Advanced Statistical Analysis of Mortality

Rhodes, Thomas E. and Freitas, Stephen A.

MIB Solutions, Inc

Braintree Hill Office Park

Suite 400

Braintree, MA 02184-8734

(781)-751-6356

fax (781)-751-6060

trhodes@mib.com

Abstract

This paper demonstrates the utility of the Poisson Distribution in advanced statistical analysis of mortality in order to allow the researcher to obtain more information from their data. The use of the Poisson Distribution allows one to compare low numbers of deaths in a strata, thereby deriving more meaningful conclusions from the information. These techniques not only demonstrate that gender, medical basis, and rating are significant factors to consider in terms of their effect on mortality, but also the degree to which these factors affect mortality.

Keywords

Poisson Distribution, Poisson Regression, A/E Ratio, Mortality Analysis, A/E Ratio Confidence Limits, Right Tail Risk.

1. Statement of Problem

Past SOA reports have presented mortality statistics with a sufficient number of deaths per cell to satisfy the Law of Large Numbers which validates using the Normal Distribution. Upcoming Society of Actuaries' (SOA) mortality studies investigate combinations of underwriting classification, smoker status and face amount that result in fewer deaths per cell. The small numbers of deaths per cell calls in question the interpretation of mortality statistics. The SOA through its Mortality Study Working Group has recommended advanced statistical analysis to be done on mortality studies.

This paper discusses statistical models and presents the Poisson Distribution as a more theoretically correct statistical approach to mortality studies than the Normal Distribution. Using a mortality study of life insurance industry experience, this paper demonstrates the application of the Poisson Distribution and discusses the results.

2. Introduction

The use of the Normal Distribution assumes that a death in the right tail of the distribution is rare and that its effect on an insurance company is minimal. Deaths in fact are rare, but the impact on an insurance company by the payout of the death benefit is not minimal. In contrast, the Poisson Distribution assumes that the event that happens in the right tail is rare, but its impact is significant.

Insurance risk classifications of preferred, standard and substandard involve factors of age, gender, policy duration and smoking status. Individual underwriting information including the number and the degree of medical impairments discovered are used to assign an individual to a risk class.

The SOA analyzes mortality experience information for the insurance public through its various committees' work. These committees collect data from life insurance companies, combine the data to assure confidentiality and perform mortality studies. In presenting information about a particular risk classification, the most commonly used mortality measure is a statistic of the ratio of actual deaths divided by expected deaths (A/E ratio). One can easily evaluate how actual mortality is greater than or less than expected mortality by noting how far it differs from 1. As the number of deaths goes beneath 35 deaths per cell studied, many SOA reports deem the A/E ratio for such cells to be increasingly unreliable.

The numerical example in this paper is based on select duration experience of a subset of companies in the SOA's experience of 1995-1996 (1995 - 1996 Select Study).

3. Overview of the Normal Distribution Method

For mortality studies that involve large numbers of deaths per cell studied, actuaries have productively used the Normal Distribution based on the Law of Large Numbers. Using the Normal Distribution, Armitage used the variance stabilizing square root transformation to derive the confidence bands about the Standardized Mortality Ratio (SMR) or A/E ratio. This confidence limit is:

$$SMR = SMR * [1 +/- (Z_{\alpha/2} / (2*A^{1/2}))]^{2}$$

where: $Z_{\alpha/2}$ = The Z Score at the $\alpha/2$ level, where α is the Type I error rate or the probability that the estimate is outside of the range stated.

A = Actual number of deaths in the cell of interest.

For mortality studies, the use of confidence bands based on the Normal Distribution raises concerns. The Normal Distribution's reliability decreases when the small number of deaths per cell studied does not support the Law of Large Numbers. Additionally, the Normal Distribution has the problematic implicit assumption that a death claim has a minimal impact on an insurance company in the right tail of the mortality distribution.

4. Poisson Model Definition

The Poisson Distribution assumes that the event that happens is rare, but its impact is significant. Therefore, the Poisson Distribution more accurately reflects both the frequency and the financial impact of a death claim. Consequently, a Poisson Regression is a powerful analytical tool for mortality distributions. Additionally, the use of the Poisson Distribution removes the restriction of a minimum of 35 deaths in a cell and allows one to analyze and predict down to no deaths in a cell. This will be useful in upcoming SOA mortality studies where cells defined by underwriting classification, smoking status, face amount band, issue age, gender and policy duration will generally contain fewer than 35 deaths.

