CLUSTER ANALYSIS ON AGE PATTERNS OF FERTILITY

Prem P. Talwar and Peter A. Lachenbruch, International Fertility Research Program

1. Introduction

The age-specific fertility rate curve, in general, is a bell-shaped, unimodal curve which first rises slowly and then sharply in the age group 15-19, attains its modal value somewhere between ages 20-29 and then declines, first slowly and then steeply, until it approaches zero around the age of 45 years. Its two dimensions are the area and the shape. The demographic term for the former dimension, area, is the total fertility rate or the completed family size, according to whether period data or cohort data are utilized for computation of such rates. The latter dimension, namely, the shape of the curve, may be termed the fertility pattern or the pattern of reproduction or the age pattern of fertility.

The shape of the age-specific fertility rate curve (subsequently referred to as ASFR) has several useful applications in demographic analysis (1,2,3,8). In a study of relationship between different fertility indices, total fertility rate has a one-to-one relationship with either crude birth rate or general fertility rate for a given age pattern of fertility (shape of the age-specific fertility rates). In comparing the levels of fertility at two time periods or between two or more populations, it may be necessary to control for differences in age patterns of fertility. Coale and Tye (2), in an attempt to explain differences in the levels of fertility in two ethnic groups in Singapore, had to consider fertility patterns of births. In population projections, where total number of births are needed for the future time periods, it may be essential to consider changes in the age pattern of fertility, especially in populations where fertility has shown decline.

Despite the need for considering age pattern of fertility in demographic analysis, the only detailed study devoted to this topic is one by the United Nations (10). They noted the variations in terms of two basic characteristics of the curves: the peak-fertility age group and the degree of concentration around this peak and determined some basic fertility patterns in which most of the populations belong. This study attempts to determine such basic fertility patterns by using more information from the individual curves. The goals of this study are:[1/]

(1) to investigate the extent to which fertility patterns vary among populations, and

(2) to investigate whether basic age patterns of fertility exist, and if so, what they are.

2. Sources of Data

The United Nations has compiled data on fertility rates specific for quinquennial age groups for 73 populations in and around the year 1960. This study uses their data (10). The UN examined the data for various inaccuracies and adjusted for most of them. For 36 low-fertility populations (TFR < 4.12)[2/], the numerator of the ASFR was derived from registration statistics on births by age of mother. It was divided by the official estimates of age-specific females population to get the fertility rates. The registration statistics were reported to be complete in these countries except for Greece where under-registration was not extensive. The 37 high-fertility populations have TFR above 4.12 and their data are of satisfactory quality for the purposes of this study. Data of doubtful accuracy were omitted.

For the present study the data on period ASFR's given in the UN publication were converted to age-specific relative fertility rates (ASRFR) by dividing them by the sum of all ASFRs (equalizing the area for each fertility curve). In other words, if $g_x\,dx$ is the area of ASFR curve between x and $x + dx$, f_x is defined as the proportion of the total fertility achieved in age x and x + dx and is obtained by making total area under the curve equal to unity. In mathematical notations

$$f_a = \frac{\int_a^{a+5} g_x\,dx}{\int_{w_0}^{w_1} g_x\,dx}$$

w_0 and w_1 are the beginning and end ages of reproductive life

which, in case the fertility rates are specific for five age groups, becomes

$$f_a = \frac{5g_a}{5\sum_1^7 g_a}$$

This transformation was done in order to investigate primarily variations in fertility patterns or age patterns of fertility. In the further description f_x will be referred to as age-specific relative fertility rates (ASRFR).

3. Variations in Individual Fertility Patterns

A large variation was observed in the individual age patterns of fertility. No two populations had exactly the same pattern; though there are some similarities in the patterns among different populations. The mean relative fertility rates for the 73 populations under study are 8.4, 26.7, 27.4, 19.5, 21.1, 4.8 and 1.0 and are different from the relative fertility rates observed in individual populations. For instance, the mean relative fertility rate for the age group 15-19,is 8.4 compared to the individual population rates which range from 1.0 for Ireland to 21.4 for Gabon. The coefficients of variantion calculated for each age group reveal that groups 15-19, 40-44 and 45-49 have relatively larger variations. The larger variation in the earlier part of the reproductive life can be attributed to variations in the patterns of marriage and in patterns and incidence of illegitimacy among countries. That in the latter part is due to several biological and socio-cultural factors which affect termination of the reproductive life.

