure standard of 17,582.5

Limited Fluctuation – Example 2 (cont.)

- Determine what the indicated factor is for a territorial factor.

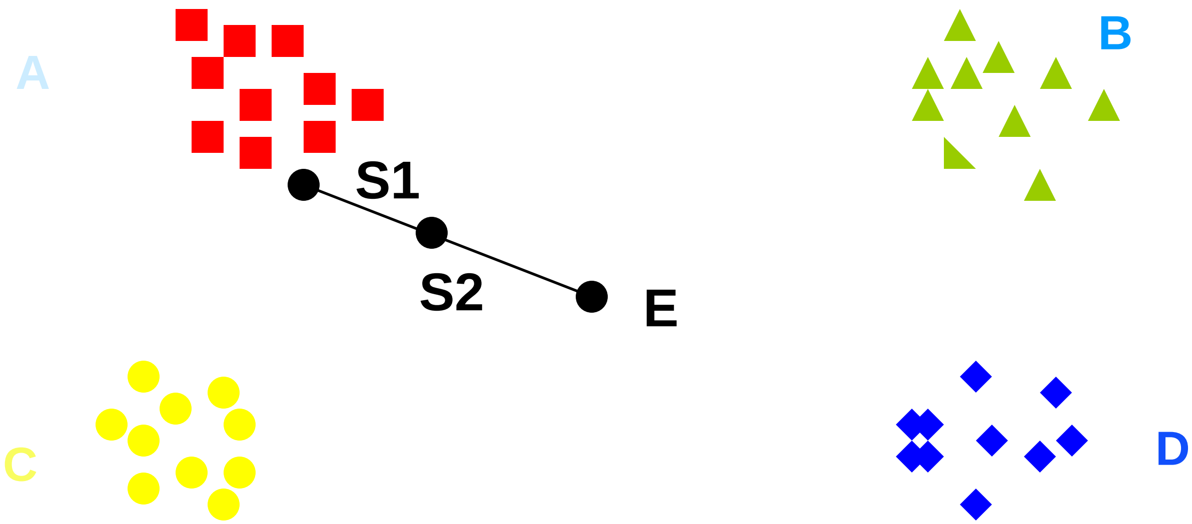
Year	Territory's Loss Ratio	Territory's Credibility	Statewide Loss Ratio	Cred. Weighted Loss Ratio
3 year	196%	71.9%	78%	162.8%
5 year	162%	92.7%	80%	156.0%

The final indicated territorial factor is (156% / 80%)*0.85 + 0.15 = 1.8075

An alternative approach would be to calculate the indicated factor prior to applying credibility, and then credibility weight the current factor with the indicated factor.

