NUMBER 8

COEFFICIENTS OF BETWEEN-GROUP INEQUALITY: A REVIEW

Bert Kestenbaum

Division of OASDI Statistics

MAY 1979

Social Security Administration
Office of Policy
Office of Research and Statistics

Coefficients of Between-Group Inequality:
A Review
by

Bert Kestenbaum
Division of OASDI Statistics

The quest for suitable indices to summarize the inequality between two groups has lagged behind the effort to obtain summary coefficients of within-group inequality. Numerous measures of within-group inequality were proposed, and their merits and shortcomings debated, by Gini, Pareto, Bowley, Dalton and other economists and statisticians of the late nineteenth and early twentieth centuries; so that by 1933 Yntema [1] could write his review paper, "Measures of the Inequality in the Personal Distribution of Wealth or Income." Yet, apparently, at the same time, there was little exploration of alternative indices to the ratio-of-medians and ratio-of-means for measuring differences between groups, despite the limitations of these two indices, particularly the former's preoccupation with the 50th percentile 1/ and the latter's sensitivity to the skewness of the underlying distributions.

Let x be a variate distributed in different quantities to members of a group or population; and let F(x) be the cumulative distribution function of x. An index or coefficient of within-group inequality is, as Woytinsky [3:1] says, a single number that "... must refer to some specific feature of the frequency distribution which is in alignment with the connotation of 'inequality'." Although any summarization discards much of the informational content of F(x), hopefully it yields a useful analytic tool to assess changes over time in within-group inequality or to compare the within-group inequality of several populations.

Clearly, the same benefit is to be had from summarization of the difference between two groups. Accordingly, let G(x) for population B be the analog of F(x) for population A. Then, if F=G, the populations are equal. Otherwise, it would be useful to summarize the extent of difference with some function H (F,G) which is, again, in alignment with the connotation of inequality.

In recent years a number of coefficients of between-group inequality have surfaced, and we believe that the time is right for a review paper. Four types of summary measures are discussed here: (1) the index of differentiation or dissimilarity, (2) the modified Gini-Lorenz index, (3) the modified relative mean difference, and (4) the rank statistic, U.

If an index of between-group inequality is truly to be in alignment with the connotation of "inequality," in addition to intituitive appeal, we ask of it, as Lorenz [4] did in discrediting certain proposed measures

---

1/ An unusually vivid illustration of the failure of the ratio-of-medians measure to reflect a significant change from 1945 to 1952 in the relationship of nonwhite-to-white family income is given by Wohlstetter and Coleman [2:47].

of within-group inequality, that it be invariant under a change of scale. Although Henderson [5:88] has argued that people spend dollars, rather than percentages, the majority opinion is that no change in equality takes place when the income of each unit in both A and B is multiplied by some constant, k. We note that each of the two traditional measures of between-group inequality, i.e., the ratio of medians and the ratio of means, has this invariance property.

Continuing to borrow from the rich literature on indices of within-group inequality, we can draw up a list of characteristics of "good" indices. Bowley [6] urged that indices of inequality be (a) sensitive to change and (b) easy to compute. Accordingly, we will be concerned with, for each of our indices, the process by which a change in F or G is registered in the index, and with ways to simplify computation whenever necessary.

Yntema [1:424] proposed that an index have a finite range of values, and that it be adaptable to the limitations of the existing data. In particular, we often face the limitation of not having access to the individual records, and having instead only published data in the form of class (interval) frequencies.

Another desirable feature in an index of inequality is that the index not be affected much by a small number of extreme values. Another is that the index have a graphical interpretation. In Appendix A we will summarize how well each of the coefficients of between-group inequality under discussion meet these criteria.

1. The index of differentiation or dissimilarity

The Bureau of the Census' Technical Paper No. 22 [7], issued in April 1970, describes in detail an "index of differentiation" which the Bureau had begun using two years earlier to summarize the gap between two groups. By December 1971, the index had already found its way into a Monthly Labor Review article on the difference between black and white earnings [8]. 2/ An article in Social Forces by Palmore and Whittington [9] in September 1970 was the first of a number of papers in sociological journals 3/ using this index to measure differences between groups in

---

2/ In comparing the gap between populations A and B with the gap between populations C and D, this article found the former gap larger when measured with the ratio-of-medians or ratio-of-means, but smaller when measured with the index of differentiation--a finding it strived unsuccessfully to account for. The explanation, in fact, only requires a good understanding of the index, as we demonstrate in Appendix B.

