ACTUARIAL RESEARCH CLEARING HOUSE
1991 VOL. 1

BAYESIAN GRADUATION

Steve Craighead
Nationwide Life Insurance Company
Columbus, Ohio

ABSTRACT: This paper is concerned with various modifications of the Kimeldorf-Jones Bayesian graduation method. These enhancements entail the use of time series and eigenvalue decomposition. By augmenting the existing method one can empirically smooth the Kimeldorf-Jones technique.

In Bayesian statistics, one begins with a prior assumption which is quantified by choosing a probability distribution that best describes that assumption. This distribution will be called the prior distribution with random variable T. Denote this distribution as $f_T(t)$. Next, one conducts various experiments, makes observations, and chooses a conditional distribution $f_{U|T}(u|t)$ where U is the random variable of the observations. Note that the parameter U is related to the prior opinion's variable T. Once $f_{U|T}(u|t)$ has been determined, the distribution

$f_U(u)$ can be obtained from summing or integrating $f_{U|T}(u|t)f_T(t)$ over the range of t.

The classic Bayes formula

$$f_{T|U}(t|u) = \frac{f_{U|T}(u|t)f_T(t)}{f_U(u)} \qquad (1)$$

produces the posterior conditional distribution. Note: Most prior distributions are chosen such that when they are linked with the conditional distribution $f_{U|T}(u|t)$, the posterior distribution has the same form of distribution as the prior distribution. These distributions are called conjugate distributions.

These techniques can be related to graduation by assuming a collection $\{T_i\}$ are 'true rates' of mortality and $\{U_i\}$ are observed rates from a mortality study. If one produces

$$f_{T_1,T_2,\ldots,T_n}(t_1,t_2,\ldots,t_n),$$

$$f_{U_1,U_2,\ldots,U_n|\,T_1,T_2,\ldots,T_n}(u_1,u_2,\ldots,u_n|t_1,t_2,\ldots,t_n),$$

and

$$f_{U_1,U_2,\ldots,U_n}(u_1,u_2,\ldots,u_n)$$

then one can find the posterior distribution

$$f_{T_1,T_2,\ldots,T_n|\,U_1,U_2,\ldots,U_n}(t_1,t_2,\ldots,t_n|u_1,u_2,\ldots,u_n)$$

So one's concept of 'true rates' would be modified by the observations. For further insight see [11].

This paper is concerned with the Kimeldorf-Jones model [8], which from this point will be referred to as the K-J model, in which the prior distribution of the rates of mortality is selected to be the multinormal distribution. Also the observed conditional distribution is selected as multinormal, which allows the conjugate posterior distribution to be multinormal. One can refer to [8] or [9] for a brief review of the use of the multinormal distribution.

In the K-J model, one sets up two covariance matrices, **A** and **B**. **A** is the covariance matrix that is associated with the prior distribution, and B is the covariance matrix associated with the observations. In [8], $\mathbf{A} = \{a_{ij}\}$ is defined to be <u>admissible</u> if the following four properties hold:

(i) $i \le j \le k$ or $i \ge j \ge k$ implies $\dfrac{a_{ij}}{a_{jj}^{1/2}} \ge \dfrac{a_{ik}}{a_{kk}^{1/2}}$. (2)

(ii) $a_{ij} \ge 0$ for all i,j. (3)

(iii) **A** is a symmetric matrix, i.e. $\mathbf{A}' = \mathbf{A}$. (4)

(iv) **A** is positive definite. That is all eigenvalues are positive. (5)

The **B** matrix associated with the observations must be admissible as well. In Kimeldorf's and Jones' orginal paper [8] they made the assumption that the random variables of the observed $\{U_x\}$ are independent, which causes **B** to be a diagonal matrix. The x'th element in this diagonal matrix was of the form $E_x/(m_x(1-m_x))$, where E_x is the exposure at age x and m_x is the mortality rate associated with the prior distribution.

