SOME SIMPLE DEVELOPMENTS IN THE USE OF THE COEFFICIENT OF STABILITY

By C. H. Forsyth

Some time ago the writer proposed[1] a coefficient of stability C_s to be used to measure the stability of a statistical series, where that coefficient is defined by the relation

$$C_s = \frac{\sigma^2}{M} \tag{1}$$

where M denotes the arithmetic mean and σ^2 the square of the dispersion of the terms of the series. It was proposed to regard series as unstable (Lexian) for which the value of the coefficient exceeded unity, and stable otherwise. The only essential way in which such a procedure differs *in results* from the traditional method is that it includes as stable those series for which the value of the coefficient lies between unity and q the probability of failure of the event under investigation—series which would be classed as unstable according to the traditional method. Stable series—according to either standard—are found so rarely in practice and therefore so many series are accepted as fairly stable which come anywhere near meeting the requirements that replacing q by unity as the line of demarcation affects the classification of no known series but adds to the effectiveness of the avowed purpose and use of the proposed coefficient—to avoid the round-about work of computing values of probabilities. Another merit of the use of the coefficient is that it enables one to measure and therefore *compare* the stability of several series—a feature which we shall illustrate later.

In brief, such a coefficient provides a means of introducing the whole Lexian theory into Federal publications such as those on vital statistics, since a comparison of the values of the coefficient for, say different communities or countries, would be readily grasped by any reader, whereas the traditional method would prove too subtle and laborious, and allow no ready comparison of results.

For purpose of orientation let us illustrate the situation by analyzing a simple series both ways—the traditional way and by the use of the coefficient of stability. As an example, let us consider the death rates of white infants under one year of age for 1919 (considered on page 89 of the Handbook) for those states whose frequencies of births are comparable or which vary little from

[1] Journal of the American Statistical Association, June, 1932.

their average of 47,830—where the number of deaths for each state has been adjusted to this average as a base.

	Adjusted Deaths X	$X - 3659$	$(X - 3659)^2$
Cal.	3350	−309	95481
Conn.	4700	1041	1083681
Ind.	3732	73	5329
Kan.	3253	−406	164836
Ky.	3686	27	729
Minn.	3159	−500	250000
N. Car.	3541	−118	13924
Va.	3732	73	5329
Wis.	3780	121	14641
	9)32933	1335–1333	)1633950
	$M = 3659$		$181550 = \sigma^2$
			$\sigma = 426$

The traditional method would be:

The mean $M = np = 3659$ where $n = 47{,}830$.

$$\text{Hence } p = \frac{3659}{47830} \text{ and } q = \frac{44171}{47830}$$

$$\text{and } \sigma_B^2 = npq = 3659\left(\frac{44171}{47830}\right) = 3378$$

$$\text{whence } \sigma_B = 58.15$$

which is the value of the dispersion we should expect if the basic probability were constant throughout. But the value of the dispersion proves to be $\sigma = \sqrt{181550} = 426$, and the comparison of the values shows that the basic probability to be very variable and therefore the series to be very unstable or Lexian.

The computation of the value of the coefficient of stability is much more simple and direct

$$C_s = \frac{\sigma^2}{M} = \frac{181550}{3659} = 49.6$$

whose excess over unity also clearly indicates the instability of the series.

Since proposing the coefficient of stability the writer has been impressed by the overwhelming proportion of existing series (such as birth rates, various kinds of death rates, etc.) which employ arbitrary bases (such as "per thousand," "per ten thousand," etc.) usually without mention of the actual base. It is obvious, of course, that such rates, or occurrences per arbitrary base, say b, can first be adjusted to give occurrences per actual base, say B (assuming that base B* can be determined) but the work can evidently be performed much easier. For, since the original series (per arbitrary base b) $X_1, X_2, \cdots X_N$ would become, on adjustment, $\frac{B}{b}X_1, \frac{B}{b}X_2, \cdots \frac{B}{b}X_N$, the mean would become $\frac{B}{b}M$ and the square of the dispersion $\left(\frac{B}{b}\sigma\right)^2$, whence the formula for the coefficient of stability would become

$$C_s = \frac{\sigma^2}{M} \cdot \frac{B}{b} \tag{2}$$

As an example, let us consider the general death rates, per 10,000, of New Zealand for the years 1921–30.

	X	$X - 86$	$(X - 86)^2$
1921	87	1	1
1922	88	2	4
1923	90	4	16
1924	83	−3	9
1925	83	−3	9
1926	87	1	1
1927	85	−1	1
1928	85	−1	1
1929	88	2	4
1930	86	0	0
	10)862	10−8	)46
	M = 86.2		4.6

This example illustrates the danger of using the coefficient of stability unless the series consists of actual occurrences or unless the actual base is given due consideration. Without due consideration of the actual base (here the population of New Zealand) one might easily fall into the error of regarding the value of the coefficient of stability as 4.6/86.2 and, therefore, the series as very stable. But the population of New Zealand is about a million and a half and, therefore the true value of the coefficient of stability is

$$C_s = \frac{4.6}{86.2} \frac{1,500,000}{10,000} = 8.0$$

* Strictly speaking, this actual base B should be constant throughout the series; otherwise the successive numbers of occurrences—the terms of the series—would not be comparable. Where, however, the base B varies little from term to term—as usually happens even in the best of series, such as a series of some kind of rates of the same community over a short interval—the variation can be ignored, in which case base B (to which the terms of the series are adjusted) usually means the arithmetic mean of the different bases. In the first treated above, the investigation was limited to certain states in an effort to comply with the rule just mentioned but the example is a poor one since the variations are still dangerously too large. The situation is saved by the conclusive results.

which shows the series to be unstable. However, before we condemn New Zealand's death rates too severely, let us compare her record with those of other important countries, including our own, for the same period.

