A GENERALIZED THEORY OF CREDIBILITY

BY

ARTHUR L. BAILEY

INTRODUCTION

Casualty insurance actuaries and statisticians have used credibility factors for many years. They have been satisfied that the use of such a consistently conservative procedure produces much more practical results than any attempt to follow one hundred percent the indications of a limited volume of their statistics. The writer will attempt to show that they have been justified in their pragmatical outlook.

Let it be necessary to obtain estimates of the population means, X_i, of N independent characteristics, for each of which K observations have been made. Let it be assumed that the population variances of all N characteristics are identical but unknown. Let the j-th observation of the i-th characteristic be z_{ij}, and let its departure from the i-th population mean be y_{ij}. Then $z_{ij} = X_i + y_{ij}$ and the unknown population variance of y_{ij} is $\sigma^2{}_y$ for all of the N populations.

The non-insurance statistician seeks the "best unbiased estimates" and imposes as the conditions to determine his estimates of the values of X_i:

(1) that the estimate *for each characteristic* be unbiased, and

(2) that the error-variance of the estimates be a minimum *for each characteristic.*

This gives him as his estimate of X_i the average of the K observations of z_{ij} for the particular value of i, or z_i. These estimates have an error-variance of $\sigma^2{}_y/K$ for each characteristic.

The insurance statistician seeks the "best unbiased *set* of estimates" and imposes as the conditions to determine his estimates:

(1′) that the average estimate *for all characteristics as a whole* (i.e. the rate level) be unbiased and

(2′) that the error-variance of the estimates be a minimum *for all characteristics combined.*

It will be shown that this gives him as his estimate of X_i the average of the K observations of z_{ij}, for the particular value of i, multiplied by a credibility of C, plus the average of the KN observations of z_{ij}, for all values of i, multiplied by the complement of the credibility, or:

$$\text{Estimated } X_i = Cz_i + (1 - C)\,\overline{z}.$$

These estimates of X_i are not unbiased for each characteristic but the average for all characteristics combined is unbiased. The value of C, as will be shown later, must be taken as equal to $\dfrac{K}{K+\dfrac{\sigma^2_y}{\sigma^2_X}}$ in order to minimize the error-variance of the estimates, and this error-variance becomes $C\sigma^2_y/K$ for all characteristics combined.

There is a further refinement of the above simple case which the insurance statistician finds it necessary to recognize which other statisticians usually neglect, namely, the number of units of exposure represented by an observation. The observations in insurance statistics are of necessity made on risks varying in size of exposure. In other statistical fields the measurements may often be limited to individuals of the same physical magnitude, although not invariably. For example, an agricultural experiment may be laid out to cover plots of a uniform size, say 50 x 50 feet. The results of such an experiment may, however, be indiscriminately compared with a similar experiment involving plots 10 x 10 feet; a comparison which may be quite improper in several respects. In the following more general development proper recognition has been given to variations in the number of units of exposure by letting y_{ij} be the sum of m independent values of u_{ijh} corresponding to the m units of exposure involved. In actual application, each y_{ij} may have a different value of m; but, as m will appear in the formulae to be developed, it will only be necessary to use the proper value of m for the case to which the formulae are applied.

In the following development of the most general case the characteristic X_i will be represented by amx_i, where m represents the number of units of exposure in each observation, a the average value of X_i per unit of exposure for all values of i or $\bar{X}/m$, and x_i the ratio of X_i to $\bar{X}$. In this way the effect of each component of the total variation in X_i will be separately considered and the treatment of special cases will be facilitated.

The Problem and Proposed Solution

Let it be desired to obtain a set of estimates of an element of variation under conditions such that it is impossible to observe that element alone. The procedure presented herein consists, where ax_i represent the elements of variation to be estimated, and where z_i represent the mean values of k observations each covering m units of time or space (i.e. exposure) of the variable z_{ij}, which is the sum of amx_i and y_{ij}, of using as the set of estimates of ax_i the values obtained from the linear regression equation of ax_i on z_i. The required regression coefficient, C, may be evaluated from an analysis of the variance of z_{ij} and, in the insurance business, where several special cases of the general theory occur, is referred to as the "credibility" of the observation, z_i.