We will first discuss the Poisson Regression model and then discuss the Poisson A/E model.

4.1 Poisson Regression Model

The technique of Poisson Regression accurately fits and predicts insurance mortality. The Poisson Regression is developed using the basic Poisson form:

$$f(y) = (\mu^y e^{-\mu})/y! \quad \text{for } y = 0, 1, 2, \ldots$$

where: y = the number of events (in this case number of deaths)

μ = mean number of events

e = exponential function

Then, one develops the Poisson Regression with a log link function taking the form:

$$\ln(\mu_i) = \ln(n_i) + x_i^{/}\beta$$

where: μ_i = mean death rate for the i^{th} interval

n_i = the number of policyholders in the i^{th} interval

x_i = 1 if the value of independent variable is present, 0 otherwise

β = the beta coefficients

The Poisson Regression is used to make an analysis of a mortality distribution through comparing it to a baseline distribution. In the example presented in this paper, we take the distribution of deaths by issue age, gender, face amount, smoking status, medical basis and rating and compare it to an industry expected mortality basis of the 2001 Valuation Basic Tables (2001 VBT). In this way we can determine the factors that affect the death rates as well as determine whether these rates are significantly different from that of the 2001 VBT. The results of the Poisson Regression analysis inform the Poisson analysis of the A/E ratios.

4.2 Poisson A/E Model

The Poisson Distribution can be used to determine confidence limits on an A/E ratio. In order to rigorously calculate confidence limits on A/E ratios, one could use a complex, recursive calculation based on each probability based on an initial number of deaths decreasing to 0 by increments of 1. This time consuming approach can be addressed by using existing tables or an approximation technique. In Brackenridge and Elder, 1998, on pages 58 and 59 are presented tables for 90% and 95% confidence bands. In the example for this paper, we will use an approximation technique derived by Byar is shown on page 71 of Breslow and Day's 1989 text. This estimate is derived separately for the lower and upper confidence bound.

Lower Bound:

$SMR_L = A/E * [1 – (1/9*A) – (Z_{\alpha/2} / 3*A^{1/2})]^3$

where: A = Actual Number of Deaths

E = Expected Number of Deaths

$Z_{\alpha/2}$ = The Z Score at the $\alpha/2$ level, where α is the Type I error rate or the probability that the estimate is outside of the range stated.

Upper Bound:

$SMR_U = (A+1/E) * [1 – (1/9*(A+1)) + (Z_{\alpha/2} / (3 * (A+1)^{1/2})]^3$

where: A, E, and $Z_{\alpha/2}$ have the same definition as SMR_L

5. Numerical Example

The numerical example uses data from the 1995 – 1996 Select Study. This study was performed on experience in the select durations for a subset of companies in a Society of Actuaries study.

The subset of the data used for the example consists of policies from individuals with issue age 40 and over and of face amount $50,000 and higher. There were 6,940 deaths in this cohort and there were many cell conditions with less than 35 deaths. A Poisson Regression was fit to this data using the natural log of the number of expected deaths based upon the 2001 VBT as a basis of comparison (also referred to as an offset). The natural log of the expected deaths is used so that no one condition type with a high number of expected deaths would dominate the basis for comparison. Various levels of gender, face amount, issue age, medical basis, smoking status, and rating were used to predict actual numbers of deaths. The results of this comparison are shown in Table 1.

5.1 Discussion of Table 1 – Poisson Regression Fit

Each parameter in Table 1 uses a baseline condition against which other values of the parameter are compared. Since the males represent the baseline condition for gender, the estimate is 0.000 and the female estimates are based on a comparison to the males. Similarly, $50,000-$99,999 is the baseline for face amount, 40 – 49 for age, non medical for medical basis, nonsmoker for smoking status, and preferred for rating.

In Table 1, the combination of the estimate and the probability of the Chi-Square provide valuable statistical information. A positive estimate means that the condition results in higher mortality than the baseline condition. Similarly, a negative estimate means that the condition results in lower mortality than the baseline condition. If the probability of the Chi-Square statistic is less than 0.05, it indicates that this condition is significantly different from the baseline. For example, the females have lower mortality rates than the males have because the sign of the estimate is negative. The result is significant because the probability of the corresponding Chi Square is less than .05 (Pr > Chi Square = 0.00).

