

MS Module 24: Actuarial risk classification – practice problems

(The attached PDF file has better formatting.)

Exercise 24.1: Least squares bias function – practice problems

The mean values and the number of observations in each cell of a 2×2 classification table are

Means	Column 1	Column 2	Observations	Column 1	Column 2
Row 1	50	30	Row 1	15	12
Row 2	20	8	Row 2	6	10

Illustration: The cell in row 1 column 1 has a mean of 50 from a sample of 15 observations.

An actuary is setting class relativities for insurance pricing using a multiplicative model and a least squares bias function with

- ! a base rate of 10
- ! a starting relativity for column 1 of 1.8
- ! a starting relativity for column 2 of 1.0

We use the following notation:

- M_{jk} = mean value for the cell with row j and column k
- N_{jk} = number of observations for the cell with row j and column k
- B = base rate
- r_1 = relativity for Row 1
- r_2 = relativity for Row 2
- c_1 = relativity for Column 1
- c_2 = relativity for Column 2

- A. What are the squared errors in each cell?
- B. What is the mean squared error?
- C. What is the partial derivative equation for the Row 1 relativity?
- D. What is the implied relativity for Row 1, given the starting relativities by column?
- E. What is the partial derivative equation for the Row 2 relativity?
- F. What is the implied relativity for Row 2, given the starting relativities by column?
- G. What is the partial derivative equation for the Column 1 relativity?
- H. What is the implied relativity for Column 1, given the computed relativities by row?
- I. What is the partial derivative equation for the Column 2 relativity?
- J. What is the implied relativity for Column 2, given the computed relativities by row?

Part A: We compute the squared error of each observation from its expected value using the cell relativities.

For the Row 1 – Column 1 cell, the observed value is 50 and the expected value is $B \times r_1 \times c_1$, so the squared error is $(50 - B \times r_1 \times c_1)^2$. This cell has 15 observations, so the squared error is $15 \times (50 - B \times r_1 \times c_1)^2$. The squared errors for the other three cells have the same format.

Question: The mean of this cell is 50, but each observation is not necessarily 50. Shouldn't we use the actual 15 observations in the cell, not the 15 instances of the mean of the observations?

Answer: Ideally, we should examine the variance of each cell, but pricing actuaries almost never have detailed data showing the observations in each cell. The solution here assumes that the variance is the same in each cell, just as analysis of variance assumes.

Computing the actual variance in cell from the sample data is much work; the computation effort exceeds the benefit of slightly more accurate rates. Actuaries sometimes assume a distribution that implies a relation of the variance to the mean, such as a Poisson distribution of claim frequency or a Gamma distribution of claim severity. The maximum likelihood bias function in the “Actuarial risk classification” reading shows how pricing actuaries use maximum likelihood to set rates for a classification table.

Some insurers use generalized linear modeling to set premium rates, using Poisson distributions for accident frequency and log-link functions in multiplicative models. The balance principle with a multiplicative model is identical to a generalized linear model with a Poisson distribution and a log-link function. The balance principle is easier for non-mathematicians to understand and it is more intuitive: balancing the observed and expected values seems the proper way to set rates.

Generalized linear models give confidence intervals and p values that are not provided by classification tables. In theory, confidence intervals and p values are useful. But almost no pricing actuary computes p values or confidence intervals when setting rates. Actuaries consider the volume of business, industry experience, and competitors’ rates to judge the significance of their work.

Part B: We sum the squared errors for the four cells to get the sum of squared errors (SSE).

$$\begin{aligned} \text{SSE} &= N_{1,1} \times (M_{1,1} - B \times r_1 \times c_1)^2 && \text{row 1, column 1} \\ &+ N_{1,2} \times (M_{1,2} - B \times r_1 \times c_2)^2 && \text{row 1, column 2} \\ &+ N_{2,1} \times (M_{2,1} - B \times r_2 \times c_1)^2 && \text{row 2, column 1} \\ &+ N_{2,2} \times (M_{2,2} - B \times r_2 \times c_2)^2 && \text{row 2, column 2} \end{aligned}$$

Part C: To minimize the sum of the squared errors, we set the partial derivative with respect to each variable equal to zero. For the *row 1* relativity (r_1), we have

$$0 = \partial \text{SSE} / \partial r_1 = 2 \times N_{1,1} \times (M_{1,1} - B \times r_1 \times c_1) \times (-B \times c_1) + 2 \times N_{1,2} \times (M_{1,2} - B \times r_1 \times c_2) \times (-B \times c_2)$$

We consider the cells only in the (r_1) row. The other cells do not have an r_1 term in the squared error, so the partial derivative with respect to r_1 is zero.

