The Philippines Statistician
Vol. XX, Nos. 1 and 2, 1971

ON STRATIFICATION USING THE LORENZ CURVE*

By ERNESTO R. RAMOS**

1. Introduction

The construction of strata poses several questions. What is the best characteristic to be used in stratification? How should the boundaries between the strata be determined?

For a single variable, the best characteristic is clearly the frequency distribution of the variable itself. The next best is the frequency distribution of some other variable highly correlated with the estimation variable.[1]

Durbin (1959) in his "Review of Sampling in Sweden" wrote:

> "On what basis should we choose the values of x which define the boundaries between strata? Treat ments of stratified sampling given in ordinary text-books ignore this problem completely; strata are conveniently regarded as *a priori*. In real life they never are and one has to construct them oneself on the base of ill-founded conjectures. Dalenius rendered a significant service by working out the theory on which rational choice of strata ought to be based."

2 The Optimum Stratification Method

The general equations defining the optimum stratification points derived by Dalenius (1957) under Neyman and proportional allocation as well as the equation that must be satisfied for equal allocation as reported by Sethi (1963) were worked out on the assumption that the estimation variable and the stratification variable are identical and that the frequency distribution is continuous.

---

* Excerpts from M.A. Thesis

** Instructor, UP Statistical Center

Assuming that the population with x_o and X_L as the smallest and largest values of x is stratified into L strata, the set of stratification points $x_1, x_2, \ldots, x_{L-1}$ is optimum that makes

$$\sigma^2(\bar{x}_{st}) = \sum_{h}^{L} \frac{W_h^2 \sigma_h^2}{N_h} \text{ a minnmum .}$$

Under Neyman allocation where $n_h = \frac{W_h \sigma_h}{\sum_h^L W_h \sigma_h}$, the general expression that must be satisfied is

$$\frac{(x_h - \mu_h)^2 + \sigma_h^2}{\sigma_h} = \frac{(x_h - \mu_{h+1})^2 + \sigma_{h+1}^2}{\sigma_{h+1}}$$

For equal allocation, $n_h = n/L$, it is

$$W_h [(x_h - \mu_h)^2 + \sigma_h^2] = W_{h+1} [(x_h - \mu_{h+1})^2 + \sigma_{h+1}^2] .$$

The general equation defining the optimum stratification points under proportional allocations or $n_h = w_h n$ is given as

$$x_h = \tfrac{1}{2} (\mu_h + \mu_{h+1}) .$$

It is quite clear that in actual survey strata construction must, of necessity, be based on the frequency distribution of a variate other than that which is yet to be measured in the survey. The stratification may be carried out using the frequency distribution of the estimation variable derived from a previous survey or enumeration. If previous data on this variable is not, however, available, one may use the frequency distribution of another variate highly correlated with the estimation variable. The use of the frequency distribution of the estimation variable itself could provide a more critical evaluation of the performances of the different stratification methods.

3. On Lorenz Curve Stratification

The general equations for determining the best stratum boundaries are, unfortunately, ill adapted to practical computations because of the difficulty and time involved in solving these equations. Quicker approximate methods have been developed due to the impracticality of solving complex simultaneous equations.

The primary concern of this study is to find out if it is possible to use the Lorenz curve as a means of determining the appropriate boundaries between the strata. The investigation will be confined to the following frequency distributions:

1. $f(x) = 1, O \leqq x \leqq 1$
2. $f(x) = 2(1-x), O \leqq x \leqq 1$
3. $f(x) = e^{-x}, x \geqq O$
4. $f(x) = x e^{-x}, x \geqq O$

These distributions except for the first one are typical forms of distributions encountered in the sampling of institutions characterized by a markedly positive skewness.