Now in the K-J model, one represents the prior distribution mortality rates $\{m_x\}$ as a vector $\mathbf{m}$ and the observed mortality rates $\{u_x\}$ as the vector $\mathbf{u}$. The graduated rates $\{v_x\}$, denoted $\mathbf{v}$

are then obtained from

$$\mathbf{v} = \mathbf{u} + (\mathbf{I} + \mathbf{A}\mathbf{B}^{-1})^{-1}(\mathbf{m} - \mathbf{u}). \tag{6}$$

This can be adjusted to the following:

$$\begin{aligned} \mathbf{v} &= \mathbf{m} - \mathbf{m} + \mathbf{u} + \mathbf{C}(\mathbf{m} - \mathbf{u}) \\ &= \mathbf{m} - (\mathbf{m} - \mathbf{u}) + \mathbf{C}(\mathbf{m} - \mathbf{u}) \\ &= \mathbf{m} + (\mathbf{C} - \mathbf{I})(\mathbf{m} - \mathbf{u}) \end{aligned} \tag{7}$$

where $\mathbf{C} = (\mathbf{I} + \mathbf{A}\mathbf{B}^{-1})^{-1}$.

The K-J method has been criticized in several ways, one of which is its subjectivity in how the matrix **A** is constructed. Another criticism is that the method is too sensitive to outliers in the observed data. This causes the K-J method to tend not to smooth out the data. This paper will address the following enhancements to correct these problems:

(i) Use the concept of autocorrelation from time series analysis to reduce the selection of the elements of the matrices **A** and **B**.

(ii) Revise the basic K-J model by using eigenvalue-eigenvector methods to eliminate the effect of the outliers.

TIME SERIES ANALYSIS

To modify the creation of the matrix **A** (or **B**) from the K-J method use the formula

$$r_k = \frac{\sum_{t=x}^{n-k} (m_t - \bar{m})(m_{t+k} - \bar{m})}{\sum_{t=x}^{n} (m_t - \bar{m})^2} \qquad (8)$$

to calculate the auto-correlation coefficents, where $\bar{m}$ is the average of all the mortality rates, and n is the total number of rates. See [12]. One can accelerate these calculations by the use of finite Fourier transforms [1]. Now the $\{r_k\}$ must be adjusted to meet the admissibility requirements. The first adjustment is to convert all negative $\{r_k\}$ to zero. This will satisfy the condition in equation (3). If any $r_k > r_{k-1}$, then let $r_k = r_{k-1}$.

In the construction of the matrix **A** (or B), let D be the diagonal matrix $\mathrm{diag}(d_1, d_2, \ldots, d_n)$ representing the variances at each age for the **A** (or B) matrix. Define the upper triangular matrix **K** by

$$k_{ij} = \begin{cases} \sqrt{d_i}\sqrt{d_j}\, r_{j-i} & \text{if } j > i \\ 0 & j \le i \end{cases} \qquad (9)$$

Define **A** = **K'** + **D** + **K**. **A** is a symmetric matrix, and one must test equation (2) for admissibility requirements. If the variances $\{d_i\}$ are constant, then **A** is a symmetric Toeplitz matrix [3] and will satisfy admissibility equation (2). This type of matrix **A** is very close to the Class 2 matrix in [8]; however, this structure does not guarantee that **A** is postive definite. To use matrix **A** or **B**, one must determine the eigenvalues of **A** or **B**, and verify that the eigenvalues are all positive. If this is not the case, one can revise the $\{r_k\}$ further by letting $r_j = 0$ for all $j > m$ for some m and then redetermine the covariance matrix. Further research may be done here to determine further restrictions on **A** (or B) to guarantee that the symmetric Toeplitz matrices are positive definite.

ITERATIVE GRADUATION

T. N. E. Greville in [7] made reference to a repeated application of a graduation process to smooth data. He proceeded to use various eigenvalue techniques to theoretically obtain a modification of the Whitaker-Henderson method, which he later claimed contained the K-J method as a subcase. See the discussion in [8]. In this paper, practical eigenvalue and eigenvector algorithms will be used to create an empirical method that allows one to control the effect of outliers on the K-J method.

If one takes the results of equation (7) and substitutes the values of v in place of u, one obtains the following formula

$$v^{(2)} = \mathbf{m} - (C-I)^2(\mathbf{m} - u) \qquad (10)$$

and in general, the r'th iteration has the formula

$$v^{(r)} = \mathbf{m} + (-1)^{r-1}(C-I)^r(\mathbf{m} - u) \qquad (11)$$

If the limit of equation (11) exists as r tends to infinity, the behavior of smoothing, as referred to in [7], holds, and the effect of outliers on the graduated data is eliminated. However, it turns out that the infinite limit of (11) goes to **m**. This is because of the following argument supplied by Dr. James Daniel:

Let **A** and **B** be positive definite symmetric matrices, and suppose $C=(I+AB^{-1})^{-1}$. Let d be an eigenvalue of C-I. Then 1+d is an eigenvalue of C, and 1/(1+d) is an eigenvalue of $C^{-1} = I + AB^{-1}$, and 1/(1+d) - 1 is an eigenvalue of AB^{-1}. This reasoning is correct in both directions, hence d being a eigenvalue of C - I is equivalent to $\mu = 1/(1+d) - 1$ being an eigenvalue of AB^{-1}.