3/ For example, Palmore and Whittington [10] in 1971, Fox and Faine [11] in 1973, and Shin [12] in 1976.

matters such as income, education, and occupation. Unlike the government publications, which credit the index to a 1966 study by Yoram Ben-Porath [13], the journal articles recognized the earlier contributions of the Duncans in 1955 [14,15], in developing what the latters called the "index of dissimilarity."

The index of differentiation or dissimilarity is computed from the relative class frequencies of the two cumulative distribution functions. If $s_i = \{F(x_i) - F(x_{i-1})\}$ and $t_i = \{G(x_i) - G(x_{i-1})\}$ are the proportions of populations A and B, respectively, in the i^{th} class, then the index, D, is defined as

$$D = 1/2 \, \Sigma \mid (s_i - t_i) - 0 \mid = 1/2 \, \Sigma \mid s_i - t_i \mid,$$

where the "1/2" is a scale factor added so that the maximum value of the index, in the case of total differentiation, will be equal to 1. When the populations are equal in the sense that F=G, then $s_i = t_i$ for all i; thus the term $\{(s_i - t_i) - 0\}$ represents the departure from equality in the i^{th} frequency class. The graphical expression of the index of differentiation is obvious from its definition.

Technical Paper No. 22 also presents an alternative summary measure based on a comparison of relative class frequences, but one which uses weighted ratios, rather than unweighted differences, namely,

$$\Sigma \, s_i \left| \frac{s_i}{t_i} - 1 \right| .$$

The term, $\frac{s_i}{t_i} - 1$, represents the departure from equality in the i^{th} frequency class; the need for weights, s_i, can be appreciated from the following argument. Suppose the i^{th} class is divided, so that the proportions of populations A and B in one subclass are rs_i and rt_i, respectively, and, in the other subclass, $(1-r)s_i$ and $(1-r)t_i$, respectively. Then, if there were no weighting, each subclass would contribute to the index the identical amount,

$$\left| \frac{rs_i}{rt_i} - 1 \right| = \left| \frac{(1-r)s_i}{(1-r)t_i} - 1 \right| = \left| \frac{s_i}{t_i} - 1 \right| .$$

Because this alternative measure is less intuitive than the index of differentiation, and because it has no finite maximum value and no graphical representation, we will address our remarks to the index of differentiation, although much of what we say is relevant to this alternative measure, as well.

The index of differentiation is not unique; its value depends on the choice of interval structure. For example, the value of the index when used to measure the gap between men and women in the occupations they hold is less if one difference $| s_i - t_i |$ is computed for the entire category of "professional, technical, and kindred workers" than if differences are computed for each of its subcategories--engineers, physicians, elementary school and kindergarten teachers, nurses, etc.

A more detailed illustration can be given with data taken from Technical Note No. 22:

> The percentages of black families with incomes of \$5,000-\$5,999 and \$6,000-\$6,999 in 1967 were 10.2% and 8.0%, respectively. The corresponding percentages for white families were 7.6% and 8.4%. The contribution over these two intervals to the index of differentiation is half of $\{ |.102-.076| + |.080-.084| \} = .015$. Compare this to the contribution to the index over a single interval from \$5,000 to \$6,999 of $1/2 \{ |(.102+.080)-(.076+.084)| \} = .011$.

The dependence of the index of differentiation on the interval structure follows, of course, from the basic theorem that the absolute value of a sum of terms equals the sum of the absolute values of each term only when the terms are of the same sign. It will be noted that as the interval structure becomes finer and finer, the index approaches its maximum value of 1 even for quite similar populations.