3.1. Stratification by Equal Partitioning of the Area of Concentration.

The Lorenz curve is defined by its cumulative distribution function

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

and

$$I(x) = \frac{1}{\mu}\int_{-\infty}^{x} tf(t)dt.$$

It is to be noted that I(x) is defined only if the mean, μ, exists. Also, assumng that the origin is to the left of the distribution, I(x) which is the incomplete first moment varies from O to 1 just as F(x) varies from O to 1.

The two equations above may be regarded as defining a relationship between the variables F and I in terms of parametric functions in X. The curve whose ordinate and abscissa are I and F respectively is called the Lorenz curve. Such a curve is shown below.

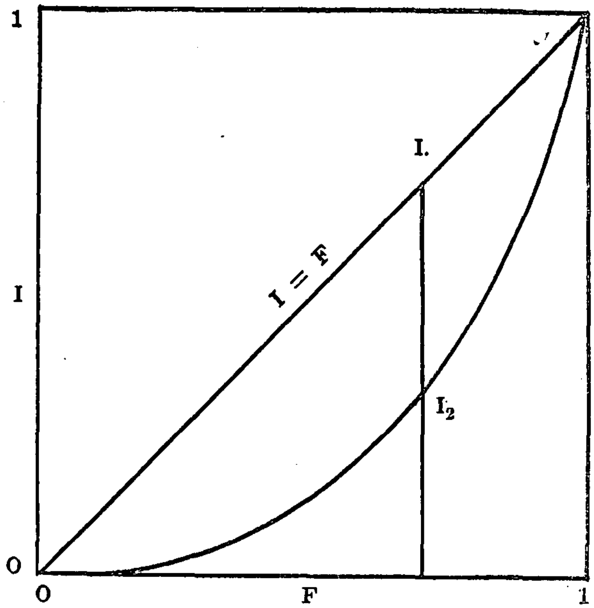

FIG. 1. The Lorenz Curve.

The area between the Lorenz curve and the line I = F is called the area of concentration.

It would be worthwhile to examine the corresponding set of values of the stratification variable obtained by dividing the area of concentration into L, the number of strata, equal parts. The division are lines vertical to the cumulative frequency axis of the Lorenz curve. The main reason for this is that areas are easily obtained through the use of a mechanical measuring device, by graphical method or by the use of a simple approximate method.

3.1.1. Stratification of the Rectangular Distribution

Let $f(x) = 1$, $O \leq x \leq 1$, be the specific rectangular distribution. The Lorenz curve of this distribution is obtained by solving for

$$F(x) = \int_0^x f(t)\, dt$$

which, in this case, is

$$F(x) = \int_0^x dt$$
$$= x,\ O \leqq x \leqq 1$$

and

$$I(x) = \frac{1}{\mu} \int_0^x t\, f(t)\, dt$$

where

$$\mu = \int_0^1 x dx = \tfrac{1}{2}\, x^2 \,\Big|_0^1 = \tfrac{1}{2}$$

Hence

$$I(x) = 2 \int_0^x t\, dt = t^2\Big|_0^x$$
$$= x^2,\ O \leqq x \leqq 1.$$

The area of concentration, A_T, is

$$A_T = \int_0^1 [I_1(x) - I_2(x)]\, dF(x)$$

$$= \int_0^1 [F(x) - F^2(x)]\, dF(x) = \tfrac{1}{2}F^2(x)\,\Big|_0^1 - \tfrac{1}{3}F^3(x)\Big|_0^1$$

$$= \tfrac{1}{2} - \tfrac{1}{3} = \tfrac{1}{6}.$$

For two strata, dividing the area of concentration into two equal parts by a line vertical to the cumulative frequency axis, the area of the lower section is

$$\int_0^{F(x_1)} [F(x) - F^2(x)]\, dF(x) = \tfrac{1}{2}\, A_T = \tfrac{1}{12}.$$

Or integrating the left hand side,

$$\tfrac{1}{2}\, F^2(x)\Big|_0^{F(x_1)} - \tfrac{1}{3}\, F^3(x_1)\Big|_0^{F(x_1)} = \tfrac{1}{12}$$

$$4\, F^3(x_1) - 6\, F^2(x_1) + 1 = O$$

or

$$[2F(x_1) - 1]\ [2F^2(x_1) - 2F(x_1) - 1] = O$$

for which

$$F(x_1) = 0.50$$

since the roots of the second factor (i.e. $F(x_1) = -0.37$ and $F(x_1) = 1.37$) are absurd as they are well beyond the range of F.