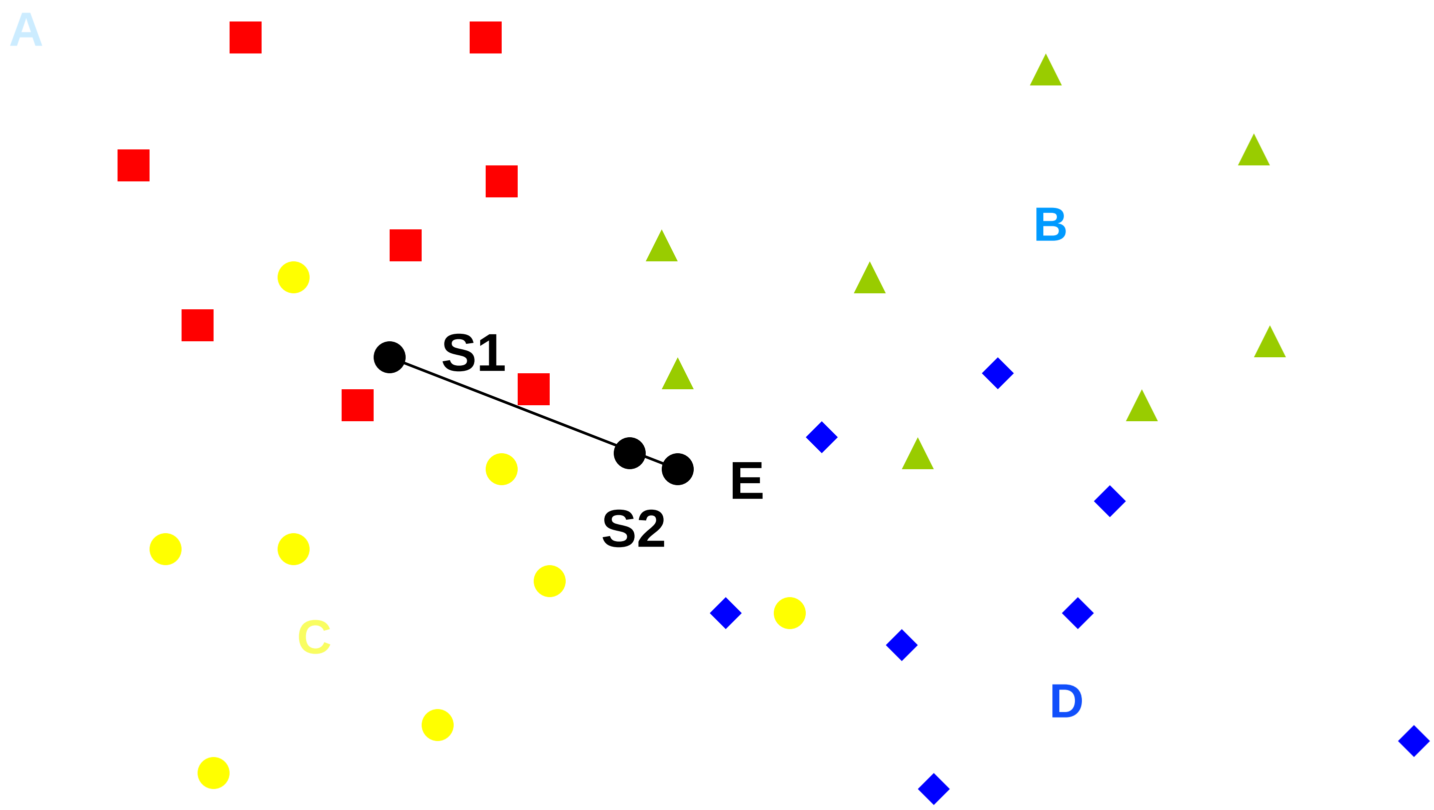
Greatest Accuracy Credibility Illustration

Steve Philbrick's target shooting example...



Greatest Accuracy Credibility Illustration (continued)

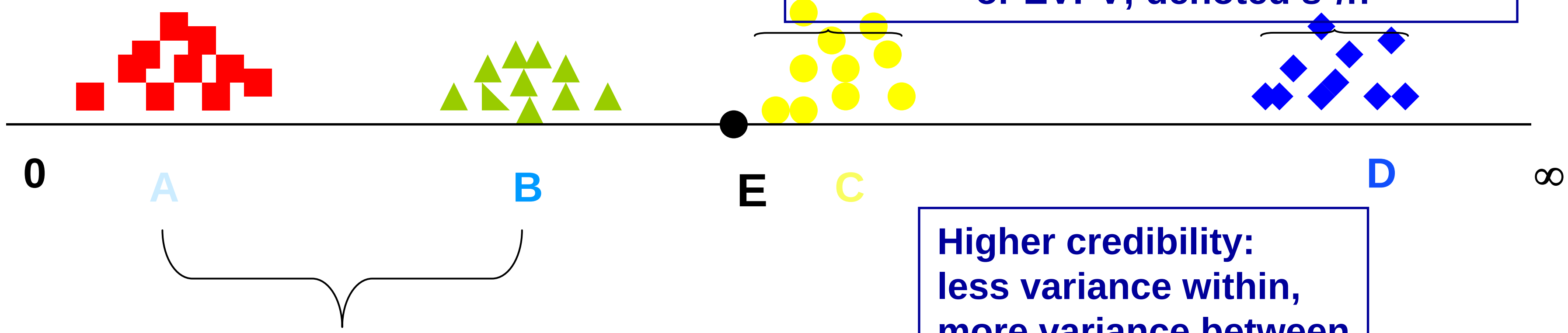
Which data exhibits more credibility?