The mortality experience for face amount bands, age, and smoking status are not statistically significant among the various levels. This makes sense for smoking and age because the expected mortality basis of the 2001 VBT presents results by age and smoking status. The consistency among face amount bands indicates that the underwriting is consistent across all face amount bands.

The mortality experience for gender, medical basis and rating contain statistically significant results among the various levels. The female results were discussed above. The results for medical basis make sense because the paramedical and medical levels represent increased underwriting which results in lower mortality than for the baseline nonmedical level. The increase in mortality from preferred to standard rating is expected. Similarly, the unknown rating has an increase in mortality compared to the mortality of the preferred rating.

5.2 Discussion of Table 2 – Poisson Confidence Limits of A/E Ratio

Table 2 contains the A/E ratios with their respective 95% confidence bounds. These results are arranged with the rating and gender identified in the rows and the medical basis in the columns. For example, the preferred rating for females shows in the column marked 'Medical' results of 24 actual deaths, 40.73 expected deaths based on the 2001 VBT, an A/E ratio of .59 with a 95% Lower Confidence Interval of .38 and a 95% Upper Confidence Interval of .88. Since the estimates of the lower and upper confidence bounds are determined separately, the deviations from the A/E ratio are not symmetric. A confidence interval that does not include the value '1' describes a condition where the mortality experience is something other than the expected basis mortality of the 2001 VBT.

In the process of analyzing a mortality study, one would use a table similar to Table 2 to analyze A/E results. The confidence interval bounds on the A/E ratio adds valuable additional information on interpreting the A/E results. Furthermore, this analysis is available even on cells with small numbers of deaths.

6. Summary of the Results

This paper demonstrates the utility of the Poisson Distribution in advanced statistical analysis of mortality in order to allow the researcher to obtain more information from their data. The use of the Poisson Distribution allows one to compare low numbers of deaths in a strata, thereby deriving more meaningful conclusions from the information. These techniques not only demonstrate that gender, medical basis, and rating are significant factors to consider in terms of their effect on mortality, but also the degree to which these factors affect mortality.

We recommend that the advanced statistical methods including the use of A/E ratio confidence intervals based on the Poisson Distribution statistics be a part of the analysis for future mortality studies.

Table 1

Poisson Regression Fit to the 1995 - 1996 Select Study
Using the Natural Log of the 2001 VBT by Policy as the Offset

Parameter	Level	Degrees of Freedom	Estimate	Chi-Square	Pr > Chi-Square
Intercept		1	0.022	0.19	0.66
Gender	Male	0	0.000	.	.
Gender	Female	1	-0.089	8.53	0.00
Face Amount	50,000 - 99,999	0	0.000	.	.
Face Amount	100,000 - 249,999	1	-0.038	2.06	0.15
Face Amount	250,000 - 499,999	1	-0.092	3.04	0.08
Face Amount	500,000 - 999,999	1	-0.068	0.70	0.40
Face Amount	1,000,000 +	1	-0.072	0.39	0.53
Issue Age	40 – 49	0	0.000	.	.
Issue Age	50 – 59	1	0.020	0.47	0.49
Issue Age	60 – 69	1	0.011	0.10	0.75
Issue Age	70 - 79	1	0.005	0.01	0.94
Issue Age	80 +	1	-0.140	0.52	0.47
Medical Basis	Non Medical	0	0.000	.	.
Medical Basis	Paramedical	1	-0.130	12.24	0.00
Medical Basis	Medical	1	-0.216	24.99	<0.001
Smoke Status	Non Smoker	0	0.000	.	.
Smoke Status	Smoker	1	0.036	1.63	0.20
Smoke Status	Unknown	1	0.073	3.64	0.06
Rating	Preferred	0	0.000	.	.
Rating	Standard	1	0.115	7.72	0.01
Rating	Unknown	1	0.126	8.83	0.00