The fertility pattern of a high-fertility country is likely to be different from that of a low-fertility country because both these groups of countries differ not only in their family size norms but also in the effects of biological and socio-cultural factors. Hence the first part of the study considered whether all 73 populations should be considered together or in sub-groups. For such study the populations were grouped into two sub-groups--those with TFR greater than 4.12 and those with less than that. This grouping was adopted in view of the U.N. observation (10) that no other social or economic indicator separates countries so well into two groups as does the TFR level of 4.12. Step-wise discriminant analysis was adopted to obtain an equation which best discriminated fertility patterns of these two groups. A posterior classification was obtained on the basis of this equation. It was found that fertility patterns of high-fertility countries are different from those for low-fertility countries. The fertility pattern of a low-fertility population typically has lower relative fertility rates at ages 45-49 and 15-19 and higher relative fertility rates at ages 25-29 and 20-24. The typical fertility pattern for a low-fertility country is 7, 29, 30, 19, 11, 3.5 and 0.5 for seven quinquennial age groups as compared to 10, 24, 24, 20, 14, 6 and 2 for a high-fertility. Since the fertility pattern for a high-fertility country is different than for a low-fertility country, the two groups were considered separately to determine basic age patterns for each.

4. Approach to the Problem

The United Nations attempted to determine basic age patterns on the basis of two characteristics of the curve: (1) modal fertility age group and (2) spread around the modal age group. We will use cluster analysis to apportion the total number of age patterns of fertility into the "best" number of homogeneous disjoint subsets(4). The average pattern of each subset is the basic age pattern for that group of countries.

Cluster analysis identifies homogeneous classes or groups in a larger set of data (7). The first step in its use is to choose a similarity index which provides a measure of closeness between two units. Such an index is calculated for all pairs of units. In the second step clustering algorithm is used to form a few clusters so that the units in one cluster are more similar than those in different clusters. A "stopping rule" is ultimately needed to terminate the clustering process. The third step in the cluster analysis applicable in our case, is to choose the "best" set of clusters if there are more than one set of clusters formed by taking more than one similarity indexes.

5. Cluster Analysis on Age Patterns of Fertility

a. Similarity Index

The indices of similarity were defined by subtracting distances between the fertility patterns of two populations from the maximum possible distance. Two types of distances between populations i and j considered in this study are:

$$d^2_{ij} = \frac{1}{7}\left[(f_i - f_j)^T(f_i - f_j)\right]$$

$$D^2_{ij} = \frac{1}{7}\left[(F_i - F_j)^T(F_i - F_j)\right]$$

where T = Transpose of the matrix
f_i = Column vector of relative fertility rates for seven quinquennial age groups for country i

These distances were converted to measures of similarity by subtracting them from

$$D^2 = \frac{100^2+100^2}{7}$$ = The maximum distance between two fertility curves.

Thus the similarity indices, for our analysis, were defined as

$$s_{ij} = D - d_{ij}$$

$$S_{ij} = D - D_{ij}$$

The first index considers the discrepancy in fertility pattern curves in each age group, while the second considers discrepancy in cumulative reproduction up to different age groups. In our analysis, we identified these similarity indices as those based on simple distance and on cumulated distance respectively. The clusters formed by these two similarity indices for low-fertility and high-fertility countries are discussed separately.

b.1 Process of Cluster Formation

The cluster formation process consisted of two parts. The first part calculated the (n x n) symmetric matrix S of similarity coefficients between every pair of units. The second part formed clusters by adopting agglomerative algorithm (4). Specifically, the cluster formation program proceeded in the following fashion:

(1) Initially, the n units were considered to constitute n clusters. Each unit was designated by the subscript i (i = 1,2 ......n)

(2) In the first step, the (n x n) symmetric similarity matrix was scanned for the maximum value of S_{ij}, i<j. We denote
$$\max_{\text{all } i<j} (S_{ij}) = S^*_{ij}$$
Then S^*_{ij} became the level of similarity at which clusters were formed at this step.

(3) In case the value of the similarity coefficient between units i and k (k = 1, 2.....n; k ≠ j) or between units j and k were tied with S^*_{ij}, then random selection decided

whether a cluster was formed by (i,j), (i,k) or (j,k) units. For this example, let us assume that the units i and j clustered together.