General Death Rates (per 10,000)

	M	C_s
New Zealand	86.2	8
Australia	94.3	90
Sweden	120.4	96
Scotland	137.3	139
Austria	151.1	536
United States	118.0	830
England-Wales	121.3	1117
France	170.3	1129
Spain	193.7	2190
Italy	163.5	2760
Germany	125.4	6040
Japan	206.4	6800

These results show how extremely unstable most series of general death rates are and that the series for New Zealand, while unstable according to our strict criterion, enjoys quite an enviable position practically in a class by itself. Parenthetically, these results also illustrate fairly well the triviality, with respect to results, of replacing q by unity as the critical value of the coefficient of stability, discussed at the beginning of this article.

The values of the coefficient listed above would, of course, be reduced somewhat in most cases if the trend of the series were first eliminated but the writer has gone though all this work and found it not worth while—that is, the series would still remain markedly unstable.

Another development proves useful when, as frequently happens, the actual base B is unknown to a degree of accuracy desirable for use in formula (2).

From the inequality $\frac{\sigma^2}{M}\frac{B}{b} \leqq 1$

we obtain

$$B \leqq \frac{Mb}{\sigma^2} \tag{3}$$

which is to be used to show how small an actual base should be for the given series to be stable. As an example, let us consider the maternal mortality, per 10,000 live births, in the so-called expanding registration area of the United States.

Maternal Deaths in the United States (per 10,000 live births) (Expanding Registration Area)

	X	$X - 66$	$(X - 66)^2$
1923	67	1	1
1924	66	0	0
1925	65	−1	1
1926	66	0	0
1927	65	−1	1
1928	69	3	9
1929	70	4	16
1930	67	1	1
1931	66	0	0
1932	64	−2	4
	10)665	9−4	)33
	66.5		3.3

Hence, by formula (3), $B \leqq \frac{66.5}{3.3}$ (10,000) or about 200,000. The number of live births varies so greatly that we should probably find it impossible to agree upon a satisfactory number[2] to use as an actual base for such an "expanding area" but we should all agree that it would be so much greater than 200,000 that the instability of the series would be unquestioned.

One must be careful in comparing the results of two or more investigations like the one just conducted. For example, the analogous result for Canada for the same period yields $B \leqq 113{,}000$ and we might conclude, too hastily, that the United States series is more stable (or less unstable) whereas any knowledge whatever of the numbers of live births of the two countries would show that Canada comes much closer to fulfilling her requirement than the United States and that the palm must go to Canada. For one thing, Canada has about the population of New York city and New York city has about 100,000 live births annually. In any case, *close* decisions in matters of this kind would be difficult without sufficient information in regard to actual bases.

There is still another situation which is interesting but of much less importance because of the rarity of its occurrence. It will be recalled that the coefficient of stability was devised mainly to avoid the use and computation of probabilities and that the only difference between the results by the traditional method and by the use of the coefficient of stability lies in the trivial replacement of the critical value q by unity. In the traditional method of analysis, but by comparing the value of the coefficient of stability with q, the coefficient is evidently always, strictly speaking, a function of the actual base B. In other words, there is no statistical series, however stable it may seem—except

[2] It was in the neighborhood of two million in 1932.

for the trivial case when all the terms of the series are exactly the same—but what would be unstable if the base were small enough. It is possible to formulate the limit once for all below which the given (otherwise seemingly stable) series would prove unstable.

If, in the relation $\sigma^2 \leqq npq$ (for stability) we replace p by M/n, q by $1 - M/n$ and then n by B, we obtain

$$\sigma^2 \leqq M - \frac{M^2}{B} \text{ or } \frac{M^2}{B} \leqq M - \sigma^2$$

whence, finally

$$B \geqq \frac{M^2}{M - \sigma^2} \tag{4}$$

where the transference of the term $M - \sigma^2$ from one side to the other should cause no apprehension since, by hypothesis, $\sigma^2 < M$ and $M - \sigma^2$ is therefore always positive. We propose to employ formula (4) in those rare cases where the value of the coefficient of stability of actual occurrences—but without reference to an actual base—is less than unity—that is, where the given series proves to be stable according to the method proposed by the writer—and determine the upper limit of the values of the base B for which the series would be unstable according to the traditional method of analysis. As an illustration, let us consider the familiar series of annual football fatalities in this country for the period 1906–1930* (omitting the years when no records were kept).

Football Fatalities

1906	11	1917	12
1907	11	1921	12
1908	13	1923	18
1909	12	1925	20
1911	11	1926	9
1912	13	1927	17
1913	5	1928	18
1914	13	1929	12
1915	15	1930	13

It is easily verified that $C_s = \frac{11.942}{13.055}$ which is clearly less than unity; whence the series clearly seems stable. Applying formula (4)

$$B \geqq \frac{13.055^2}{13.055 - 11.942} \quad \text{or } 153$$

which shows that the given series is stable as long as the total number of football players exceeds the number 153. A recent news item quoted an estimate of the number players participating in games of four hundred colleges as about

13,000 and over 600,000 including high schools and all. We can then definitely say that the series just considered is stable. Such a conclusion has no bearing, of course, upon what might happen if other terms were added to the series. It happens that adding the records for the next five years—1931(33), 1932(32), 1933(27), 1934(25), 1935(30)—would change the whole series to an unstable one with $C_s = 56.9/16.6 = 3.4$; but, obviously, the additional records belong to a new regime of collection.