The non-uniqueness of the index of differentiation makes difficult deliberation on whether or not the index possesses the important property of variance under a change in scale. We do note, however, that, if k is the constant of scale, the value of the index will not change if the old interval structure with endpoints of ...,K,L,M,N,... is replaced by a structure with endpoints of ...,kK,kL,kM,kN,....

The distributions of black and white family income which constitute the subject matter of Technical Note No. 22 are illustrative of a fairly common situation wherein the proportion of one population exceeds the corresponding proportion of the other in all of the first m frequency classes, and the reverse is true in all of the remaining (n-m) classes. 4/ Under these circumstances, only movement between the first m and remaining (n-m) classes would be registered by the index, but the index would not be sensitive to movement within the first m or within the remaining (n-m) intervals.

4/ This is always the situation when the index of differentiation is used as a measure of segregation, as the Duncans [14] do, because the frequency classes are ordered according to their composition.

2. The modified Gini-Lorenz index

In 1905, Max Lorenz [4] proposed a graphical expression for within-group inequality. Let V(x) denote the fraction of the total aggregate belonging to members of the population with no more than x units; and let F(x) be, as before, the relative cumulative distribution function for x. Then the locus of points (F(x), V(x)) is the parametric representation of the Lorenz curve. Lorenz suggested that the inequality within a population could be appreciated with this curve if the diagonal, i.e., the "line of equality" emanating from the origin with slope of 1, is the frame of reference.

The Lorenz curve, by its definition, always lies below the diagonal, and has several other interesting properties. It is convex: the second derivative is always positive [16:49]. The first derivative (slope) at any x equals the quotient of x by its mean [17:48]; in particular, its slope at the mean equals 1. The maximum vertical distance between the Lorenz curve and the diagonal is at the mean of x [18:307].

In 1914, Corrado Gini suggested an "index of concentration" to quantify the graphical expression of inequality embodied in the Lorenz curve. The Gini coefficient, G, is defined as the quotient of the area between the Lorenz curve and the diagonal by the area below the diagonal. It has the property of invariance under a change of scale, 5/ which Lorenz had insisted on.

The Gini index of concentration has evolved as the most popular summary measure of within-group inequality. It is easy to show [20:274-6] that the index can be computed from:

$$G = \Sigma \{F(x_i) V(x_{i+1}) - F(x_{i+1}) V(x_i)\}$$

If the data are presented as class frequencies, then the x_i in the above formula are the class endpoints, and the value obtained for G is a slight underestimate--for which a "grouping correction" is available [21:10-11].

The index of concentration was not the first measure proposed by Gini to measure within-group inequality. The earlier measure, developed in 1912, and known as the "relative mean difference," is defined, in the discrete case, as:

$$\frac{1}{N \cdot N} \cdot \Sigma \Sigma \mid x_k - x_j \mid / \bar{x} ,$$

where $\bar{x}$ denotes the mean value in the population. This measure, expressing

---

5/ This is one of the properties of the Gini index of concentration demonstrated in [19].

inequality as an average of differences between individuals, has a very strong intuitive appeal. Indeed, for Gini, the attraction of the index of concentration rested upon a relationship to the earlier measure, as given in the following theorem which has been called "remarkable" by both Dalton [22:354] and Woytinsky [3:11]: the index of concentration is equal to one-half the relative mean difference. A proof for the continuous case is given by Kendall and Stuart [16:49]. For the discrete case, Woytinsky [3:251-3] presents a proof by geometry, and it is demonstrated algebraically in a number of ways--by von Bortkiewicz [23:196-7], by Theil [24:121-3], and in Appendix C of this paper.

In a 1971 paper, James Sweet [25:2] undertakes to summarize the inequality of the income distributions of blacks and whites with a modification of the Gini-Lorenz index. As he describes it, "...rather than comparing the cumulative distributions of families and money income, we are comparing the cumulative distributions of black and white families ordered with respect to incomes." In our notation, the modified Lorenz curve is the locus of points (F(x),G(x)). The modified index is the quotient of the area between the modified Lorenz curve and the diagonal by the area below the diagonal. The index can be computed from:

$$\Sigma \{ F(x_i)\ G(x_{i+1}) - F(x_{i+1})\ G(x_i) \}.$$

An alternate form, which emphasizes the contribution of each class to the index, is:

$$\Sigma \ \frac{F(x_i)}{G(x_i)} \cdot \frac{t_{i+1}}{s_{i+1}} - 1 \quad .$$

Clearly, the index equals zero when the distributions are identical and equals one when there is no overlap.