The General Case

Let it be assumed that:

(a) $z_{ij} = amx_i + y_{ij}$;

(b) $y_{ij} = \Sigma u_{ijh}$ for values of h from 1 to m;

(c) x_i has a mean of unity and a variance of σ^2_x;

(d) u_{ijh} has a mean of zero and a variance of $\sigma^2_{u_i}$ for all values of j;

(e) $\sigma^2_{u_i}$ has a mean of σ^2_u for all values of i;

(f) the m values of u_{ij} are random and independent;

(g) z_i is the average of k random and independent values of z_{ij}.

(h) Except that u_{ijh} and x_i are finite real numbers, no restrictions are placed on their distribution. (In most applications x_i will assume only discrete values).

Under these assumptions it can be shown that:*

(1) $$\bar{y} = E\,(y_i) = 0,\ \bar{z} = am, \text{ and } E\,(x_i\, y_i) = 0;$$

(2) $$\begin{aligned}\sigma^2_z &= E\,(z_i - \bar{z})^2 = E\,[am\,(x_i - \bar{x}) + (y_i - \bar{y})]^2 \\ &= a^2\,m^2\,\sigma^2_x + \frac{\sigma^2_y}{k} + 2am\,[E\,(x_i\,y_i) - \bar{y}E\,(x_i) - \bar{x}E\,(y_i) + \bar{x}\bar{y}] \\ &= a^2\,m^2\,\sigma^2_x + \frac{m\,\sigma^2_u}{k};\end{aligned}$$

(3) $$\begin{aligned}U_{xz} &= E\,(x_i - \bar{x})(z_i - \bar{z}) = E\,(x_i - \bar{x}\,)\,[am\,(x_i - \bar{x}\,) + (y_i - \bar{y}\,)] \\ &= amE\,(x_i - \bar{x}\,)^2 + E\,(x_i\,y_i) - \bar{y}E\,(x_i) - \bar{x}E\,(y_i) + \bar{x}\bar{y} \\ &= am\sigma^2_x.\end{aligned}$$

It will be seen from (2) that, if z_i/m is used as an estimate of ax_i, the variance of such an estimate will exceed the variance of ax_i by σ^2_u/km. The best unbiased set of estimates of ax_i will be the estimates obtained from the regression of ax_i on z_i/m. Such a regression will be linear if $\sigma^2_{u_i}$ is constant for all values of i; otherwise the form of the regression will depend on the relationship between $\sigma^2_{u_i}$ and ax_i and on the form of the distribution of x_i.

The best unbiased set of *linear* estimates of ax_i will be the estimates obtained from the equation of linear regression of ax_i on z_i/m, irrespective of the form of the true regression. This linear regression equation, using x'_i to indicate the value of x_i obtained from the regression equation, is:

* E is used to indicate the expected value of that which follows.

$$ax'_i = \frac{amU_{xz}}{\sigma^2_z} \cdot \frac{z_i}{m} + a - \frac{amU_{xz}}{\sigma^2_z} \cdot \frac{\bar{z}}{m} \text{ or, as } a = \frac{\bar{z}}{m} \text{ from (1)}$$

$$(4) \qquad ax'_i = C\,\frac{z_i}{m} + (1 - C)\,\frac{\bar{z}}{m}$$

where, from (2) and (3) :

$$(5) \qquad C = \frac{amU_{xz}}{\sigma^2_z} = \frac{a^2\, m^2\, \sigma^2_x}{a^2\, m^2\, \sigma^2_x + \dfrac{m\sigma^2_u}{k}} = \frac{km}{km + \dfrac{\sigma^2_u}{a^2\,\sigma^2_x}}.$$