The corresponding x value is

$$x_1 = 0.50$$

since $F(x) = x$.

The results for $L = 3$ and 4, similarly derived, are shown in the table below together with the optimum stratification points under the different types of allocation.

TABLE 1. STRATIFICATION POINTS FOR THE RECTANGULAR DISTRIBUTION

$$f(x) = 1,\ O \leq x \leq 1$$

Number of Strata	Stratification Code	Method of Stratification			
		Lorenz Curve	Optimum by type of Allocation*		
			Neyman	Equal	Proportional
2	x_1	0.50	0.50	0.50	0.50
3	x_1	0.39	0.33	0.33	0.33
	x_2	0.61	0.67	0.67	0.67
4	x_1	0.33	0.25	0.25	0.25
	x_2	0.50	0.50	0.50	0.50
	x_3	0.67	0.75	0.75	0.75

* The optimum stratification points for the different allocations were obtained using the general equations given in Section 2.

The table above reveals that the Lorenz curve may be used for stratification as a rough approximation to the optimum stratification points of the rectangular distribution. It will be noted from the table that for $L = 2$, stratification by equal partitioning of the area of concentration yields a stratification point identical to the stratification point obtained by the optimum method. However, for L = 3 and 4, the strata constructed thru this method are wider at the lower and upper ends of the x scale in comparison with the strata formed by optimum stratification method. As an improvement to this method, adjustments which shall be called end corrections, should be applied when the number of strata is greater than two. The end corrections should be so devised such that the areas of the lower and upper portions of the area of concentration when it is divided are smaller than the areas of the middle portions. A formula, derived empirically, for the end corrections is suggested in Section 3.3.2.

Of greater import and relevance to the study is the conduct of this method of stratification when applied to a skewed distribution. For an illustration, consider the simple distribution $f(x) = 2(1 - x)$, $O \leqq x \leqq 1$, which has a relative measure of skewness equal to 0.566. The equations defining the Lorenz curve are

$$F(x) = 2x - x^2, \; O \leqq x \leqq 1$$

and

$$I(x) = 3x^2 - 2x^3, \; O \leqq x \leqq 1$$

for which the area of concentration is

$$A_T = \int_0^1 [I_1(x) - I_2(x)] \, dF(x)$$

$$= \int_0^1 [2x - x^2 - 3x^2 + 2x^3] \, 2(1 - x) \, dx$$

$$= \frac{1}{5} .$$

Partitioning the area of concentration into L equal parts, the following results, tabulated below, were obtained.

TABLE 2. STRATIFICATION POINTS FOR
$f(x) = 2(1-x), 0 \leqq x \leqq 1$

Number of Strata	Stratification Code	Method of Stratification			
		Lorenz Curve	Optimum by type of Allocation* Neyman	Equal	Proportional
2	x_1	0.31	0.35	0.50	0.38
3	x_1	0.23	0.23	0.39	0.25
	x_2	0.40	0.50	0.61	0.53
4	x_1	0.19	0.18	0.33	0.20
	x_2	0.31	0.37	0.50	0.42
	x_3	0.45	0.62	0.67	0.64

* The optimum stratification points for proportional allocation were obtained using the general equation given in Setion 2.3. while the optimum stratification points under Neyman and equal allocations were derived from Raj Des. "On Forming strata of Equal Aggregate Size," Journal of the *American Statistical Association*, 59 (1964), p. 485.