Now what is the form of the eigenvalues of AB^{-1}? Since **B** is positive definite and symmetric, one can write $B=R^TR$ for some real non-singular matrix **R**. Then $AB^{-1}x = \mu x$ or $AR^{-1}R^{-1T}x = \mu x$, so by multiplying by R^{-1T}, $R^{-1T}AR^{-1}(R^{-1T}x) = \mu(R^{-1T}x)$ and $R^{-1T}AR^{-1}y = \mu y$ with $y = R^{-1T}x$. That is, μ is an eigenvalue of the symmetric positive definite matrix $A' = R^{-1T}AR^{-1}$. This is positive definite since $x^TA'x = (R^{-1}x)^TA(R^{-1}x) > 0$ if $x \neq 0$. So all $\mu > 0$. Thus

$$\frac{1}{1+d} - 1 > 0$$

and finally

$$-1 < d < 0.$$

So C - I cannot have negative one as an eigenvalue.

So to find the limit of equation (11) as r tends to infinity, the C - I will converge to the zero matrix. So the graduated rates will converge to the prior table. However, intermediate results will still supply a viable graduation technique. First some preliminary definitions are necessary.

Let $\mathbf{w} = (w_i)$ be a vector. Define the euclidean norm of $\mathbf{w}$ to be

$$\| \mathbf{w} \| = \left(\sum_{i=1}^{n} |w_i|^2 \right)^{1/2}$$

and define the 2-norm of a n x n matrix $\mathbf{F}$ to be

$$\| \mathbf{F} \| = \sup_{\mathbf{x} \neq 0} \frac{\| \mathbf{F}\mathbf{x} \|}{\| \mathbf{x} \|}$$

where $\mathbf{x}$ is a n x 1 vector. The same double-bar notation is used for the matrix norm as for the vector norm. The context will determine whether the notation is for the vector or matrix form. Define the condition number of a n x n matrix $\mathbf{F}$ to be

$$c(\mathbf{F}) = \| \mathbf{F} \| \, \| \mathbf{F}^{-1} \|$$

In [4] matrices with small condition numbers are said to be well-conditioned (less thatn 10^6). If $c(\mathbf{F})$ is large, then $\mathbf{F}$ is said to be ill-conditioned.

To determine the eigenvalues and eigenvectors of C - I, the practical QR or QZ algorithm can be used. See [5]. These algorithms are available in EISPACK (see [15] and [2]), or MATLAB (see [13]). When these algorithms are used, a eigenvalue matrix D and a eigenvector matrix $\mathbf{X}$ are generated, If the $c(\mathbf{X})$ is well-conditioned, one can assume that the eigenvectors are independent. See [5].

Let the eigenvalue and eigenvector matrices of C - I be designated as **D** and **X** respectively.

So $C - I = XDX^{-1}$. Now

$$(C - I)^r = XD^rX^{-1}$$

where $D^r = \mathrm{diag}(d_1^r, d_2^r, \ldots, d_n^r)$. In fact,

$$(-1)^{r-1}(C - I)^r = \mathrm{diag}((-1)^{r-1} d_1^r, (-1)^{r-1} d_2^r, \ldots, (-1)^{r-1} d_n^r)$$

Hence, a useful approximation to D^r is to create a matrix D' which is determined by setting the eigenvalues very near negative one, to exactly negative one, and all others to zero. This allows one to selectively decide the level of smoothing to the prior table. Once this matrix D' is obtained, create a new matrix $C' = XD'X^{-1}$. Instead of using equation (11) for D^r, use:

$$V = m + C'(m - u) \qquad (12)$$

This will be demonstrated in the example below.

EXAMPLE

The data presented in Table 1 is an approximation to the data underlying the 1971 Group Annuity Mortality Table, Female Lives. See [6] and [10]. The vector **m** of prior means was specified as the mortality rates of the 1951 Group Annuity Table, Female lives from [14]. Using formula (8) one obtains the autocorrelation coefficents, $\{r_k\}$, as displayed in Table 2. The $\{r_k\}$ have been adjusted, by setting all negative values to zero.