(4) The group which clustered $(i,j)^{th}$ units together is represented by the i^{th} unit (note: $i < j$) whose row and column elements were replaced by those of

Min (S_{ik}, S_{jk}); (k = 1,2,....n; ≠ i or j)
Elements corresponding to unit j were replaced by those of the n^{th} unit. Thus each cluster was represented by a unit in the next step of the similarity matrix whose dimension was reduced by one.

(5) The value of S_{ij}^* and various clusters and their constituent units were recorded

(6) Steps (2) to (5) were repeated until all the units were included in one cluster.

This algorithm was used to form two sets of clusters from two indexes of similarity considered here. The two indexes are the simple distance and the cumulated distance between two units. It was noted that the process of clusters formation was slower for the low-fertility counties for both these sets. This might be an indicator of more heterogeneity in the fertility patterns among the low-fertility countries because of greater control on their reproductive lives.

b. 'Stopping rules' for decision on the optimum number of clusters

The clustering process started from the stage when each unit was an individual cluster to the last stage when all units formed one cluster. Some rule was needed to decide on the optimum number of 'homogeneous' clusters. One possible approach is to use distribution theory, although fully aware that it is not applicable in the usual statistical sense. This approach looks at some statistics of within-cluster and between-cluster variations at each step of the clustering process.3/ Though these statistics are generally used to test the statistical significance, they were used here to study their changed pattern when the clustering process passes from one stage to the next. An abrupt change in the expected changing pattern can be suggestive of a stopping stage.

The statistics based on between clusters and within clusters variations considered but not pursued further were (1) those generally used in multivariate testing--largest root, trace or the likelihood criterion, and (2) multivariate outlier test statistics, because of smaller number of units per cluster compared to the number of variates (six corresponding to seven quinquennial age groups). The other statistics computed at different stages of the clustering process and whose results were used to decide on the optimum number of clusters will be described here. The purpose of looking at several statistics was to check the consistency in decision about the stopping stage since no one statistic appeared to be better than the other. Two types of measures were used: (1) statistics based on unweighted distance, and (2) statistics based on weighted distance. The word distance has been used here though analogy with univariate analysis of variance is very close. Like analysis of variance, total variance is being partitioned into within-cluster and between-clusters. The first type of measure is F-ratio test in the analysis of variance and the second measure is similar to the generalized D^2 given by Rao(6).

The measure based on unweighted distance at a step when there were g clusters is defined as

$$UD_1 = \frac{(B/(g-1))}{W/(n-g)}$$

where

$$W = \text{overall within-cluster variance} = \sum_{k=1}^{g} W_k$$

and

W_k = within-cluster variance for k^{th} cluster and is calculated as

$$\sum_{k=1}^{n_k} (\underset{\sim}{f}_i^{(k)} - \underset{\sim}{\bar{f}}^{(k)})^T (\underset{\sim}{f}_i^{(k)} - \underset{\sim}{\bar{f}}^{(k)})$$

where $f_i^{(k)}$ = column vector for relatively fertility rates for seven age groups for country i which belongs to cluster k, and

$\bar{f}^{(k)}$ = column vector for mean relative fertility rates for seven age group in cluster k.

T = transpose of the column vector, and

n_k = number of countries in cluster k

B = overall between-cluster variance calculated as

$$\sum_{k=1}^{n_g} n_k (\underset{\sim}{\bar{f}}^{(k)} - \underset{\sim}{\bar{f}}^{(.)})^T (\underset{\sim}{\bar{f}}^{(k)} - \underset{\sim}{\bar{f}}^{(.)})$$

where $\bar{f}^{(.)}$ = column vector for overall mean relative fertility rates for seven age groups.

g = number of clusters in which n countries have been grouped with $n_1, n_2 \ldots n_g$ countries in each cluster

n = total number of countries

The measure based on weighted distance at a step when there are g clusters is defined as

$$WD = \frac{WD_1}{(g-1)\ (p-1)}$$

where

$$WD_1 = \sum_{k=1}^{g} n_k\ (\bar{\underset{\sim}{f}}^{(k)} - \bar{\underset{\sim}{f}}^{(.)})^T\ (S_w^{-1})(\bar{\underset{\sim}{f}}^{(k)} - \bar{\underset{\sim}{f}}^{(.)})$$

$$\text{and } S_w = \frac{\sum_{k=1}^{g} \sum_{i=1}^{n_k} (\underset{\sim}{f}_i^{(k)} - \bar{\underset{\sim}{f}}^{(k)})(\bar{\underset{\sim}{f}}_i^{(k)} - \bar{\underset{\sim}{f}}^{(k)})^T}{n-g}$$

P = number of independent variates = 6 in our case

Obviously, the larger values of $\underline{UD_1}$ or $\underline{WD}$ would mean larger distance between clusters and smaller distances within clusters. Thus larger values of these indices are preferred for the choice of stopping stage.