It is evident that the Lorenz curve stratification may be employed as an aid in the determination of strata boundaries even for mildly skewed distribution. Table 2 reveals that the Lorenz curve stratification deviates from the optimum method even for the two strata case. This situation demands that adjustment for skewness be made if the stratification is to be improved. Section 3.3.1 suggests an adjustment factor which defines the appropriate proportion of the area of concentration corresponding approximately to the optimum stratification points for Neyman allocation.

Further examination of the table reveals the need for end corrections as the strata constructed thru this method are wider at the top in comparison with the optimum method under Neyman allocation. These improvements suggested necessitates unequal partitioning of the area of concentration.

3.1.2. Stratification of the Gamma Distribution

Two simple Gamma distributions which approximates fairly well the distributions encountered in practice will be considered here. The first is the exponential distribution

$f(x) = e^{-x}, x \geqq O$. The other distributiton is defined as $f(x) = x\,e^{-x}, x \geqq O$.

Solving for the equations defining the Lorenz curve of the exponential distribution, the following are obtained.

$$F(x) = 1 - e^{-x}, x \geqq O$$

and

$$I(x) = 1 - e^{-x} - x\,e^{-x}, x \geqq O.$$

The area of concentration is obtained as

$$A_T = \int^{1} [I_1(x) - I_2(x)]\, dF(x)$$

$$= \int_0^{\infty} [(1 - e^{-x}) - (1 - e^{-x} - x\,e^{-x})]\, e^{-x}\, dx$$

$$= \int_0^{\infty} x\,e^{-2x}\, dx = - \tfrac{1}{2} x\,e^{-2x} \Big|_0^{\infty} - \tfrac{1}{4}\, e^{-2x} \Big|_0^{\infty}$$

$$= \tfrac{1}{4}$$

TABLE 3. STRATIFICATION POINTS FOR THE EXPONENTIAL DISTRIBUTION $f(x) = e^{-x}, x \geqq O$

Number of Strata	Stratification Code	Method of Stratification		
		Lorenz Curve	Optimum by type of Allocation*	
			Neyman	Equal
2	x_1	0.84	1.30	1.68
3	x_1	0.59	0.76	1.18
	x_2	1.15	2.07	2.30
4	x_1	0.48	0.55	0.96
	x_2	0.84	1.30	1.68
	x_3	1.35	2.64	2.69

* The optimum stratification points were obtained from Raj Des. "On Forming Strata of Equal Aggregate Size," *Journal of the American Statistical Association,* 59 (1964), p. 485.

The results of stratifying the distribution into 2, 3 and 4 strata are shown in Table 3. Note that for this particular distribution, a markedly skewed one with a relative measure of skewness equal to 2, the Lorenz curve stratification method resulted in a very rough approximation to the optimum stratification points. This suggests that adjustment for skewness and end corrections must be applied when the disttribution is highly skewed in order to improve the stratification points determined thru this method of stratification.

Consider another Gamma distribution, $f(x) = x\, e^{-x},\ x \geqq O$. The Lorenz curve of this distribution is defined by

$$F(x) = 1 - e^{-x} - x\, e^{-x},\ x \geqq O$$

and

$$I(x) = 1 - e^{-x} - x\, e^{-x} - \tfrac{1}{2}x^2\, e^{-x},\ x \geqq O\,.$$

with an area of concentration equal to

$$A_T = \int_O^1 [I_1(x) - I_2(x)]\ dF(x)$$

$$= \tfrac{1}{2} \int_O^\infty (x^2\, e^{-x})\ (xe^{-x})\ dx$$

$$= \tfrac{1}{2} \left\{ - \tfrac{1}{2}\, x^3\, e^{-2x} \Big|_O^\infty - \tfrac{3}{4}\, x^2\, e^{-2x} \Big|_O^\infty - \tfrac{3}{4}\, x\, e^{-2x} \Big|_O^\infty \right.$$