When z_i/m is used as an estimate of ax_i the error of the estimate, $\xi_i = ax_i - z_i/m$, will have a variance, for all values of i, of:

$$(6) \qquad \sigma^2_\xi = a^2\,\sigma^2_x + \frac{\sigma^2_z}{m^2} - \frac{2a}{m} U_{xz}$$

$$= a^2\,\sigma^2_x + \left[\, a^2\,\sigma^2_x + \frac{\sigma^2_u}{mk}\right] - 2a^2\,\sigma^2_x$$

$$= \frac{\sigma^2_u}{mk}.$$

When $ax'_i = C\dfrac{z_i}{m} + (1 - C)\dfrac{\bar{z}}{m}$ is used as an estimate of ax_i the error of the estimate, $\xi'_i = ax_i - C\,\dfrac{z_i}{m} - (1 - C)\dfrac{\bar{z}}{m}$, will have a variance for all values of i, of:

$$(7) \qquad \sigma^2_{\xi'} = a^2\,\sigma^2_x + \frac{C^2}{m^2}\,\sigma^2_z - 2\frac{Ca}{m}\,U_{xz}$$

$$= a^2\,\sigma^2_x + C^2\left(a^2\,\sigma^2_x + \frac{\sigma^2_u}{mk}\right) - 2Ca^2\,\sigma^2_x$$

$$= a^2\,\sigma^2_x\,(1 - C)^2 + C^2\frac{\sigma^2_u}{mk}$$

$$= C\,\frac{\sigma^2_u}{mk},$$

$$\text{as } a^2\,\sigma^2_x = \frac{C}{1 - C} \cdot \frac{\sigma^2_u}{mk} \text{ from (5).}$$

It will be noted that $\sigma^2_{\xi'}$ in (7) is less than σ^2_ξ in (6) except in the limiting case when $C = 1$.

An analysis of variance calculated* from K observations of z_{ij}, for which $m = M$, in each of N classifications of i, produces a variance within classification, W, and a variance between classifications, B, for which the expected values are:

* The capitals K, M, and N are used to represent the values of k, m, and n in data used to evaluate C in order to maintain a distinction between the values in such data and the values in data to which C is to be applied.

(8) $$E\,(W) = \frac{K}{K-1} E\,(z_{ij} - z_i)^2 = \frac{K}{K-1} \mathrm{E}\,(y_{ij} - y_i)^2$$
$$= E\,(\sigma^2{}_{y_i}) = M\,E\,(\sigma^2{}_{u_i}) = M\,\sigma^2{}_u\,;$$

(9) $$E\,(B) = \frac{NK}{N-1} E\,(z_i - \bar{z}\,)^2 = \frac{NK}{N-1} E\,[aM\,(x_i - \bar{x}) + (y_i - \bar{y})]^2$$
$$= \frac{NK}{N-1}\,[a^2\,M^2\,E\,(x_i - \bar{x})^2 + E\,(y_i - \bar{y}\,)^2$$
$$+ 2aM\,\{E\,(x_i\,y_i) - \bar{y}\,E\,(x_i) - \bar{x}\,E\,(y_i) + \bar{x}\bar{y}\}]$$
$$= Ka^2\,M^2\,\sigma^2{}_x + M\,\sigma^2{}_u\,;$$

and $\dfrac{MK\,E\,(W)}{E\,(B) - E\,(W)} = \dfrac{\sigma^2{}_u}{a^2\,\sigma^2{}_x}$, so that C may be evaluated by using in (5) the value:

(10) $$\frac{\sigma^2{}_u}{a^2\,\sigma^2{}_x} = \frac{MKW}{B-W}.$$

It is interesting to note that, when $k = K$ and $m = M$, as is the case when C is to be applied to the same data from which it is evaluated, the calculated value of C reduces to:

(11) $$C = 1 - \frac{W}{B},$$

and that, when $k = 1$, C is the intraclass correlation coefficient, and when $m = M$, the calculated value of C becomes the value given by R. A. Fisher* as the unbiased estimate of the intraclass correlation coefficient:

(12) $$C = \frac{B - W}{B + (K-1)\,W} = \rho.$$

When K_i varies for different values of i in the actual observations used for the analysis of variance, it is possible to calculate W and B, where $\bar{K}$ represents the average value of K_i, as follows:

(13) $$W = \frac{\Sigma \frac{K_i}{K_i - 1}(z_{ij} - z_i)^2}{\bar{K}N} \qquad B = \frac{\Sigma\,K_i\,(z_i - \bar{z})^2}{\bar{K}\,(N-1)},$$

and to evaluate C from (10) using $\bar{K}$ in place of K.

