# USING CLUSTERING ALGORITHMS TO STRATIFY PRIMARY SAMPLING UNITS

D. R. Judkins and R. P. Singh
U.S. Bureau of the Census

## INTRODUCTION

The Current Population Survey (CPS) is a probability sample survey of the U.S. population conducted monthly by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics. Its primary purpose is to provide monthly estimates of labor force characteristics. The sample design is a stratified multistage design. This paper is an analysis of various techniques investigated for stratifying the primary sampling units (PSUs), also known as first stage units.

The current stratification of PSUs is based on a national stratification that was formed by hand in the early 1950's using 1950 Decennial Census data. Since then it has been periodically revised and expanded to provide at least some coverage in every state and to reflect changes in population size and distribution and changes in industrial mix among areas. (See Technical Papers 7 and 40.) The strata were formed to provide national and regional estimates of labor force characteristics. As a result, most of these national strata cross state boundaries.

However, in the mid 1970's, the data needs changed significantly when estimates were required at the state and substate level. In order to obtain state and substate estimates with sufficient reliability, about three-quarters of the states were either supplemented, restratified, or both. The between-PSU variances for state estimates are very large in those states that were not restratified, but even in the states that were restratified, the variances are not as small as they could be. This is because of the severe constraints that were imposed upon the restratification process to retain all existing sample PSUs.

When the changes in requirements are considered together with the general deterioration in the national strata caused by demographic and labor force shifts, it becomes clear that considerable gains in efficiency could be obtained by restratifying all of the states. Since computer time is less expensive now, and new data from the 1980 Decennial Census will soon be available, it was decided to do this as part of a general redesign of the CPS and other surveys conducted by the Bureau.

After reviewing several methods, it was decided that an appropriate approach would be to modify a clustering algorithm developed by Friedman and Rubin (1967). (The modification is described in the section, "Definition of a Good Stratification.") Work was begun on the modifications in the summer of 1980. Since it was then unclear whether the modifications could be completed within our time constraints, we decided to simultaneously compare some additional methods in greater detail. There was also the consideration that until the modifications had been completed and the algorithm had been tested, it could not be determined whether the algorithm would adequately fulfill our needs. Thus the primary purpose of this study was to identify alternate methods in case the modified Friedman-Rubin Algorithm was not ready or did not work satisfactorily.

While the study was in progress, it became apparent that it would be possible to modify at least the major section of the Friedman-Rubin Algorithm. So the scope of the study was expanded to also determine whether it was sufficient to modify the major section (the hill-climbing pass) or it was necessary to modify two additional sections (the forcing and reassignment passes) as well. (For a discussion of these passes, see the section, "Non-Hierarchical Algorithms.")

## STRATIFICATION METHODS AND CLUSTER ANALYSIS

In this section we first discuss what seems to be the traditional approach to stratification and our reasons for not using it. We then discuss cluster analysis and how it has come to be used as a tool for stratification.

Aside from using personal judgment, the method of stratification referred to most often in the texts on sampling theory is what we refer to as rectangular stratification. (See Cochran (1977) and Kish (1965).) The range of each stratification variable is partitioned. Then the Cartesian product of the partitions is formed. Each n-dimensional cell which contains at least one of the sampling units is then defined to be a stratum. In the case of a single continuous variable and optimal or proportional allocation of sample, the optimal boundary points were determined by Dalenius (1950). His work has been extended to two continuous variables by Ghosh (1963). A non-linear programming approach for several continuous or discrete variables has been developed by Schneeberger (1970).

It was decided that the rectangular methods are inappropriate for our problem for several reasons. The first is that the minimum number of strata that can be formed is quite large if there are very many stratification variables. Suppose that there are J variables and that the range of the $j^{th}$ variable is to be partitioned into $L_j$ cells. Then the number of strata to be formed is $(L_1)(L_2)\cdots(L_J) \geq 2^J$. This is too many strata for most states. (One possible solution to this problem might be to take the principle components of the data matrix as suggested by Hagood and Bernert (1945) and to stratify using the factor scores of the units.) The second reason is that there are not enough PSUs in most states for the distribution of the stratification variables to be considered approximately continuous. This condition is more likely to be satisfied when stratifying second or third stage units. The final reason is that roughly equal stratum populations are necessary in order to have both equal weights and efficient utilization of interviewers. This would have been difficult to achieve with rectangular strata.