We note that Duncan's [26, 27:61-63, 28:70-71] index of industrialization, as well as segregation index #3 proposed by Jahn, Schmid, and Schrag [29:298], 6/ both are examples of the modified Gini-Lorenz index.

Nevertheless, the device of adapting the Lorenz-Gini technique to measure between-group inequality in terms of how far removed from a "line of equality" is the locus of paired values from cumulative distribution functions, represents a substantial departure from the standard after which it was modelled. In its most general sense, the curve may even cross over the diagonal; and this is, in fact, the case with the index of industrialization. Even if we restrict ourselves to situations where $F \gtrless G$, there is no guarantee--except in special cases such as the measurement of segregation, in which the population is ordered according to its composition--that the curve is even convex. Most importantly, the simple, remarkable relationship between the index and the relative mean difference measure is, of course, gone.

6/ See the Duncans' [14] discussion of this index.

The modified relative mean difference

A most intuitive approach to the measurement of inequality is to compute an average of pairwise differences. Hence Gini's recommendation of his relative mean difference coefficient to measure within-group inequality. The need to store all the data in order to compute all pairwise differences, as required for the relative mean difference measure, can be circumvented by resorting to alternative formulae for the relative mean difference, such as Kendall and Stuart's [16:50-51],

$$\frac{2}{N \cdot N} \quad \Sigma\ (i)(N-i)(x_{i+1} - x_i) \quad ,$$

where the x_i are arranged in ascending order. Other formulae are given by Wold [19:48] and Mendershausen [30:162].

The relative mean difference measure of within-group inequality is inherently an average of absolute differences. In fact, the sum of all signed differences is zero. On the other hand, if we were to seek a parallel to the relative mean difference to use for measuring the gap between group A, composed of M values $\{x_i\}$, and group B, consisting of N values $\{y_j\}$, we would think in terms of signed differences, $x_i - y_j$, and obtain

$$\{1 \div ((\Sigma^{M} x + \Sigma^{N} y)\ /(M + N))\} \cdot \{\Sigma\Sigma_{ji}^{NM}\ (x_i - y_j)\ /\ (M \cdot N)\}$$

However, we have:

$$\Sigma\Sigma_{ji}^{NM}\ (x_i - y_j) = \Sigma_j^{N}\{\ \Sigma_i^{M}\ x_i - \Sigma_i^{M}\ y_j\}$$

$$= \Sigma_j^{N}\ (M\bar{x} - My_j)$$

$$= \Sigma_j^{N}\ M\bar{x} - \Sigma_j^{N}\ My_j$$

$$= NM\bar{x} - MN\bar{y} = MN(\bar{x} - \bar{y}).$$

Thus our proposed measure is mothing more than the quotient of $[MN\ (\bar{x}-\bar{y})\ /\ MN]$, or $(\bar{x}-\bar{y})$, by the overall mean, $(M\bar{x} + N\bar{y})/(M+N)$.

A measure of between-group inequality utilizing an average of pairwise differences was also proposed by Joseph Gastwirth [31] in 1973, and it is best described by comparison to the index presented above. Our method consists of these steps:

(i) computing a difference for each pair (x_i, y_j);

(ii) weighting each difference by the reciprocal of M•N;

(iii) standardizing each weighted difference by the reciprocal of the overall mean;

(iv) cumulating those standardized weighted differences for which $x_i \geq y_j$ into one sum, and cumulating those standardized weighted differences for which $x_i < y_j$ into another sum; and,

(v) defining the index as the difference of the two sums.

Gastwirth's method is different in two ways. First, it replaces the overall mean in step (iii) with the average of x_i and y_j. Then, in step (v), it defines the index as the quotient of the first sum by the sum of the two sums.