The variances for the prior covariance matrix **A** are all set to the constant one. The off diagonal elements are created as discussed above. When the eigenvalues for **A** are calculated, they are all positive, and so in this model, **A** is positive definite. To simplify the process, we assume that the observations in **B** are independent; therefore **B** will be a diagonal matrix with variances determined by the formula $E_x/(m_x(1-m_x))$. Applying equation (6) one obtains the graduated values $\{v_x\}$. These values are displayed in Figures 1 and 2, and in Table 4. The numerical fit measure F is determined from the formula

$$F = \sum_x \frac{n_x(v_x - u_x)^2}{v_x(1-v_x)}$$

where the smoothness measure S is determined by

$$S = \sum_x (\Delta^3 v_x)^2$$

One may note from these measures that there is very little deviation from the observed $\{u_x\}$.

In the determination of the $\{v'_x\}$, the matrix **C** - **I** from equation (7) is obtained from the above matrices **A** and **B**. Then the matrix **C** - **I** was factored into a eigenvalue matrix **D** and an eigenvector matrix **X**. The condition number on the eigenvector matrix **X** was 17.2; therefore the matrix is well conditioned. Listed in Table 3 is the set of the eigenvalues obtained by the QR algorithm. Due to roundoff error, one obtains eigenvalues equaling negative one. Next, a diagonal matrix **D'** was created with zeros for the first sixteen diagonal entries and negative one for the remaining entries. Using **D'** and **X**, let $\mathbf{C'} = \mathbf{XD'X}^{-1}$. Using the equation (12) one finally obtains the $\{v'_x\}$, which is displayed in Figures 3 and 4 and in Table 4.

Now create a diagonal matrix **D''** with zeros for the first twelve entries and negative one for the remaining entries. In Figures 5 and 6 and v''_x in Table 4, one can observe the effect of the matrix **D''** in graduating the data. In this method, by the use of autocorrelation coefficents and constant variances for the prior distribution, the only subjectivity arises in the decision of whether to set the eigenvalues to zero or negative one.

In summary, these enhancements remove a great deal of subjectivity from the selection of the matrices **A** and B. Also, they allow the actuary to control the effect of sensitivity to low exposure. However, it would be desirable to be able to work directly with equation (7) in determining a weighted average between an established table **m** and the observed data **u** without recourse to the Bayesian techniques of the K-J method. However, the conditions on **C**, which will guarantee this modification, require further research and study.

APPROXIMATION TO 1971 GAM FEMALE LIVES - TABLE 1

x	θ_x	E_x	u_x	x	θ_x	E_x	u_x
55	1	84	.01190	80	374	6140	.06091
56	2	418	.00478	81	348	4718	.07376
57	10	1066	.00938	82	304	3791	.08019
58	21	2483	.00846	83	249	2806	.08874
59	35	3721	.00941	84	167	2240	.07455
60	62	5460	.01136	85	192	1715	.11195
61	50	6231	.00802	86	171	1388	.12320
62	55	8061	.00682	87	126	898	.14031
63	88	9487	.00928	88	86	578	.14879
64	132	10770	.01226	89	97	510	.19020
65	267	24267	.01100	90	93	430	.21628
66	300	26791	.01120	91	75	362	.20718
67	432	29174	.01481	92	84	291	.28866
68	491	28476	.01724	93	31	232	.13362
69	422	25840	.01633	94	75	196	.38265
70	475	23916	.01986	95	29	147	.19728
71	413	21412	.01929	96	25	100	.25000
72	480	20116	.02386	97	20	161	.12422
73	537	18876	.02845	98	5	11	.45455
74	566	17461	.03242	99	3	10	.30000
75	581	15012	.03870	100	2	8	.25000
76	464	11871	.03909	101	0	5	.00000
77	461	10002	.04609	102	2	4	.50000
78	433	8949	.04839	103	0	2	.00000
79	515	7751	.06644	104	1	2	.50000

AUTOCORRELATION COEFFICENTS - TABLE 2

1	0.9089
2	0.8589
3	0.8033
4	0.7428
5	0.6780
6	0.6095
7	0.5497
8	0.4973
9	0.4511
10	0.4101
11	0.3733
12	0.3400
13	0.3097
14	0.2818
15	0.2563
16	0.2327
17	0.2109
18	0.1908
19	0.1721
20	0.1548
21	0.1385
22	0.1233
23	0.1091
24	0.0956
25	0.0829
26	0.0708
27	0.0593
28	0.0484
29	0.0380
30	0.0281
31	0.0186
32	0.0095
33	0.0008