Taking partial derivatives with respect to each of the class relativities gives four equations in four unknowns. The equations are not linear, so we solve them iteratively.

Part D: Using the means and number of observations in each cell and the starting values of $B = 10$, $c_1 = 1.8$, and $c_2 = 1$, we have

$$\begin{aligned} &2 \times 15 \times (50 - 10 \times r_1 \times 1.8) \times (-10 \times 1.8) \\ + &2 \times 12 \times (30 - 10 \times r_1 \times 1.0) \times (-10 \times 1.0) = 0 \end{aligned}$$

We factor out the “2” at the beginning of each line and the second “-10” from the equation, giving

$$r_1 = (15 \times 50 \times 1.8 + 12 \times 30 \times 1.0) / (10 \times 15 \times 1.8^2 + 10 \times 12 \times 1.0^2) = 2.821782$$

Part E: For the *row 2* relativity (r_2), we have

$$0 = \partial \text{SSE} / \partial r_2 = 2 \times N_{2,1} \times (M_{2,1} - B \times r_2 \times c_1) \times (-B \times c_1) + 2 \times N_{2,2} \times (M_{2,2} - B \times r_2 \times c_2) \times (-B \times c_2)$$

We consider the cells only in the (r_2) row. The other cells do not have an r_2 term in the squared error, so the partial derivative with respect to r_2 is zero.

Part F: Using the means and number of observations in each cell and the starting values of $B = 10$, $c_1 = 1.8$, and $c_2 = 1$, we have

$$2 \times 6 \times (20 - 10 \times r_2 \times 1.8) \times (-10 \times 1.8) + 2 \times 10 \times (8 - 10 \times r_2 \times 1.0) \times (-10 \times 1.0) = 0$$

We factor out the “2” at the beginning of each line and the second “-10” from the equation, giving

$$r_2 = (6 \times 20 \times 1.8 + 10 \times 8 \times 1.0) / (10 \times 6 \times 1.8^2 + 10 \times 10 \times 1.0^2) = 1.005435$$

Part G: For the *column 1* relativity (c_1), we have

$$0 = \partial \text{SSE} / \partial c_1 = 2 \times N_{1,1} \times (M_{1,1} - B \times r_1 \times c_1) \times (-B \times r_1) + 2 \times N_{2,1} \times (M_{2,1} - B \times r_2 \times c_1) \times (-B \times r_2)$$

We consider the cells only in the (c_1) column. The other cells do not have a c_1 term in the squared error, so the partial derivative with respect to c_1 is zero.

Part H: Using the means and number of observations in each cell, the base value $B = 10$, and the computed values of $r_1 = 2.821782$, and $r_2 = 1.005435$, we have

$$2 \times 15 \times (50 - 10 \times 2.821782 \times c_1) \times (-10 \times 2.821782) + 2 \times 6 \times (20 - 10 \times 1.005435 \times c_1) \times (-10 \times 1.005435) = 0$$

We factor out the “2” at the beginning of each line and the second “-10” from the equation, giving

$$c_1 = (15 \times 50 \times 2.821782 + 6 \times 20 \times 1.005435) / (10 \times 15 \times 2.821782^2 + 10 \times 6 \times 1.005435^2) = 1.782430$$

Part I: For the *column 2* relativity (c_2), we have

$$0 = \partial \text{SSE} / \partial c_2 = 2 \times N_{1,2} \times (M_{1,2} - B \times r_1 \times c_2) \times (-B \times r_1) + 2 \times N_{2,2} \times (M_{2,2} - B \times r_2 \times c_2) \times (-B \times r_2)$$

We consider the cells only in the (c_2) column. The other cells do not have a c_2 term in the squared error, so the partial derivative with respect to c_2 is zero.

Part J: Using the means and number of observations in each cell, the base value $B = 10$, and the computed values of $r_1 = 2.821782$, and $r_2 = 1.005435$, we have

$$2 \times 12 \times (30 - 10 \times 2.821782 \times c_2) \times (-10 \times 2.821782) + 2 \times 10 \times (8 - 10 \times 1.005435 \times c_2) \times (-10 \times 1.005435) = 0$$

We factor out the “2” at the beginning of each line and the second “-10” from the equation, giving

$$c_2 = (12 \times 30 \times 2.821782 + 10 \times 8 \times 1.005435) / (10 \times 12 \times 2.821782^2 + 10 \times 10 \times 1.005435^2) = 1.037566$$