While the dissimilarity of the two methods in combining sums in step (v) to yield an index is unimportant, the different ways of standardizing deserve comment. Our strategy was to design an index of between-group inequality analogous to the relative mean difference measure of within-group inequality; and the latter standardizes with respect to the population mean. In his paper, Gastwirth, however, proposes a variant of the relative mean difference coefficient itself, recommending that within-group inequality be measured by,

$$\frac{1}{N \cdot N} \quad \Sigma\Sigma \quad \frac{|x_k - x_j|}{1/2\ (x_k + x_j)} \ ,$$

rather than by,

$$\frac{1}{N \cdot N} \quad \Sigma\Sigma \quad | x_k - x_j | / \bar{x} \ .$$

Gastwirth's rationale for replacing the relative mean difference measure, despite its singular relationship to the Lorenz curve's area of concentration, with an index devoid of geometric interpretation, is that, "the standard measure, the Gini index, does not allow the overall inequality to be split into parts such as the differences between blacks and whites." The fact is, however, that, as Soltow [32] has demonstrated, the Gini index does admit such a decomposition into the components of inequality--though not as readily. 7/

7/ Soltow's paper is an interesting assessment of the response of the Gini index of personal income inequality to changes in the age, educational, and occupational structure of the population. The relative mean difference is decomposed into three components, as follows:

$$\frac{W_A}{M \cdot M} \overset{MM}{\Sigma\Sigma} |x_k - x_j| / \bar{x} + \frac{W_B}{N \cdot N} \overset{NN}{\Sigma\Sigma} |y_k - y_j| / \bar{y} + \frac{W_C}{M \cdot N} \overset{NM}{\Sigma\Sigma} |x_k - y_j| / u$$

where,

u = the population mean, $(M\bar{x} + N\bar{y}) / (M+N)$;

$W_A = \bar{x}\ MM/u(M+N)(M+N)$;

$W_B = \bar{y}\ NN/u(M+N)(M+N)$; and

$W_C = M \cdot N / (M+N)(M+N)$.

A better recommendation for Gastwirth's innovation, relevant to both the measurement of within-group inequality and the measurement of between-group inequality, is that the resultant index is less affected by a small number of very large values. Moreover, the innovation responds to the reservations expressed by some to the appropriateness of the Gini index for measuring income inequality, given that the index attaches equal weights to equal differences in income with no regard for the size of the incomes being differenced. As Kravis [33:179] has articulated:

> There are, however, both economic and statistical grounds for casting the analysis in terms of percentage rather than absolute differences in income. Equal absolute differences in income of say $100 may have very different significance if one difference is found around the $1,000 level and the other around the $10,000 level. Equal percentage differences in income at two different points on the income scale are apt, we believe, to have more nearly the same welfare significance than equal absolute differences.

Gastwirth's index can be used to approximate between-group inequality when the data are in the form of class frequencies, if it is assumed that all units in a class have the value of the midpoint (or some other point) in the class. The computation of the index from individual records appears to be a formidable task, unless some alternative formula is discovered which avoids the need for storing all the data.

The rank statistic, U

Considering the recent popularity of nonparametric methods, it is not surprising to find that a rank statistic be recommended as a coefficient of between-group inequality. The statistic, U/MN, where M and N are, as before, the sizes of populations A and B, respectively, was proposed by Gastwirth [34] in a 1973 paper.

Often called the Mann-Whitney statistic, after the statisticians who introduced it in 1947 [35], and used primarily for testing the null hypothesis of identical distributions, the U-statistic is defined as the number of times a member of population B precedes a member of populations A in a combined, ordered arrangement. Then the fraction, U/MN, gives the probability that a randomly selected member of population B precedes a randomly selected member of population A; so that small values of U/MN are indicative of a concentration of members of A in the lower ranks, while large values indicate a large concentration of members of B there. Gastwirth gives the analog of U/MN, for the continuous case, as $\int G(x)f(x)dx$.

The U-statistic actually is a linear function of Wilcoxon's [36] "sum of ranks" statistic, proposed two years earlier; thus a simple way for computing U. The precise relationship is:

$$U = MN + 1/2\ N\ (N+1) - T,$$

where T is the sum of the ranks occupied by members of B.