$$\left. - \tfrac{3}{8}\, e^{-2x} \Big|_O^\infty \right\}$$

$$= \frac{3}{16}$$

Let it be assumed that it was decided to have four strata. Now, the first stratification point is obtained by taking one-fourth of the area of the area of concentation, that is

$$\tfrac{1}{2} \int_O^{x_1} x^3\, e^{-2x}\, dx = \tfrac{1}{4}\, A_T = \tfrac{1}{4}\, (\tfrac{3}{16})$$

$$- \tfrac{1}{2}\, x_1^3\, e^{-2x1} - \tfrac{3}{4}\, x_1^2\, e^{-2x1} - \tfrac{3}{4}\, x_1\, e^{-1x2} - \tfrac{3}{8}\, e^{-2x1} + \tfrac{3}{8} = \tfrac{3}{16}$$

$$- \tfrac{1}{8}\, e^{-2x1}\, [4x_1^3 + 6x_1^2 + 6x_1 + 3] = - \tfrac{12}{32} + \tfrac{3}{32} = - \tfrac{9}{32}$$

or

$$4[4x_1^3 + 6x_1^2 + 6x_1 + 3] = 9\ e^{2x_1}$$

for which

$$x_1 \doteq 1.27$$

The second stratification is secured by dividing the area of concentration into two equal parts, which is

$$\tfrac{1}{2} \int_0^x x^3\ e^{-2x}\ dx = \tfrac{1}{2} A_T = \tfrac{1}{2}\ (\tfrac{3}{16})$$

Integrating and substituting limits,

$$\tfrac{1}{8}\ e^{-2x_2}\ 4x_2^3 + 6x_2^2 + 6x_2 + 3 = \tfrac{3}{8} - \tfrac{3}{16}$$

or

$$2\ [4x_2^3 + 6x_2^2 + 6x_2\ 3] = 3\ e^{2x_2}$$

Solving for x_2 graphically (that is taking the point of intersection of $y = 2[4x_2^3 + 6x_2^2 + 6x_2 + 3]$ and $y = 3e^{2x_2}$)

$$x_2 \doteq 1.84$$

The last stratification point, similarly derived, is equal to

$$x_3 \doteq 2.55$$

The stratification points after applying this method for 2 and 3 strata are shown in Table 4.

TABLE 4. STRATIFICATION POINTS FOR $f(x) = x\ e^{-x},\ x \geqq 0$

Number of Strata	Stratification Code	Method of Stratification		
		Lorenz Curve	Optimum by Type of Allocation*	
			Neyman	Equal
2	x_1	1.84	2.32	2.66
3	x_1	1.46	1.56	2.04
	x_2	2.88	3.28	3.44
4	x_1	1.27	1.22	1.72
	x_2	1.84	2.29	2.66
	x_3	2.55	3.96	3.92

* The optimum stratification points were obtained from Raj, Des. "On Forming Strata of Equal Aggregate Size," *Journal of the American Statistical Assocition,* 59 (1964) p. 485.

3.2. *Relative Precision of the Stratification Methods*

To render a better assessment of the performance of the different stratification methods, stratification was carried out using the frequency distribution of the estimation variable itself.

TABLE 5A. RELATIVE PRECISION OF THE STRATIFICATION METHODS OVER SIMPLE RANDOM SAMPLING

σ_1^2/σ_2^2 WITH $f(x) = 1, 0 \leqq x \leqq 1$

Number of Strata	Optimum Stratification			Lorenz Curve Stratification		
	Neyman	Equal	Propor-tional	Neyman	Equal	Propor-tional
2	1.92	1.92	1.92	1.92	1.92	1.92
3	8.96	8.96	8.96	8.01	10.28	7.71
4	16.02	16.02	16.02	13.02	19.37	12.07