FIGURE 1

MORTALITY RATES

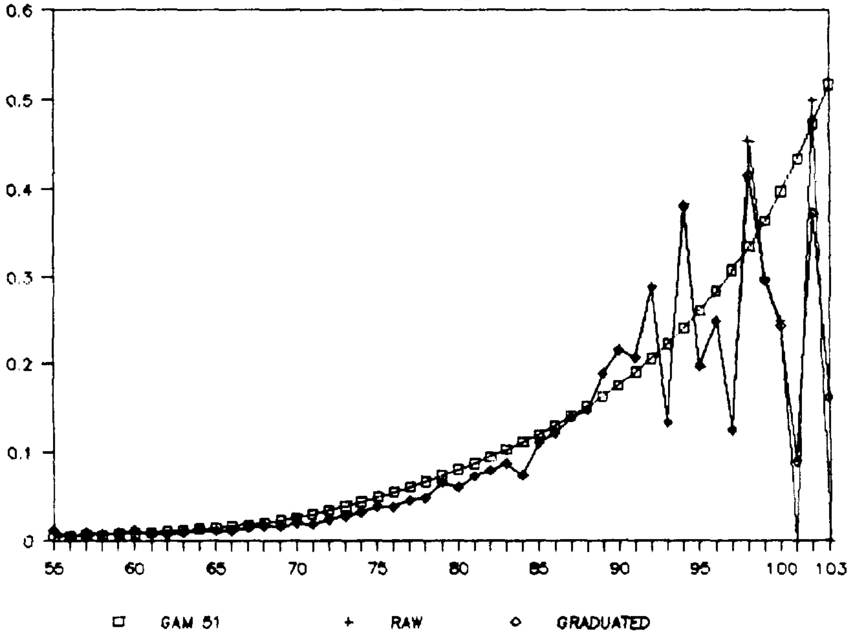
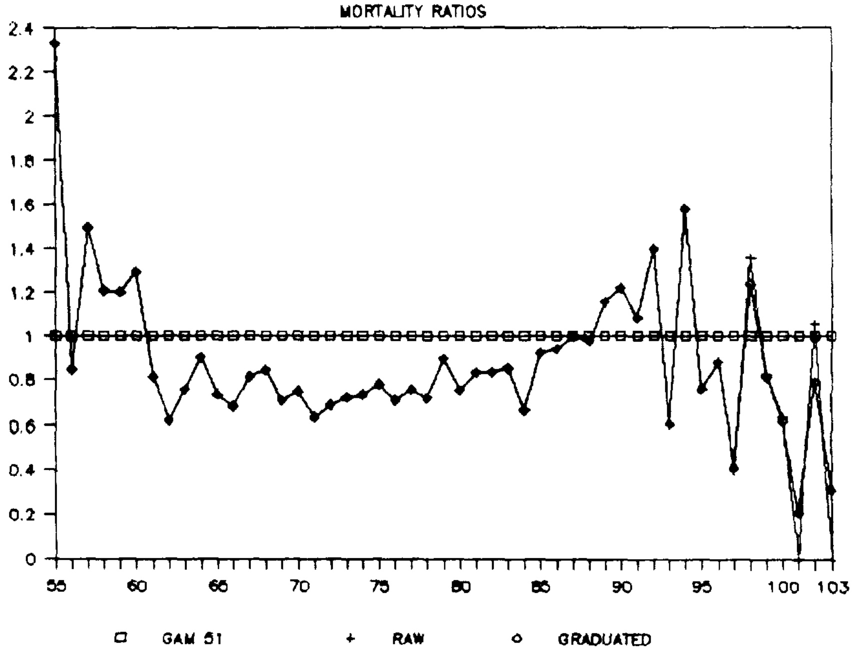


FIGURE 2

MORTALITY RATIOS



EIGENVALUES - TABLE 3

1	-0.483
2	-0.615
3	-0.733
4	-0.788
5	-0.865
6	-0.970
7	-0.976
8	-0.986
9	-0.990
10	-0.993
11	-0.995
12	-0.996
13	-0.997
14	-0.998
15	-0.999
16	-0.999
17	-1.000
18	-1.000
19	-1.000
20	-1.000
21	-1.000
22	-1.000
23	-1.000
24	-1.000
25	-1.000
26	-1.000
27	-1.000
28	-1.000
29	-1.000
30	-1.000
31	-1.000
32	-1.000
33	-1.000
34	-1.000
35	-1.000
36	-1.000
37	-1.000
38	-1.000
39	-1.000
40	-1.000
41	-1.000
42	-1.000
43	-1.000
44	-1.000
45	-1.000
46	-1.000
47	-1.000
48	-1.000
49	-1.000