Since it is easy to compute, it registers changes in F or G well, and, as a rank statistic, it is not sensitive to the skewness of distributions, the U-statistic has much to recommend it as an index of between-group inequality. If the data are given as class frequencies, an approximation can be obtained by assigning all members of the class the mean of the ranks which they jointly occupy and proceeding to calculate T.

On the other hand, the U-statistic is less appealing intuitively than indices which take account of the magnitude of differences between members of A and members of B. Also, the U-statistic has no evident graphical expression.

Appendix A: A summary comparison of indices of between-group inequality

The table below summarizes the strengths and weaknesses of several indices of between-group inequality. A check mark (✓) indicates that a given index possesses a particular desirable property. No attempt is made to distinguish between indices in the degree to which they possess some characteristic; the categorization is dichotomous. There is, admittedly, an element of subjectivity in this characterization.

Seven indices are considered, as follows:

- I1 = the index of differentiation or dissimilarity;
- I2 = the modified Gini-Lorenz index;
- I3 = the modified relative mean difference;
- I4 = the modified relative mean difference, with Gastwirth's innovation;
- I5 = the U-statistic;
- I6 = the ratio of medians; and
- I7 = the ratio of means.

All these have been discussed in the paper, and all are intuitively in alignment with the connotation of between-group inequality.

Desired properties	Index						
	I1	I2	I3	I4	I5	I6	I7
Very intuitively appealing			✓	✓			✓
Invariant under change in scale	✓ a/	✓	✓	✓	✓	✓	✓
Quite sensitive to change		✓	✓	✓	✓		✓
Easy to compute	✓	✓	✓		✓	✓	✓
With a finite range	✓	✓	✓	✓	✓	✓	✓
Serviceable even if data are grouped	✓ b/	✓	✓	✓	✓	✓	✓
Resistant to effects of positive skewness		✓		✓	✓	✓	
Possessing a geometric interpretation	✓	✓					

a/ This is qualified in the text.

b/ Index I1 can be used only if the data are grouped.

Appendix B: Comparing the black-white male indices of earnings differentiation for all and 4-quarter workers

A Monthly Labor Review study of the earnings gap between black men and white men in 1966 [8] finds that, when restricted to workers with earnings in each calendar quarter, the gap is smaller than for all workers if measured with the ratio of medians or means, but, surprisingly, larger if measured with the index of differentiation. It observes, "Contrary to expectation, the distributional differences between black and white workers with earnings in each quarter of the year is greater than the difference which exists between the earnings of all workers."

It goes on to make this analysis: "Available data sources shed only limited light on the causes behind the differences in the areas of overlap among all and four quarter workers. It seems clear, however, that part of the explanation must be that a substantially greater proportion of blacks than of whites do not have earnings in all four quarters of the year and that there are substantial differences in the occupations of blacks and whites...." We demonstrate here that the explanation is simple if the index of differentiation is well understood; and that, in fact, if there were a substantially greater proportion of whites than of blacks who worked part-year, the excess of the 4-quarter worker index over the all-worker index would be even greater.

As is often the situation with such subject matter, the index of differentiation is here nothing more than the difference between the proportions of the two groups falling in the lower-order frequency classes. Using the data in table 1 of the Monthly Labor Review study, we find the index to be the difference in the proportions of black men and white men with earnings below $6,000--which for 4-quarter workers is $\{ .738 - .378 \} = .360$.

Now, for part-year workers, the proportions of black and white men with earnings under $6,000 were .991 and .978, respectively. The proportions among all workers who work part-year were .343 for blacks and .267 for whites.

When including part-year workers with 4-quarter workers, the fraction of white men with earnings below $6,000 swells to:

$$\{ (.378)(.733)+(.978)(.267) \} = .538.$$

For black men, because the proportion of 4-quarter workers with earnings below $6,000 was already quite high, the inclusion of part-year workers raises it only modestly, to

$$\{ (.738)(.657)+(.991)(.343) \} = .825.$$

The resulting index of differentiation for all workers therefore was only $\{.825 - .538\} = .287$, despite the fact that black men were more often part-year workers than whites.