TABLE 5B. RELATIVE PRECISION OF THE STRATIFICATION METHODS OVER SIMPLE RANDOM SAMPLING

σ_2^2/σ_2^2 WITH $f(x) = 2(1 - x), 0 \leqq x \leqq 1$

Number of Strata	Optimum Stratification			Lorenz Curve Stratification		
	Neyman	Equal	Propor-tional	Neyman	Equal	Propor-tional
2	3.68	3.63	3.59	3.63	3.43	3.35
3	8.06	7.94	7.72	6.95	5.40	5.79
4	14.26	13.56	13.24	10.69	6.95	8.06

TABLE 5C. RELATIVE PRECISION OF THE STRATIFICATION METODS OVER SIMPLE RANDOM SAMPLING

σ_1^2/σ_2^2 WITH $f(x) = e^{-x}, x \geqq 0$

Number of Strata	Optimum Stratification		Lorenz Curve Stratification	
	Neyman	Equal	Neyman	Equal
2	3.50	2.93	3.12	2.44
3	7.50	5.04	5.41	3.10
4	12.99	7.09	6.34	3.41

TABLE 5D. RELATIVE PRECISION OF THE STRATIFICATION METHODS OVER SIMPLE SAMPLING

σ_1^2/σ_2^2 WITH $f(x) = x\,e^{-x},\ x \geqq 0$

Number of Strata	Optimum Stratification		Lorenz Curve Stratification	
	Neyman	Equal	Neyman	Equal
2	3.17	2.87	2.96	2.69
3	6.47	4.78	3.95	2.78
4	11.12	7.14	7.42	3.85

For the rectangular distribution where the optimum stratification points for all the three types of allocation were identical, the relative precision of stratified sampling for two strata over simple random sampling is 192%. Dividing the distritbution into four strata further increases the relative precision to 1602% as disclosed by Table 5.A.

It is interesting to note that the Lorenz curve stratification is equally efficient as the optimum stratification method for the case of two strata. Further investigation of Table 5.A. reveals that the Lorenz curve stratification under proportional allocation is the least efficient among the three types of allocation. Moreover, the Lorenz curve stratification with equal allocation resulted in a higher relative precision for three and four strata.

Table 5.B. suggests that even for moderately skewed distribution ($\alpha_3 = 0.566$), Lorenz curve stratification is approximately as efficient (99% as efficient) as the optimum stratification method under Neyman allocation with two strata. It is also evident that Neyman allocation is the best type of allocation and proportional allocation is the most inferior sample allocation among the three types of allocation.

Examination of Tables 5C and 5D confirms the suggestion made earlier that adjustment for skewness should be effected in order to improve this type of stratification. This also strengthens the need for end corrections when the number of strata is greater than two.

3.3. *On the Possible Improvement of Stratification by Equipartition of the Area of Concentration*

A distinctive feature of stratification by equal partitioning of the area of concentration is that the set of stratification points obtained thru this method are consistently smaller in magnitude than the optimum stratification points on positively skewed distributions. In fact, for a mildly skewed distribution, $f(x) = 2(1 - x), O \leqq x \leqq 1$, the deviation is greater at the upper strata compared to the disparity at the bottom strata. Also, for the case of two strata, the variation from optimum increases as a_3, the relative measure of skewness, increases. This observation suggest that it might be possible to relative measure of skewness in the improvement of the Lorenz curve stratification method.

3.3.1. *Adjustment for Skewness*

Table 6 shows that the ratio of the area corresponding to the optimum stratification points to the total area of concentration increases as the distribution becomes more skewed. A closer examination of the table reveals that for $L = 2$, the difference between the ratios secured for the optimum and Lorenz curve stratification is highly correlated with the relative measure of skewness. In fact, if we divide these differences the corresponding relative measure of skewness, the quotients obtained are nearly invariant;

Distribution	Quotient
$f(x) = 2(1 - x), O \leqq x \leqq 1$	0.126
$f(x) = x e^{-x}, x \geqq O$	0.127
$f(x) = e^{-x}, x \geqq O$	