These two measures were calculated at each stage of the clustering process. It may be noticed that both these measures consider the number of clusters at each stage which is essential since these measures are calculated at different stages of the clustering process when there are different numbers of clusters. Tables 1 and 2 give values of $\underline{UD_1}$ and $\underline{WD}$ at different stages of the clustering process.

Table 1 gives results for the case when the simple distance was taken as the similarity index and Table 2 when the cumulated distance was used. In general the values of UD_1 and WD in Tables 1 and 2 decline with the decline in the number of clusters. But our choice of the "best" number of clusters was based on two desirable properties. First,the number of clusters should not be too large and secondly, the within clusters distance should not be too large (or between clusters distance should be large)i.e. larger value of the indices were to be preferred. These two considerations suggest five clusters for the low-fertility countries from Table 1. Both UD_1 and WD suggest a great increase in the within clusters variance compared to between clusters if one considers a choice of less than five clusters (sharp decrease in UD_1 and WD). The choice has to be on eight clusters if the within clusters variance has to be further reduced. The same considerations suggest three clusters for the high-fertility countries in Table 1.

In the case of Table 2, $\underline{UD_1}$ suggests seven clusters for the developed countries but WD suggests six, although it seems that the value of $\underline{WD}$ remains stable from step 3 to 6 except for random fluctuations. Hence, the choice in this case was seven clusters for the low-fertility countries. For the high-fertility countries, the value of $\underline{UD_1}$ remains stable from steps 4 to 7, while that of $\underline{WD}$ is stable from steps 3 to 6 except for the random fluctuations. Hence, the choice was three clusters for the high-fertility countries.

In reality, there are only five clusters for the low-fertility countries based on Table 2, since Japan and Ireland have unique fertility patterns and form clusters of their own. The same argument suggests three clusters for the high-fertility countries in Table 1. The basic pattern for this cluster is obtained by averaging the fertility patterns of all populations forming that particular cluster. Although some countries belong to the same cluster on the basis of the two similarity indices considered here, there are other countries which belong to different clusters if different similarity indices are used.

c. The Choice of"best set of age patterns of fertility"

The choice of the two similarity indices lead to the two sets of age patterns of fertility both for the low-fertility and the high-fertility countries. The next decision was to make a choice of the best set out of the two available. The first set based on simple distance as a measure of similarity had five fertility patterns for the low-fertility and three for the high-fertility countries. The second set (based on cumulated distance as a matter of similarity) had seven fertility patterns for the low-fertility and three for the high fertility countries.

The choice of the "best" set was based on statistical and demographic considerations. It was desired that the basic fertility patterns should be distinct. In other words, countries within a pattern (cluster) should be more homogeneous than those in different patterns. Hence the statistical considerations were based on smaller within-clusters distance or larger between-clusters distance. The demographic considerations required that the "best" set should cover extreme shapes of the age-specific relative fertility curves. The following paragraphs discuss the results of various statistical and demographic techniques utilized for investigations into the choice of the "best" set. A method based on ranks of the pair-wise distances was also used for this purpose.

Step-wise discriminant analysis was used to study whether different fertility patterns within a set were distinct and non-overlapping. In Set I all low-fertility countries were correctly classified in the three clusters; only one of the 37 high-fertility countries was misclassified; it had 0.45 posterior probability of belonging to the correct group. For Set II, all high-fertility countries were correctly classified, but one of the low-fertility countries was misclassified with 0.37 probability of correct classification. If the number of misclassified cases and their posterior probabilities are taken as an index of clear-cut and distinct fertility patterns in a set, then Set I seems to be slightly better by this analysis.

A statistical test for outliers was used to determine whether all countries forming a cluster have come from some basic pattern. It provided a test for homogeneity of all fertility patterns in a cluster. In its application, we have assumed that the basic distributional assumptions are fulfilled by our data. Wilks' (11) test for multivariate outliers was applied to individual clusters within the two sets[4/]. In Set I, no outliers were found. In Set II, there was one outlier. If the number of outliers is taken as an index of heterogeneity of clusters, then this test, too, tends to favor Set I.