Appendix C: An algebraic proof of Gini's theorem

The theorem states: the quotient of the area (A) between the Lorenz curve given by (F(x),V(x)) and the diagonal by the area below the diagonal (i.e., 1/2) is equal to one-half the relative mean difference of the N values $\{x_i\}$. If we let d denote any difference between a pair of values; and if we arrange the values in ascending order and define $X_m = \sum^{m} x_i$, then the assertion of the theorem is:

$$A \div 1/2 \;=\; 1/2 \left\{ \frac{\Sigma|d|}{N \cdot N} \div \frac{X_N}{N} \right\}, \text{ or}$$

$$(*) \quad 4A \cdot X_N \cdot N \;=\; \Sigma\, |d|$$

Now, the area A equals the difference between the area below the diagonal (=1/2) and the area B below the Lorenz curve. Using approximating trapezoids, the latter area is: $1/2 \cdot 1/N \cdot \{X_o/X_N + 2X_1/X_N + \ldots + 2X_{N-1}/X_N + X_N/X_N\}$.

Therefore,

$$(**) \quad 4A \cdot X_N \cdot N = 2 \cdot (N-1)X_N - 4\,\{X_1 + X_2 + \ldots + X_{N-1}\} \quad .$$

The method of proof consists of observing that the numbers of times each value x_i appears on the left hand side and the right hand side of equation (*) are the same.

Because $X_m = \sum^{m} x_i$, we see from (**) that x_i appears on the left-hand side of (*) exactly $\{2 \cdot (N-1) - 4(N-i)\} \;=\; \{4i - 2(N+1)\}$ times.

As for the right hand side of (*), we have:

$$\Sigma|d| \;=\; \sum_k \left\{ \sum_{j=1}^{k} (x_k - x_j) + \sum_{j=k+1}^{N} (x_j - x_k) \right\}$$

$$= \sum_k \left\{ (kx_k - X_k) + (X_N - X_k - (N-k)x_k) \right\}$$

$$= \sum_k \left\{ (2k-N)x_k + X_N - 2X_k \right\} \; .$$

Summing over k, we find that x_i appears exactly $\{(2i-N)+N-2(N-i+1)\} = \{4i-2(N+1)\}$ times.

Q.E.D.

BIBLIOGRAPHY

[1]Yntema, Dwight B., "Measures of the Inequality in the Personal Distribution of Wealth or Income," *Journal of the American Statistical Association*, December 1933, pp. 423-433.

[2]Wohlstetter, Albert and Sinclair Coleman, *Race Differences in Income*, Santa Monica (Calif.), Rand Corporation, October 1970.

[3]Woytinsky, W.S., *Earnings and Social Security in the United States*, Washington, D.C., Committee on Social Security, Social Science Research Council, 1943.

[4]Lorenz, Max O., "Methods of Measuring the Concentration of Wealth," *Publications of the American Statistical Association*, New Series, No. 70, June 1905, pp. 209-219.

[5]Henderson, Vivian, "Regions, Race, and Jobs," in Ross, A. and H. Hill, ed., *Employment, Race, and Poverty*, Harcourt, Brace, & World, 1967.

[6]Bowley, A.L., *Elements of Statistics*, 4th edition, London, P.S. King and Son, 1920.

[7]U.S. Bureau of the Census, *Measures of Overlap of Income Distributions of White and Negro Families in the United States*, Technical Paper No. 22, Washington, D.C., Government Printing Office, April 1970.

[8]Strasser, Arnold, "Differentials and Overlaps in Annual Earnings of Blacks and Whites," *Monthly Labor Review*, December 1971, pp. 16-26.

[9]Palmore, Erdman and Frank J. Whittington, "Differential Trends Toward Equality Between Whites and Nonwhites," *Social Forces*, September 1970, pp. 108-116.

[10]Palmore, Erdman and Frank Whittington, "Trends in the Relative Status of the Aged," *Social Forces*, September 1971, pp. 84-91.

[11]Fox, William S. and John R. Faine, "Trends in White-Nonwhite Income Equality," *Sociology and Social Research*, April 1973, pp. 288-299.