It should be noted that no assumption is made in the general case as to the form of the distributions of u_i and x_i except that all values be finite real numbers. When the form of either distribution or the variance of either distribution is known or assumed, special cases arise in many of which more efficient estimates of $\sigma^2{}_u/a^2\,\sigma^2{}_x$ can be obtained by methods other than the analysis of variance.

* In Section 40 of "Statistical Methods for Research Workers."

Special Cases

The simplest special case occurs when $\sigma^2_{u_i}$ is assumed to be constant for all values of i. An interesting special case is met in life insurance statistics, where $amx_i + y_{ij}$ is assumed to be distributed according to the binomial distribution with a mean of amx_i and the variance, $\sigma^2_{u_i} = ax_i(1 - ax_i)$. With casualty insurance statistics the special case of the Poisson distribution arises for the number of losses, the special multiplicative case arises for the average amounts of losses, and a combination of these two arises in dealing with the total amount of all losses incurred. The Poisson and multiplicative special cases are outlined below in order to indicate the method of approach to other special cases.

The Poisson Case

The number of events, z_i, occurring in m_i units of space are observed for n values of i. It is assumed that the number of events occurring in single units of space are distributed in a Poisson distribution for each value of i but with different means and with variances equal to these means. Denoting the means by ax_i, we have the general case but with $k = 1$ and $\sigma^2_{u_i} = ax_i$, from which $\sigma^2_u = a$. Also m_i is not constant and $\bar{z}/m$ in the regression equation will have to be replaced by $\bar{z}/\bar{m}$.

The regression of ax_i *on* z_i/m_i will not be linear, but the linear regression equation will afford a very close approximation when $am_i x_i$ exceeds 10. The best unbiased set of linear estimates of ax_i, from (4) and (5), is obtained from:

$$ax'_i = C\,\frac{z_i}{m_i} + (1 - C)\,\frac{\bar{z}}{\bar{m}} \tag{14}$$

where:

$$C = \frac{m_i}{m_i + \dfrac{1}{a\,\sigma^2_x}}. \tag{15}$$

An analysis of variance is obviously impossible in this case. Even if it were possible, it would not provide the most efficient estimate of $1/a\,\sigma^2_x$, as the assumption of the Poisson distribution has completely determined $\sigma^2_{u_i}$ in terms of ax_i, thereby making other methods more efficient. If $v_i = z^2_i/m_i$, the expected values of z_i and v_i, for a specified value of m_i, are:

$$E(z_i) = am_i \qquad E(v_i) = \frac{E(z^2_i)}{m_i} = a^2 m_i(\sigma^2_x + 1) + a$$

and, for the observed values of m_i, are:

$$E(z_i) = a\bar{m}\; E(v_i) = a^2\,\bar{m}(\sigma^2_x + 1) + a,$$

so that $1/a\,\sigma^2_x$ can be evaluated from:

$$\frac{1}{a\,\sigma^2_x} = \frac{\bar{z}\bar{m}}{\bar{v}\bar{m} - \bar{z} - \bar{z}^2}. \tag{16}$$

The Multiplicative Case

The yields of k_i units of variety i are observed as z_{ij} for n varieties. It is assumed that the expected yield per unit of variety i is ax_i and that the variance of the yields of individual units of a variety is proportional to the square of the expected yield and is $ba^2 x^2_i$. This is referred to as the multiplicative case because, if $w_{ij} = 1 + \frac{y_{ij}}{ax_i}$, z_{ij} could be represented as: $z_{ij} = ax_i w_{ij}$, where x_i has a mean of unity and a variance of σ^2_x and w_{ij} has a mean of unity and a variance of b.