FIGURE 3

MORTALITY RATES

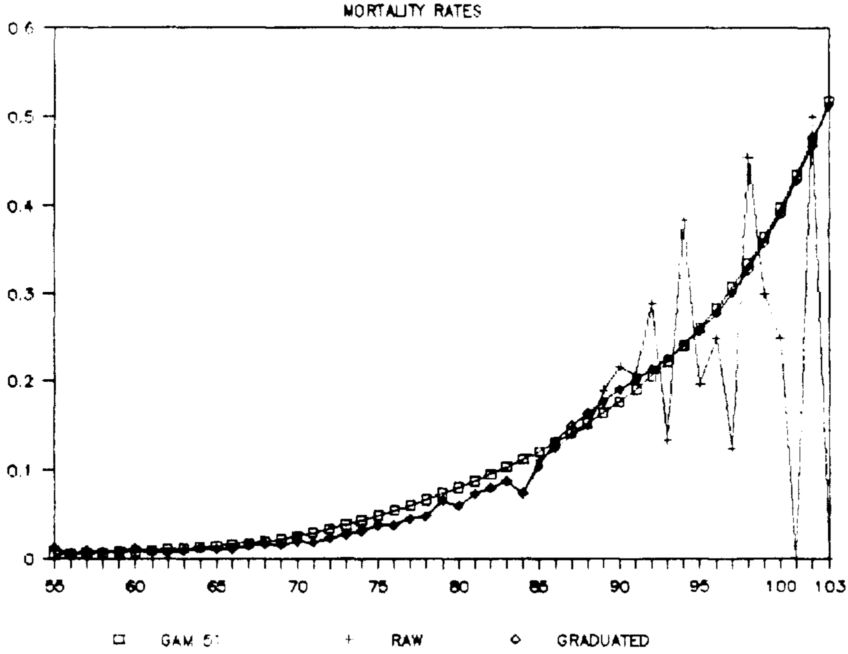


FIGURE 4

MORTALITY RATIOS

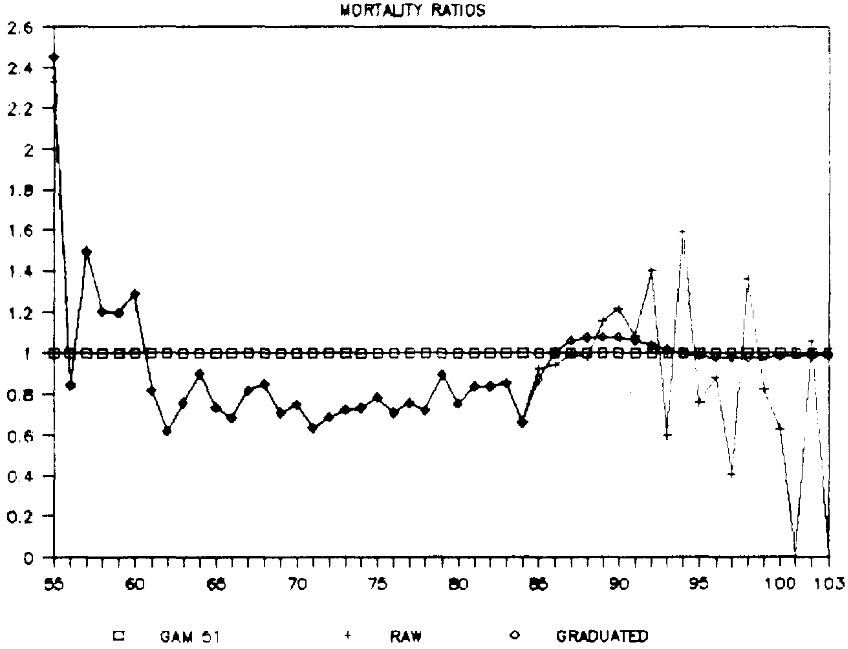