Demographically, the fertility patterns should cover the extreme shapes since they have effect on the population birth rates and the growth rates. Stable population parameters were estimated for each fertility pattern within the different sets under consideration. It is found that the fertility patterns in Set I covered the larger range of the stable population parameters.

The rank order of distances method considers all the $\binom{n}{2}$ pair-wise distances between countries. The principle underlying this technique is that two countries with shorter distances between them are more likely to belong to the same cluster. That is, pair-wise distances between units in the same cluster should be smaller than between those in different clusters. Let there be $\underline{g}$ clusters $\underline{C}_1$, $\underline{C}_2 \ldots \underline{C}_g$ with $\underline{n}_1, \underline{n}_2 \ldots \underline{n}_g$ number of units such that $n = \sum_{k=1}^{g} n_k$. The total number of pair-wise, within-cluster distances are given by

$$N = \binom{n_1}{2} + \binom{n_2}{2} + \ldots + \binom{n_g}{2}$$

where N is less than $\binom{n}{2}$ unless g = 1. Now if all $\binom{n}{2}$ distances are ranked in an array of descending order then ideally, the last N distances should be from countries which belong to same clusters. Operationally, all the $\binom{n}{2}$ distances were ranked in descending array; 'S' was marked for distance between those countries both of which were in the same cluster and 'D' for those in the different, then the per cent of S's in the last N distances were taken as an index of distinct and clear-cut grouping in clusters within a set. The ideal value of this index is 100. This index was calculated for both the sets; Table 3 shows the results for the developed and the developing countries. Set I shows the largest percentage of S's though they are far from the ideal value of 100 per cent. Thus this method also suggests Set I as the "best" choice.

Different techniques adopted in the investigation on the "best" set have generally shown results in favor of Set I.

e. Results:

The analysis discussed above shows that the basic age patterns of fertility, i.e. percent of the total fertility rates accounted for by different quinquennial age groups, for the developing countries are:

Pattern I= UP_1[5/]: 19.7, 24.1, 21.2, 16.6, 11.5, 4.8, 2.2

Pattern II= UP_2: 12.0, 25.9, 24.1, 18.4, 12.6, 5.4, 1.6

Pattern III = UP_3: 5.9, 21.7, 25.8, 22.2, 15.7, 7.0, 1.6

and those [6/] for the developed countries are given as follows:

Pattern I = DP_1: 13.1, 37.9, 26.6, 13.6, 6.6, 2.0, 0.2

Pattern II = DP_2: 6.5, 30.9, 30.4, 18.8, 9.7, 3.3, 0.3

Pattern III = DP_3: 3.4, 23.4, 31.5, 22.6, 13.6, 5.0, 0.6

The subscripts 1,2 and 3 have been assigned with a view of suggesting typology to the basic fertility patterns. The patterns with subscripts 1's (UP_1 and DP_1) had lower mean, median and modal ages at reproduction compared to the other subscripts in the same group(U or D).

The results obtained by cluster analysis technique were confirmed when individual fertility patterns were graphically represented by

(a) Parameters of a mathematical curve fitted to the fertility patterns(5), and

(b) First two factors in scores in factor analysis on six-variate fertility patterns

Different basic age patterns of fertility occupied different and non-overlapping sub-spaces and thus reinforced the fertility patterns obtained in our earlier analysis.

6. Summary

Age-specific fertility rate data for the period around 1960 from 36 low-fertility and 37 high-fertility countries were converted to age-specific relative fertility rates by equalizing the area for each fertility curve. This was done primarily to investigate variations in fertility patterns.

It was found that the fertility patterns for the high-fertility countries were different from those for the low-fertility countries. Thus these two groups of countries were considered separately in the investigation of the basic age patterns. A cluster analysis technique was utilized to determine a "reasonable" number of patterns. It was found that both the low-fertility countries, excluding Japan and Ireland, and the high-fertility countries could be grouped into three basic patterns each. Their basic patterns, i.e. percent of the total fertility rates accounted for by different quinquennial age groups, are given in the paper.

TABLE 1

Some Descriptive Statistics Calculated at Different Steps, to Determine the Stopping Stage, of the Clustering Process with Simple Distance as the Similarity Index, for the Low-Fertility and the High-Fertility Countries

Low-Fertility Countries				High-Fertility Countries			
Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)	Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)
1	19	46.2	35.0	1	21	25.2	25.9
2	11	45.2	24.1	2	11	17.4	13.4
3	8	41.5	20.9	3	6	22.2	15.5
4	6	36.7	19.5	4	3	28.1	23.1
5	5	39.2	19.4	5	2	26.2	17.2
6	3	20.6	9.1				
7	2	30.0	10.4				

[a]See text for the definitions of these notations.