Greatest Accuracy Credibility Illustration (continued)

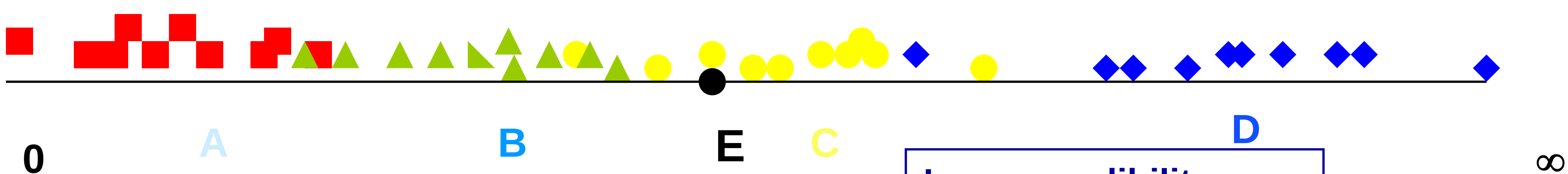
Class loss costs per exposure...

Average "within" class variance =
"Expected Value of Process Variance"
= or EVPV; denoted s^2/n



Higher credibility:
less variance within,
more variance between

Variance between the means =
"Variance of Hypothetical Means"
or VHM; denoted t^2



Lower credibility:
more variance within,
less variance between

Greatest Accuracy Credibility Derivation
(with thanks to Gary Venter)

- Suppose you have two independent estimates of a quantity, **x** and **y**, with squared errors of **u** and **v** respectively
- We wish to weight the two estimates together as our estimator of the quantity:

$$a = z*x + (1-z)*y$$

- The squared error of a is

$$w = z^2 u + (1-z)^2 v$$

- Find Z that minimizes the squared error of **a** – take the derivative of **w** with respect to **z**, set it equal to 0, and solve for **z**:
  - $dw/dz = 2zu + 2(z-1)v = 0$, so
  - $Z = u/(u+v)$

Greatest Accuracy Credibility Derivation
(with thanks to Gary Venter) (Continued)

- Using the formula that establishes that the least squares value for Z is proportional to the reciprocal of expected squared errors:

$$Z = (n/s^2)/(n/s^2 + 1/t^2)$$
$$Z = n/(n + s^2/t^2)$$
$$Z = n/(n+k)$$

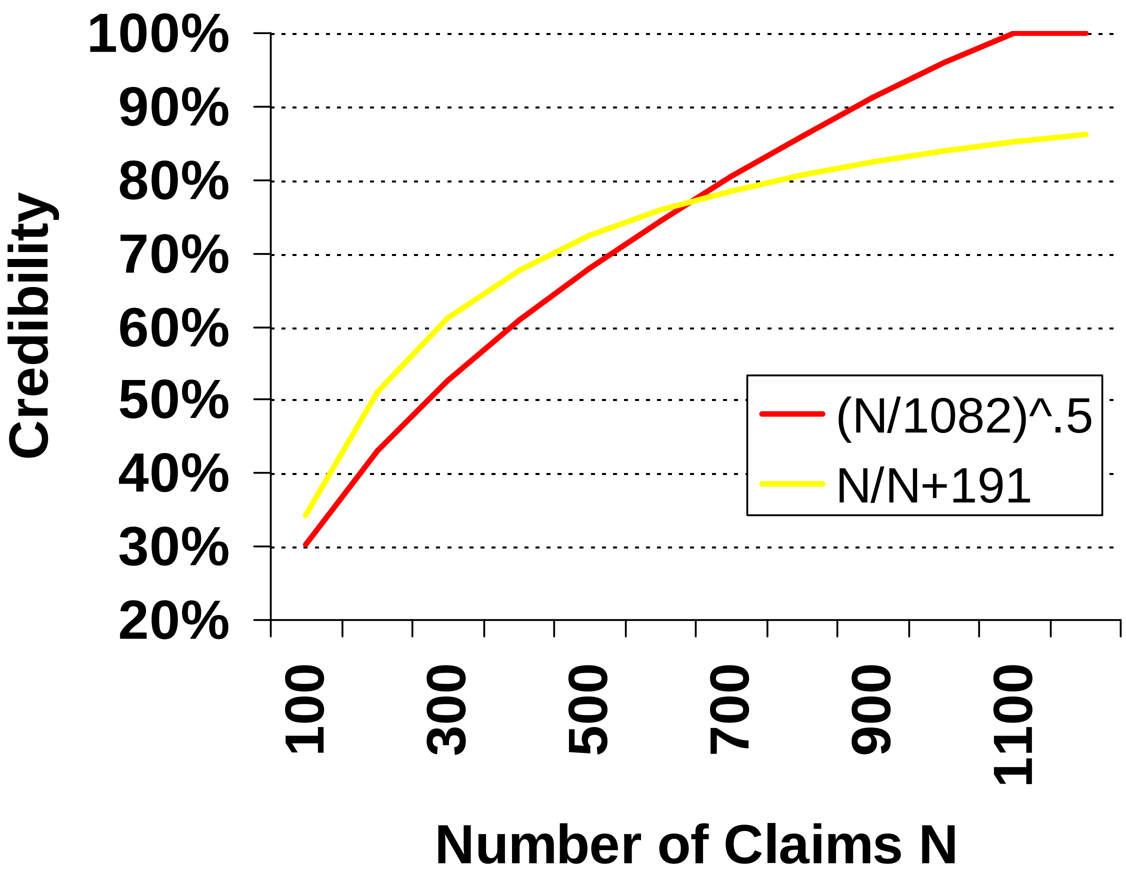
- Credibility Z can be increased by:
  - Getting more data (increasing n)
  - Getting less variance within classes (e.g., refining data categories) (decreasing s^2)
  - Getting more variance between classes (increasing t^2)

Greatest Accuracy Credibility – Strengths and Weaknesses

- The greatest accuracy or least squares credibility result is more intuitively appealing.
  - It is a relative concept
  - It is based on relative variances or volatility of the data
  - There is no such thing as full credibility

- Issues
  - Greatest accuracy credibility can be more difficult to apply. Practitioner needs to be able to identify variances.
  - The Credibility Parameter K, is a property of the *entire* set of data. So, for example, if a data set has a small, volatile class and a large, stable class, the credibility parameter of the two classes would be the same.

Comparison of Limited Fluctuation and Greatest Accuracy



Greatest Accuracy Credibility Example

Business Problem

- Personal Automobile Bodily Injury loss emergence patterns are influenced by both the limit profile of the book as well as the state profile of the book.
- The data is not large enough to split it both ways (i.e. by State and Limit)
- Which split is more credible thus providing more reliable estimates of ultimate loss?

Solution

- Test emergence patterns by State Group (Four State Groups)
  - No-Fault versus Tort Law
  - High versus Low Liability environment
- Test emergence patterns by Limit Group (Three Limit Groups)
  - Low
  - Medium
  - High
- Calculate the value of $\mathbf{K = s^2/t^2}$ both ways and compare resulting credibility estimates

Greatest Accuracy Credibility Example
Data Sample by State Group over all Limit Groups

State Group=1, Limit Group=All	qtr_1	qtr_2	qtr_3	qtr_4	qtr_5
PQ 15 LDF to ULT	3.8820	1.2822	1.7165	1.3517	1.0864
PQ 14 LDF to ULT	2.3053	1.9036	1.2038	1.4716	1.2357
PQ 13 LDF to ULT	2.1553	1.6172	1.6991	1.1884	1.4664
PQ 12 LDF to ULT	2.1045	1.5362	1.2893	1.5848	1.1546
. . .					
PQ 4 LDF to ULT	2.1814	1.4757	1.2541	1.3063	1.1901
PQ 3 LDF to ULT	2.0628	1.1631	1.3862	1.2218	1.1730
PQ 2 LDF to ULT	1.6577	1.2594	1.1577	1.3171	1.0902
Prior Quarter 1 LDF to ULT	3.1211	1.4010	1.2681	1.1772	1.2987
Current Quarter LDF to ULT	1.9756	1.5606	1.3548	1.3559	1.0443