FIGURE 5

MORTALITY RATES

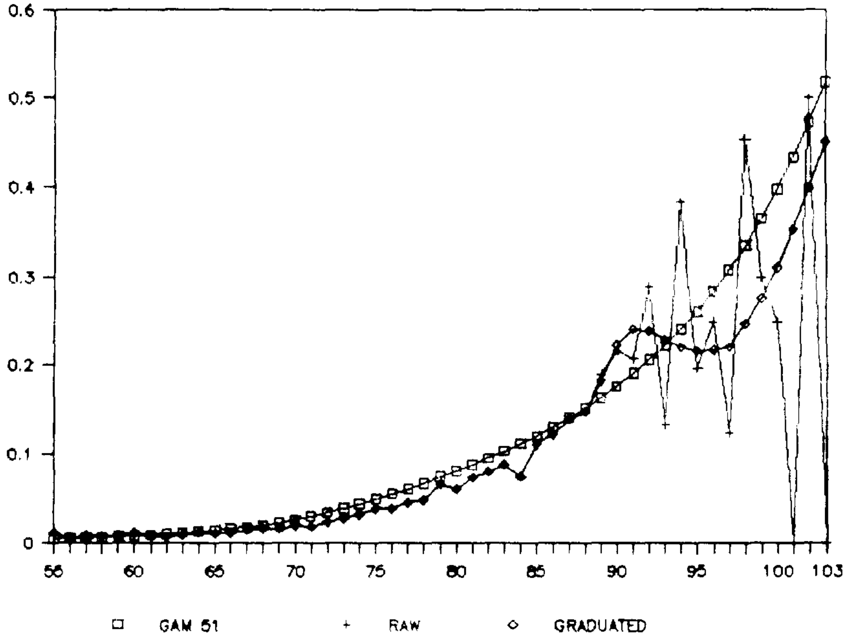
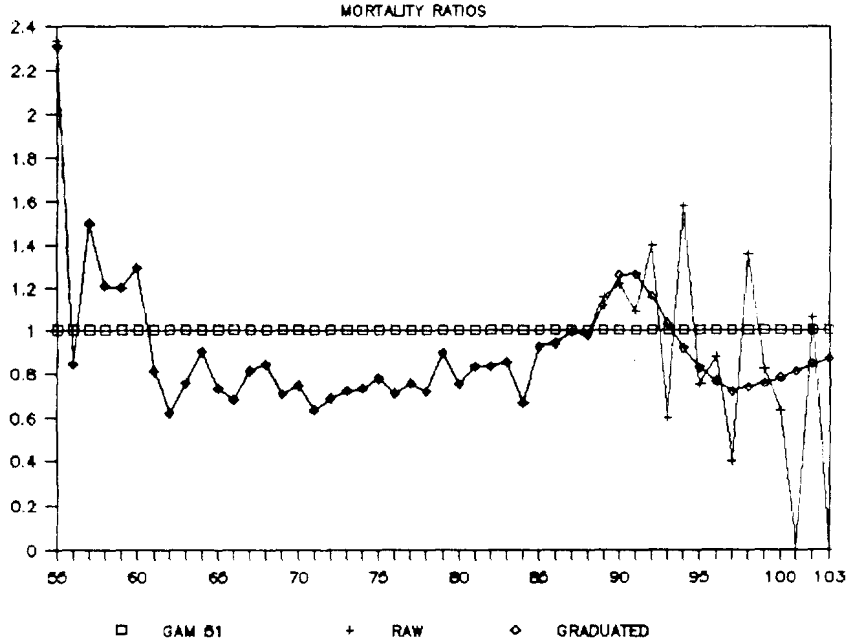