Here we have the general case with $m = 1$ and $\sigma^2_{u_i} = ba^2 x^2_i$, from which $\sigma^2_u = ba^2 (\sigma^2_x + 1)$. The best unbiased set of linear estimates of ax_i, from (4) and (5), is obtained from:

$$ax'_i = Cz_i + (1 - C)\bar{z}, \tag{17}$$

where:

$$C = \frac{k_i}{k_i + \frac{b(\sigma^2_x + 1)}{\sigma^2_x}}. \tag{18}$$

If an analysis of variance is made in accordance with (13), C can be evaluated by using in (18) the value:

$$\frac{b(\sigma^2_x + 1)}{\sigma^2_x} = \frac{\bar{k}W}{B - W}. \tag{19}$$

It frequently occurs in this special case that the value of b is known, in which case the evaluation of C can be made as follows. Let $t_i = k_i z_i$ and $v_i = k_i z^2_i$. For a specific value of k_i:

$$E(t_i) = ak_i \qquad E(v_i) = k_i E(z^2_i)$$
$$= k_i a^2 (\sigma^2_x + 1) + ba^2 (\sigma^2_x + 1)$$

and, for the observed values of k_i:

$$E(t_i) = a\bar{k} \qquad E(v_i) = a^2 (\sigma^2_x + 1)(\bar{k} + b),$$

so that C can be evaluated by using in (18) the value:

$$\frac{(\sigma^2_x + 1)}{\sigma^2_x} = \frac{\bar{v}\bar{k}^2}{\bar{v}\bar{k}^2 - \bar{t}^2(\bar{k} + b)}. \tag{20}$$

In this multiplicative case consideration might be given to the desirability of making the sum of the squares of the percentage errors a minimum rather than the sum of the squares of the arithmetic errors.

Conclusion

The conditions which underly and justify the use of credibilities in making a set of estimates from observations based on a limited amount of exposure have been presented. A review of the general case shows such a procedure to be directly applicable in many fields of practical statistics outside of as well as in the insurance business. Methods of evaluating the proper credibilities have been presented both from an analysis of variance and, in the special applications considered, from more directly available statistics.

One warning should be given the reader. It has *not* been shown that the use of any arbitrarily selected schedule of credibilities is justified. As a matter of fact, if such an arbitrarily chosen credibility, C', is less than:

$$\frac{km - \dfrac{\sigma^2_u}{a^2\,\sigma^2_x}}{km + \dfrac{\sigma^2_u}{a^2\,\sigma^2_x}}$$

it can be shown from equation (7) that $\sigma^{2\prime}_{\xi}$, is greater than σ^2_u/mk and that the use of the arbitrary credibility has produced a greater error-variance than would have resulted from giving each observation 100% credibility. The converse of this is that, as long as km is less than $\sigma^2_u/a^2\,\sigma^2_x$, any credibility between 0 and 1 will produce a lesser error-variance than the use of 100% credibility.

It should also be mentioned that the presentation herein has been limited to cases where the estimates are to be made for the i-th variable from the observations of the i-th variable and from the observed mean of all variables. The same type of technique is applicable and the theoretical development is similar in the case where a set of estimates based on a large volume of data are to be modified to obtain a set of estimates which reflect the peculiarities of a comparatively limited volume of data for which unknown changes have occurred in the values of the variables x_i. Such techniques become necessary in the revision of insurance rates to reflect the most recent conditions affecting hazards and in the modification of insurance rates to reflect the peculiarities of individual insureds and have been developed specifically for such applications.* Corresponding procedures would appear to be proper in many other fields of statistical application where sets of estimates are desired having a minimum error variance.

* See "Sampling Theory in Casualty Insurance" by A. L. Bailey P.C.A.S. Vol. XXIX, page 69—The Fundamentals of Experience Rating, and Vol. XXX, page 64—The Greatest Accuracy Credibility.