TABLE 2

Some Descriptive Statistics Calculated at Different Steps, to Determine the Stopping Stage, of the Clustering Process with Cumulated Distance As the Similarity Index, for the Low-Fertility and the High-Fertility Countries

Low-Fertility Countries				High Fertility Countries			
Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)	Step	Number of Clusters	Unweighted Distance[a] (UD_1)	Weighted Distance[a] (WD)
1	21	38.8	59.9	1	25	21.5	55.2
2	13	36.3	54.6	2	12	18.9	24.4
3	10	38.7	22.2	3	9	16.0	27.3
4	8	38.8	27.5	4	6	24.8	34.5
5	7	38.8	22.3	5	4	24.2	23.5
6	6	26.8	22.7	6	3	21.4	27.5
7	5	27.4	17.1	7	2	24.5	10.6
8	4	26.7	11.9				
9	3	29.7	10.7				
10	2	27.8	9.5				

[a]See text for the definitions of these notations.

TABLE 3

Number of S's And D's in the Last N of the $\binom{n}{2}$ Pair-Wise Ranked (Descending) Distances by Sets of Clusters, for the Low-Fertility and High Fertility Countries.

Country Type		Set No. I	Set No. II
Low-fertility	N	213	136
Percent S of total N		81.2	72.1
High-fertility	N	284	252
Percent S of total N		75.7	65.9

FOOTNOTES

1/ This is a part of the first author's Ph.D. dissertation, which goes beyond what has been reported here. Various factors responsible for differential fertility patterns have been studied. Many demographic applications have also been shown.

2/ The U.N. divides all 73 populations into two groups, those with GRR>2.0 and those with GRR < 2.0. We translated these into total fertility rates (TFR).

3/ The cluster analysis program groups n units into a few clusters and the total variance into two components: between-clusters variation and within-clusters variation. Each successive step of the program tends to increase the within-clusters component (pooled within variance) and thus reduces the one corresponding to the between-clusters variation.

4/ A cluster must have seven or more countries for this test to be applicable.

5/ "U" stands for underdeveloped populations. Similarly "D: will be for the developed ones.

6/ Japan and Ireland are omitted, since both of them have unique fertility patterns and form clusters of their own. Japan's unique fertility pattern is explained by a very high incidence of abortion. A very high age at marriage and a very high incidence of spinsterhood in Ireland is responsible for its unique fertility pattern.

REFERENCES

1. Coale, A.J.(1965)"Birth rates, death rates and rates of growth in human population". Public Health and Population Change. Eds. Mindel C. Sheps and Jeanne C. Ridley. University of Pittsburgh Press, Pittsburgh,242-65.

2. Coale, A.J. and Tye, C.Y.(1961). The Signifi- of Afe Patterns of Fertility in High Fertility Population. Milbank Memorial Fund Quarterly 39(4), 631-46.

3. Coale, A.J. and Zelnik, M. (1963). New Estimates of Fertility and Population in the United States. Princeton University Press. Princeton.

4. Gower, J.C.(1967)"A Comparison of Some Methods a Cluster Analysis", Biometrics 23(4).

5. Johnson, N.L. (1949). Systems of Frequency Curves Generated by Methods of Translation. Biometrika 36(1&2), 149-76.

6. Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. John Wiley & Sons, New York.

7. Sokal, R.R., and Sneath P.H. (1963). Principles of Numerical Taxonomy. W.H. Freeman and Company, San Francisco and London.

8. Talwar, P.O.(1970). Age Patterns of Fertility. Institute of Statistics Mimeo Series No. 656, University of North Carolina, Chapel Hill, North Carolina.

9. Talwar, P.P. (1971). "Variations in Fertility Patterns Among Populations". Presented in All-India Seminar on Demography and Statistics, Demographic Research Center, Banaras Hindu University, Varanasi, April 23-25, 1971.

10. United Nations (1965). Population Bulletin of the United Nations No. 7, 1963: Conditions and Trends of Fertility in the World. New York

11. Wilks, S.S. (1963). "Multivariate Statistical Outliers". Sankhya Series A, 25(1).