FIGURE 6

MORTALITY RATIOS



GRADUATION RESULTS - TABLE 4

	m_x	u_x	v_x	v'_x	v''_x
55	0.0051	0.0119	0.0119	0.0125	0.0118
56	0.0056	0.0048	0.0048	0.0047	0.0048
57	0.0062	0.0094	0.0094	0.0094	0.0094
58	0.0069	0.0085	0.0085	0.0084	0.0085
59	0.0078	0.0094	0.0094	0.0094	0.0094
60	0.0087	0.0114	0.0114	0.0114	0.0114
61	0.0098	0.0080	0.0080	0.0080	0.0080
62	0.0110	0.0068	0.0068	0.0068	0.0068
63	0.0122	0.0093	0.0093	0.0093	0.0093
64	0.0136	0.0123	0.0123	0.0123	0.0123
65	0.0149	0.0110	0.0110	0.0110	0.0110
66	0.0164	0.0112	0.0112	0.0112	0.0112
67	0.0182	0.0148	0.0148	0.0148	0.0148
68	0.0203	0.0172	0.0172	0.0172	0.0172
69	0.0231	0.0163	0.0163	0.0163	0.0163
70	0.0265	0.0199	0.0199	0.0199	0.0199
71	0.0304	0.0193	0.0193	0.0193	0.0193
72	0.0347	0.0239	0.0239	0.0239	0.0239
73	0.0394	0.0285	0.0285	0.0285	0.0285
74	0.0443	0.0324	0.0324	0.0324	0.0324
75	0.0495	0.0387	0.0387	0.0387	0.0387
76	0.0551	0.0391	0.0391	0.0391	0.0391
77	0.0610	0.0461	0.0461	0.0461	0.0461
78	0.0674	0.0484	0.0484	0.0484	0.0484
79	0.0741	0.0664	0.0664	0.0665	0.0664
80	0.0811	0.0609	0.0609	0.0611	0.0609
81	0.0883	0.0738	0.0738	0.0738	0.0738
82	0.0959	0.0802	0.0802	0.0801	0.0802
83	0.1039	0.0887	0.0887	0.0883	0.0888
84	0.1123	0.0746	0.0746	0.0740	0.0747
85	0.1213	0.1120	0.1119	0.1056	0.1120
86	0.1309	0.1232	0.1232	0.1320	0.1227
87	0.1412	0.1403	0.1403	0.1496	0.1399
88	0.1523	0.1488	0.1488	0.1637	0.1482
89	0.1643	0.1902	0.1902	0.1771	0.1823
90	0.1771	0.2163	0.2162	0.1907	0.2234
91	0.1911	0.2072	0.2073	0.2029	0.2418
92	0.2063	0.2887	0.2881	0.2142	0.2394
93	0.2230	0.1336	0.1348	0.2264	0.2299
94	0.2413	0.3827	0.3810	0.2408	0.2216
95	0.2615	0.1973	0.1984	0.2579	0.2168
96	0.2836	0.2500	0.2493	0.2779	0.2173
97	0.3080	0.1242	0.1266	0.3008	0.2215
98	0.3348	0.4546	0.4152	0.3277	0.2477
99	0.3644	0.3000	0.2963	0.3575	0.2759
100	0.3971	0.2500	0.2439	0.3906	0.3109
101	0.4332	0.0000	0.0909	0.4272	0.3523
102	0.4729	0.5000	0.3740	0.4678	0.3992
103	0.5182	0.0000	0.1636	0.5137	0.4521
			F= 0.73	F=82.95	F=62.23
			S=4.253	S=0.014	S=0.023

Bibliography

[1] Chatfield, C. (1985). The Analysis of Time Series, An Introduction. New York: Chapman and Hall. pp. 145-149.

[2] Garbow, B., et al. (1972). Matrix Eigensystem Routines: EISPACK Guide Extension. New York: Springer-Verlag.

[3] Golub, G. and Van Loan, C. (1983). Matrix Computations. Maryland: The Johns Hopkins University Press. pp. 125-135.

[4] ————, Ibid., pp. 24-29.

[5] ————, Ibid., pp. 228-266.

[6] Greenlee, H. and Keh, A. (1971) The 1971 Group Annuity Mortality Table. Trans. Soc. Actuaries. 23, p. 569.

[7] Grenville, T. (1957). On smoothing a finite table: A matrix approach. Soc. Indust. Appl. Math. Vol. 5, No. 3, pp 137-153.

[8] Kimeldorf, G. and Jones, D. (1967). Bayesian graduation. Trans. Soc. Actuaries. 19, pp. 66-112.

[9] London, D. (1985). Graduation: The Revision of Estimates. Connecticut: ACTEX Publications. pp. 71-85.

[10] ————, Ibid., p. 162.

[11] Jones, D. (1965). Bayesian graduation. Trans. Soc. Actuaries. 17, p. 33.

[12] Miller, R., and Wichern, D. (1977). Intermediate Business Statistics. New York: Holt, Rinehart and Winston. p. 225.

[13] Moler, C. (1980). MATLAB User's Guide. New Mexico: Technical Report CS81-1, Department of Computer Science, University of New Mexico. 87131.

[14] Peterson, R. (1952). Group annuity mortality. Trans. Soc. Actuaries. 7, pp. 246-307

[15] Smith, B, et al. (1970). Matrix Eigensystem Routines: EISPACK Guide, 2nd ed. New York: Springer-Verlag.