TABLE 6. PROPORTION OF THE AREA OF CONCENTRATION, A_{xh}/A_T, CORRESPONDING TO THE STRATIFICATION POINTS

Method of Stratification	L	Distribution and Corresponding Relative Measure of Skewness											
		$f(x) = 1, 0 \leq x \leq 1$ $a_3 = 0$			$f(x) = 2(1 - x), 0 \leq x \leq 1$ $a_3 = 0.566$			$f(x) = xe^{-x}, x \geq 0$ $a_3 = 1.414$			$f(x) = e^{-x}, x \geq 0$ $a_3 = 2$		
		Stratum Code			Stratum Code			Stratum Code			Stratum Code		
		1	2	3	1	2	3	1	2	3	1	2	3
Optimum	2	0.500	—	—	0.572	—	—	0.681	—	—	0.732	—	—
(Neyman	3	0.255	0.745	—	0.325	0.812	—	0.379	0.892	—	0.449	0.918	—
Allocation)	4	0.156	0.500	0.844	0.222	0.610	0.928	0.229	0.671	0.955	0.301	0.732	0.968
	2	0.500	—	—	0.500	—	—	0.500	—	—	0.500	—	—
Lorenz Curve	3	0.333	0.667	—	0.333	0.667	—	0.333	0.667	—	0.333	0.667	—
	4	0.750	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750

Since the quotients obtained are not significantly different, one may take the average and use it as a constant multiplier of α_3, the resulting product of which should be added to h/L as an adjustment for skewness. That is, the corresponding stratification point for the two strata case is derived not by halving the area of concentration but rather by some other proportion related to the relative measure of skewness defined by

$$\frac{A_{x_1}}{A_T} = \tfrac{1}{2} + 0.12\,\alpha_3$$

3.3.2. End Corrections

The Lorenz curve stratification for L = 3 and 4 tends to deviate from the optimum even for the simple rectangular distribution. Further, this method tends to magnify the departure from optimum as the distribution becomes more skewed. It is therefore essential that refinements be applied to render the Lorenz curve stratification method adequate foı highly skewed distributions.

The improvement on Lorenz curve stratification by equipartition of the area of concentration calls for, aside from adjustment for skewness, end corrections when the number of strata is greater than two. Table 6 reveals that the optimum stratification points for Neyman allocation may be approximated by unequal partitioning of the area of concentration. Furthermore, it might be possible to relate the relative measure of skewness to the end corrections since the proportion of the area of concentration corresponding to the optimum stratification points for Neyman allocation increases as α_3 increases, as shown in the table. Several attempts to relate the stratum, the number of strata and the relative measure of skewness empirically to the necessary end correction were made. The following addition proves to be the best.

$$\frac{0.193\ L\ (16 - L)}{100} + \frac{0.422\ [L - (h + 1)]}{1000(1.707 - \alpha_3)^2} - (h - 1)\ \frac{\alpha_3^2}{100}\left[(1.5)^{h-2} + (4 - L)\ [1.64 - (\alpha_3 - 2.217)^2]\right]$$

The final formula defining the proportion of the area of concentration corresponding to the appropriate stratification points for L = 2, 3 and 4 is therefore

$$\frac{A_{x_h}}{A_T} = \frac{h}{L} + 0.12\, a_3 + \left(h - \frac{L}{2}\right)(5 - L)\left[\frac{0.193\, L\, (16 - L)}{100}\right.$$

$$+ \frac{0.422\, [L - (h + 1)]}{100(1.707 - a_3)^2} - (h - 1)\, \frac{a_3^2}{100}\, [(1.5)^{h-2}$$

$$\left. + (4 - L)\, [1.64 - a_3 - 2.217)^2]\right]$$

3.4. Final Remarks on Lorenz Curve Stratification

The preceeding investigation discloses the possible use of the Lorenz curve in the determination of strata boundaries. The first method suggested (i.e., equipartition of the area of concentration) is limited only to moderately skewed distributions as departure from optimum stratification is observed for highly skewed distributions.