Cluster analysis is a collection of techniques for exploratory data analysis. It works with sets of objects and tries to find natural groupings among them. After it has been performed, other types of analysis can be performed on the groups to see if they are really different and to better understand the whole population. The origins of cluster analysis are to be found in taxonomy. With the development of high speed computers, a number of algorithms were developed in the sixties by MacQueen (1967), Johnson (1967), Beale (1969), Friedman and Rubin (1967), as well as by numerous others. Though the purpose of cluster analysis is to identify natural groupings of objects, it frequently occurs that if the objects are taken to be PSUs and the clusters to be strata, then the between-PSU variances resulting from sampling are small. This effect of cluster analysis is what led us to consider using it as a method of stratification.

After we started work, we discovered that we were not the first to try this approach. Starting in the late sixties there have been several previous applications of cluster analysis to the problem of stratification of sampling units. The first efforts were made by marketing scientists, Green, et al. (1967), Day and Heeler (1971), and Golder and Yeomans (1973). The first effort by sample survey statisticians was the work of Dahmström and Hagnell (1974, 1978) with the Swedish Labor Survey as an example. The modified Friedman-Rubin algorithm is quite similar to the one they developed in 1978. The modifications to Friedman and Rubin's algorithm were completed despite the similarity because of the uncertainty in how long it might take to obtain a copy of their program.

## THE ALGORITHMS

For the reasons cited above, it was decided that using a clustering algorithm to stratify is the most promising approach. After research and consultation, we decided on the algorithms which we wanted to test. The main criteria were potential effectiveness, availability and ease of implementation. The main sources were Mezzich (1975), Hartigan (1975), and Anderberg (1973). There are many algorithms described in these sources in addition to the ones studied. None of the authors were interested, though, in using clustering algorithms for stratification purposes. So we give below brief descriptions of the algorithms from the point of view of stratification.

There are two principal types of clustering algorithms: Hierarchical and Non-hierarchical. The hierarchical algorithms seek a family of stratifications such that there is a stratification for every possible number of strata and such that every stratum in each stratification is contained in exactly one stratum of every higher stratification.[1] The non-hierarchical algorithms seek only one stratification with a prespecified number of strata.

### Hierarchical Algorithms

The hierarchical algorithms which we studied were linkage algorithms. We chose them because of their availability and simplicity of use as part of BMDP. The linkage algorithms start out assuming that each PSU is self-representing. They then collapse the most similar strata, two at a time, until there is only one stratum. They differ only on the definition of similarity.

The complete linkage algorithms (also known as farthest neighbor and maximum distance algorithms) define the similarity of two strata to be the similarity of the two least similar PSUs contained in them. The average linkage algorithms define the similarity of two strata to be either the similarity of their centers or the average similarity of their PSUs. The single linkage algorithms (also known as nearest neighbor and minimum distance algorithms) define the similarity of two strata to be the similarity of the two most similar PSUs contained in them.

The similarity of PSUs is defined in a number of ways. Define for each PSU a vector consisting of the values of the stratification variables for the PSU. The most common measure of similarity between PSUs is the squared euclidean distance between their vectors. Also used are a measure of association between vectors[2], the absolute value of this association between vectors, the angle between vectors, as well as others. We studied complete, average, and single linkage using both squared euclidean distances and associations between vectors. Results are presented in Table 2.

### Non-hierarchical Algorithms

The two non-hierarchical algorithms which we studied were an algorithm by Beale (1969) called euclidean cluster analysis, and Friedman and Rubin's algorithm (1967), without any of our modifications.

Both algorithms are characterized by the iterative reallocation of PSUs to strata in such a way as to optimize some criterion function. We give a brief description of the algorithms below. For more detail, we refer you to the respective articles.

Friedman and Rubin's algorithm can optimize any one of three different criteria, two of which are invariant under non-singular linear transformations of the data. It has three different methods for determining which reallocations to try. These methods are referred to as the hill-climbing pass, the forcing pass, and the reassignment pass. Use of the hill-climbing pass guarantees the finding of a "one move local optimum" of the criterion function. A "one move local optimum," hereafter referred to simply as a local optimum, is a stratification which cannot be improved by the reallocation of any single PSU. The forcing and reassignment passes are heuristic devices for "getting past" inferior local optima. None of the three methods guarantees the finding of an absolute optimum.

Beale's algorithm optimizes a criterion that is the same as one of the three criteria optimized by Friedman and Rubin's algorithm using a method of reallocation similar to the hill-climbing pass. At this point it will be helpful to set forth some notation.