Greatest Accuracy Credibility Example
Calculations for each subset of data

- We have divided the data two ways and calculated the implied Loss Development Factors to Ultimate (LDF to ULT) for each subset of the data:
  - Four State Groups across all Limit Groups
  - Three Limit Groups across all State Groups
- Calculate for each subset of the data (seven ways):
  - Variance of the LDF to ULT for each evaluation period (interim calculation for EVPV)
  - Mean LDF to ULT (interim calculation for VHM)
  - Square Error between the Mean for the subset and the overall mean for each grouping (a.k.a. state groups or limit groups) (interim calculation for VHM)

State Group=1, Limit Group=All	**qtr_1**	**qtr_2**	**qtr_3**	**qtr_4**	**qtr_5**
LDF to ULT Variance (within a State Group)	0.3133	0.0563	0.0400	0.0257	0.0136
Loss Weighted Average LDF to ULT (Mean of a State Group)	2.2099	1.4535	1.3566	1.2812	1.2081
LDF to ULT Square Error across State Groups	0.0547	0.0225	0.0120	0.0066	0.0030

Greatest Accuracy Credibility Example
Credibility Calculation for each grouping of data

Calculations for each view of the data (State Groups versus Limit Groups)

- **EVPV** = *Within Variance* = Mean of the variance for each subset
- Mean LDF to ULT= Mean across all subsets (used to calculate squared error)
- **VHM** = *Between Variance* = Weighted average of the squared error term for each subset
- Use reported incurred losses as weights for each average

State Group Summary	**qtr_1**	**qtr_2**	**qtr_3**	**qtr_4**	**qtr_5**
Expected Value of Process Variance (EVPV)	**0.3813**	0.0418	0.0264	0.0181	0.0100
Mean across Groups (used to calculate squared errors)	2.4437	1.6035	1.4661	1.3622	1.2629
Variance of Hypothetical Means (VHM)	**0.6741**	0.1292	0.0650	0.0355	0.0201
K = EVPV / VHM	0.57	0.32	0.41	0.51	0.50
Bühlmann Credibilty Estimate Y / (Y+K)	**97%**	**98%**	**98%**	**97%**	**97%**
Y	16	16	16	16	16

Greatest Accuracy Credibility Example
Credibility Calculation for each grouping of data

Limit Group Summary	**qtr_1**	**qtr_2**	**qtr_3**	**qtr_4**	**qtr_5**
Expected Value of Process Variance (EVPV)	**0.1888**	0.0233	0.0170	0.0164	0.0106
Mean across Groups (used to calculate squared errors)	2.5213	1.6514	1.5050	1.3918	1.2871
Variance of Hypothetical Means (VHM)	**0.0490**	0.0145	0.0092	0.0078	0.0032
K = EVPV / VHM	3.85	1.61	1.84	2.10	3.32
Bühlmann Credibilty Estimate Y / (Y+K)	**81%**	**91%**	**90%**	**88%**	**83%**
Y	16	16	16	16	16

Conclusion

- State Groups have higher credibility than Limit Groups
- K Limit Groups > K State Groups
- Variance between State Groups was much higher than the Variance between Limit Groups

Credibility – Bibliography

- Herzog, Thomas. Introduction to Credibility Theory.
- Longley-Cook, L.H. "An Introduction to Credibility Theory," PCAS, 1962
- Mayerson, Jones, and Bowers. "On the Credibility of the Pure Premium," PCAS, LV
- Philbrick, Steve. "An Examination of Credibility Concepts," PCAS, 1981
- Venter, Gary and Charles Hewitt. "Chapter 7: Credibility," Foundations of Casualty Actuarial Science.
- ___________. "Credibility Theory for Dummies," CAS Forum, Winter 2003, p. 621
- Mahler, H.C. and Dean, C.G., "Credibility," Foundations of Casualty Actuarial Science (Fourth Edition), 2001, Casualty Actuarial Society, Chapter 8.
- Dean, C.G., "Topics in Credibility Theory," 2004 (SOA Study Note)