Adjustment for skewness and end corrections has brought about an improved Lorenz curve stratification applicable even for highly skewed distributions. In fact, examination of Table 7 shows that the improved Lorenz curve stratification is almost as efficient as the optimum stratification method under Neyman allocation. The proportion of the area of concentration secured by the improved method is nearly equal to that of the optimum method. They differ by only a trivial amount. Consequently, the same set of stratification points are obtained.

Finally, it should be noted that this method of stratification does not consider the type of sample allocation. However, the adjustment for skewness and end corrections were worked out based on Neyman allocation. This suggests that it is possible to have another set of adjustment factors for equal allocation and one more for proportional allocation. It might also be possible to incorporate the type of allocation in the correction factors. No attempt was made, however, inasmuch as the aim of this study is to find out if it is possible to use the Lorenz curve of the stratification variable as a means of determining the appropriate strata buondaries.

TABLE 7. COMPARISON OF THE STRATIFICATION POINTS OBTAINED BY THE IMPROVED LORENZ CURVE STRATIFICATION METHOD WITH THE OPTIMUM STRATIFICATION POINTS FOR NEYMAN ALLOCATION

DISTRIBUTION	Optimum Stratification Method for Neyman Allocation						Improved Lorenz Curve Stratification					
	L = 2	L = 3		L = 4			L = 2	L = 3		L = 4		
	x_1	x_1	x_2	x_1	x_2	x_3	x_1	x_1	x_2	x_1	x_2	x_3
1. $f(x) = 1, 0 \leq x \leq 1$	0.50	0.33	0.67	0.25	0.50	0.75	0.50	0.33	0.67	0.25	0.50	0.75
2. $f(x) = 2(1 - x), 0 \leq x \leq 1$	0.35	0.23	0.50	0.18	0.37	0.62	0.35	0.23	0.50	0.18	0.35	0.59
3. $f(x) = e^{-x}, x \geq 0$	1.30	0.76	2.07	0.55	1.30	2.64	1.32	0.76	2.07	0.55	1.32	2.63
4. $f(x) = x e^{-x}, x \geq 0$	2.32	1.56	3.28	1.22	2.29	3.96	2.29	1.56	3.28	1.22	2.29	3.95

BIBLIOGRAPHY

1. Cochran, W.G. *Sampling Techniques.* 2nd ed. New York: John Wiley and Sons, Inc., 1968.

2. Cochran, W.G. "Comparison of Methods for Determining Stratum Boundaries." *Bulletin of the International Statistical Institute,* 38(2), Tokyo (1961), pp. 345-358.

3. Dalenius, T. *Sampling in Sweden.* Stockholm: Almqvist and Wichksell, 1957.

4. Dalenius, T., and Hodges, J.L., Jr. "Minimum Variance Stratification," *Journal of the American Statistical Association,* 54 (1959) pp. 88-101.

5. Durbin, J. "Review of 'Sampling in Sweden,'" *Journal of the Royal Statistical Society,* A. 122 (1959), pp. 246-248.

6. Hansen, M.L., Hurwitz, W.N., and Madow, W.G. *Sample Survey Methods and Theory,* Vol. 1, New York: John Wiley and Sons, Inc., 1958.

7. Kendall, M. *The Advanced Theory of Statistics,* Vol. 1, 4th ed. London: Charles Griffin and Company Limited, 42 Drury Lane, 1948.

8. Mood, A., and Graybill, F. *Introduction to the Theory of Statistics.* 2nd ed. New York: McGraw-Hill Book Company, Inc., 1963.

9. Raj, Des. "On Forming Strata of Equal Aggregate Size," *Journal of the American Statistical Association,* 59 (1964) pp. 481-486.

10. Sethi, V.K. "A Note on Optimum Stratification of Populations for Estimating the Population Means," *The Australian Journal of Statistics,* 5 (1963), pp. 20-23.