Table 2

Poisson Confidence Limits of A/E Ratios for the 1995 - 1996 Select Study

Rating	Gender	Category				
			Medical Basis			
Preferred	Male		Medical	Paramedical	Nonmedical	Total
		Actual Deaths	140	371	90	601
		2001 VBT Expected	187.49	411.97	82.38	681.84
		A/E Ratio	0.75	0.90	1.09	0.88
		95% Lower CI	0.63	0.81	0.88	0.81
		95% Upper CI	0.88	1.00	1.34	0.95
	Female					
		Actual Deaths	24	95	22	141
		2001 VBT Expected	40.73	100.24	27.44	168.41
		A/E Ratio	0.59	0.95	0.80	0.84
		95% Lower CI	0.38	0.77	0.50	0.70
		95% Upper CI	0.88	1.16	1.21	0.99
	Both					
		Actual Deaths	164	466	112	742
		2001 VBT Expected	228.22	512.21	109.82	850.25
		A/E Ratio	0.72	0.91	1.02	0.87
		95% Lower CI	0.61	0.83	0.84	0.81
		95% Upper CI	0.84	1.00	1.23	0.94
Standard	Male					
		Actual Deaths	960	1330	447	2737
		2001 VBT Expected	1129.45	1298.14	328.97	2756.56
		A/E Ratio	0.85	1.02	1.36	0.99
		95% Lower CI	0.80	0.97	1.24	0.96
		95% Upper CI	0.91	1.08	1.49	1.03
	Female					
		Actual Deaths	217	314	170	701
		2001 VBT Expected	242.88	355.28	135.49	733.65
		A/E Ratio	0.89	0.88	1.25	0.96
		95% Lower CI	0.78	0.79	1.07	0.89
		95% Upper CI	1.02	0.99	1.46	1.03
	Both					
		Actual Deaths	1177	1644	617	3438
		2001 VBT Expected	1372.33	1653.42	464.46	3490.21
		A/E Ratio	0.86	0.99	1.33	0.99
		95% Lower	0.81	0.95	1.23	0.95
		95% Upper	0.91	1.04	1.44	1.02
Unknown	Male					
		Actual Deaths	818	1240	172	2230
		2001 VBT Expected	769.31	1210.66	204.66	2184.63
		A/E Ratio	1.06	1.02	0.84	1.02
		95% Lower	0.99	0.97	0.72	0.98
		95% Upper	1.14	1.08	0.98	1.06
	Female					
		Actual Deaths	199	270	61	530
		2001 VBT Expected	204.69	308.87	83.08	596.64
		A/E Ratio	0.97	0.87	0.73	0.89
		95% Lower	0.84	0.77	0.56	0.81
		95% Upper	1.12	0.98	0.94	0.97
	Both					
		Actual Deaths	1017	1510	233	2760
		2001 VBT Expected	974.00	1519.53	287.74	2781.27
		A/E Ratio	1.04	0.99	0.81	0.99
		95% Lower	0.98	0.94	0.71	0.96
		95% Upper	1.11	1.05	0.92	1.03

Table 2 continued
Poisson Confidence Limits of A/E Ratios for the 1995 - 1996 Select Study
For all Rating Classes Combined

Rating	Gender	Category				
			Medical Basis			
All Rating Classes	Male		Medical	Paramedical	Nonmedical	Total
		Actual Deaths	1918	2941	709	5568
		2001 VBT Expected	2086.25	2920.77	616.01	5623.03
		A/E Ratio	0.92	1.01	1.15	0.99
		95% Lower	0.88	0.97	1.07	0.96
		95% Upper	0.96	1.04	1.24	1.02
	Female					
		Actual Deaths	440	679	253	1372
		2001 VBT Expected	488.30	764.39	246.01	1498.70
		A/E Ratio	0.90	0.89	1.03	0.92
		95% Lower	0.82	0.82	0.91	0.87
		95% Upper	0.99	0.96	1.16	0.97
	Both					
		Actual Deaths	2358	3620	962	6940
		2001 VBT Expected	2574.55	3685.16	862.02	7121.73
		A/E Ratio	0.92	0.98	1.12	0.97
		95% Lower	0.88	0.95	1.05	0.95
		95% Upper	0.95	1.01	1.19	1.00

References

Armitage P. and Berry G. (1994) *Statistical Methods in Medical Research* (3rd edition). Blackwell, London.

Brackenridge, R.D.C and Elder W. John, editors (1998) *Medical Selection of Life Risks* (4th edition), pp58 – 59, Stockton Press, New York.

Breslow and Day (1989) *Statistical Methods in Cancer Research, Volume 2 – The Design and Analysis of Cohort Studies*. Oxford University Press, New York: International Agency for Research on Cancer.

Thomas E. Rhodes, FSA, MAAA
Actuarial Director
MIB Solutions, Inc.
Braintree Hill Office Park
Suite 400
Braintree, MA 02184-8734

Stephen A. Freitas, FLMI, ACS
Senior Statistician
MIB Solutions, Inc.
Braintree Hill Office Park
Suite 400
Braintree, MA 02184-8734