[12]Shin, Eui Hang, "Earnings Inequality Between Black and White Males by Education, Occupation, and Region," *Sociology and Social Research*, January 1976, pp. 161-172.

[13]Ben-Porath, Yoram, *The Labor Force in Israel*, Jerusalem, Maurice Falk Institute of Economic Research in Israel, 1966.

[14]Duncan, Otis Dudley and Beverly Duncan, "A Methodological Analysis of Segregation Indexes," *American Sociological Review*, April 1955, pp. 210-217.

[15]Duncan, Otis Dudley and Beverly Duncan, "Residential Distribution and Occupational Stratification," *American Journal of Sociology*, July 1955, pp. 493-503.

[16]Kendall, Maurice G. and Alan Stuart, *The Advanced Theory of Statistics*, vol. 1, New York, Hafner, 1958.

[17]Bronfenbrenner, Martin, *Income Distribution Theory*, Aldine-Atherton, Chicago, 1971.

[18]Gastwirth, Joseph, "The Estimation of the Lorenz Curve and Gini Index," *Review of Economics and Statistics*, August 1972, pp.306-316.

[19]Wold, Herman, "A Study on the Mean Difference, Concentration Curves and Concentration Ratio," *Metron*, v. 12 no. 2, 1935, pp. 39-58.

[20]Miller, Herman P., *Rich Man, Poor Man*, New York, Thomas Y. Crowell, 1971.

[21]Goldsmith, S., G. Jaszi, H. Kaitz, and M. Liebenberg, "Size Distribution of Income Since the Mid-Thirties," Review of Economics and Statistics, February 1954, pp. 1-32.

[22]Dalton, Hugh, "The Measurement of the Inequality of Incomes," Economic Journal, vol. 30, 1920, pp. 348-361.

[23]von Bortkiewicz, L., "Die Disparitatsmasse Der Einkommensstatistik," Bulletin De L'Institut International De Statistique, 1931, pp. 189-291.

[24]Theil, Henri, Economics and Information Theory, Amsterdam, North-Holland, 1967.

[25]Sweet, James A., "The Employment of Wives and the Inequality of Family Incomes," Proceedings of the Social Statistics Section of the American Statistical Association, 1971, pp. 1-5.

[26]Duncan, Otis Dudley, "Urbanization and Retail Specialization," Social Forces, March 1952, pp. 267-271.

[27]Duncan, Otis Dudley and Albert J. Reiss, Jr., Social Characteristics of Urban and Rural Communities, 1950, New York, Wiley, 1956.

[28]Duncan, Otis Dudley, W. Richard Scott, Stanley Lieberson, Beverly Duncan, and Hal H. Winsborough, Metropolis and Region, Baltimore, Johns Hopkins Press, 1960.

[29]Jahn, Julius, Calvin F. Schmid, and Clarence Schrag, "The Measurement of Ecological Segregation," American Sociological Review, June 1947, 293-303.

[30]Mendershausen, Horst, Changes in Income Distribution During the Great Depression, New York, National Bureau of Economic Research: Studies in Income and Wealth, vol. 7, 1946.

[31]Gastwirth, Joseph, "A New Index of Income Inequality," Bulletin of the International Statistical Institute: Proceedings of the 39th Session, 1973, pp. 437-441.

[32]Soltow, Lee, "The Distribution of Income Related to Changes in the Distribution of Education, Age, and Occupation", Review of Economics and Statistics, November 1960, pp. 450-454.

[33]Kravis, Irving B., The Structure of Income, Philadelphia, University of Pennsylvania, 1962.

[34]Gastwirth, Joseph, "Measurement of Earnings Differentials Between the Sexes," Proceedings of the Social Statistics Section of the American Statistical Association, 1973, pp. 133-137.

[35]Mann, H.B. and D.R. Whitney, "On a Test Whether One of Two Random Variables is Stochastically Larger than the Other," Annals of Mathematical Statistics, 1947, pp. 50-60.

[36]Wilcoxon, Frank, "Individual Comparisons by Ranking Methods," Biometrics Bulletin, December 1945, pp. 80-83.