Let g be the number of strata, $n_k$ be the number of PSUs in the k-th stratum, and $y_{ki}$ be the population and $u_{jki}$ be the estimated number of inhabitants who have the j-th characteristic in the i-th PSU of the k-th stratum. Let a dot signify summation on a subscript. Then the within stratum sum of squared errors (WSS) for the estimated average proportion of the population which have the j-th characteristic is

$$S_j^2 = \sum_{k=1}^{g} \left[ \sum_{i=1}^{n_k} \left( \frac{u_{jki}}{y_{ki}} \right)^2 - \frac{1}{n_k} \left( \sum_{i=1}^{n_k} \frac{u_{jki}}{y_{ki}} \right)^2 \right].$$

Let J be the number of stratification variables, and let $m_j$ be a scale factor for the j-th variable. (This can be used to equalize the total sums of squared errors or to stress one variable over another.) Let

$$S^2 = \sum_{j=1}^{J} m_j S_j^2 \quad .$$

This is one of the criterion functions which Friedman and Rubin's algorithm minimizes. (The other two criterion functions are of less interest to us and were not tested.) It is also the criterion function which Beale's algorithm minimizes. Let me stress that $S^2$ is of no intrinsic interest to us; it is merely a function which is used in these and other clustering algorithms.

## DEFINITION OF A GOOD STRATIFICATION

As mentioned earlier, the purpose of stratification is to minimize between-PSU variances. Using the same notation, the between-PSU variance for the j-th variable is

$$V_j^2 = \sum_{k=1}^{g} \sum_{i=1}^{n_k} \frac{y_{ki}}{y_{k\cdot}} \left( \frac{y_{k\cdot}}{y_{ki}} u_{jki} - u_{jk\cdot} \right)^2 ,$$

where it is assumed that one sample PSU is chosen from each stratum with probability proportionate to size (PPS). (Note the similarity between $S_j^2$ and $V_j^2$.) Since PPS sampling is used for the CPS, it is clear that the best criterion function to minimize is

$$V^2 = \sum_{j=1}^{J} m_j V_j^2 \quad .$$

This is in fact the principal modification made to Friedman and Rubin's algorithm. Instead of minimizing $S^2$ or one of the other two criterion functions, the modified algorithm minimizes $V^2$. (See Kostanich, et al.) The original purpose of this study though was to find possible alternatives, should the modifications not be completed in time. Thus, in comparing the algorithms, we evaluated the performance of each by how much it reduced $V^2$.

Once it became clear that the hill-climbing pass of Friedman and Rubin's algorithm would be modified in time to minimize $V^2$, our efforts were directed towards determining whether the forcing and reassignment passes should be modified to minimize $V^2$ as well. We did this by evaluating how effective the forcing and reassignment passes were at reducing $S^2$ beyond the amount it was reduced by the original hill-climbing pass. We then used this information as an indication of how effective modified forcing and reassignment passes might be in reducing $V^2$ beyond the amount it is reduced by the modified hill-climbing pass.

## PROCEDURES

The testing was done on Colorado, Pennsylvania, and Mississippi. These states were chosen to have representation from three of the four census regions and to have a mix of demographic characteristics. Portions of each state were excluded because they are major population centers which would be selected with certainty in the sample design. Each county was treated as a separate PSU. The number of strata was somewhat arbitrarily taken to be 12 in Colorado, 10 in Pennsylvania, and 12 in Mississippi. Estimates of population and numbers of inhabitants with specific characteristics were taken from the 1970 Decennial Census. The characteristics used for stratification in Colorado were unemployed, civilian labor force (CLF), Spanish unemployed, and Spanish CLF. In Pennsylvania and Mississippi, black unemployed and black CLF were substituted for the Spanish characteristics. Total population 16 years and older was used to calculate proportions and PSU probabilities of selection. The data was initially scaled to equalize the total sum of squared errors for each characteristic.

To test the linkage algorithms, we used the BMDP 77 package. We had to transpose the data since the BMDP programs are meant to cluster variables, not cases. To test Beale's algorithm, we used a FORTRAN subroutine written by Sparks (1973). To test Friedman and Rubin's algorithm, we used the program written by Rubin. We obtained a copy of it from the IBM Program Information Center. For a detailed description of the program, see Rubin (1967).

Both Beale's and Friedman and Rubin's algorithm require an initial stratification as input. We felt that the most valid comparison between the two would be to input the same random initial stratification to both. Unfortunately, this was not possible. Sparks' version of Beale's algorithm assumes that each PSU belongs to the stratum that has the closest center in J-dimensional euclidean space, whereas Rubin's program allows each PSU to belong to any stratum. It happens quite frequently that the center of a randomly generated stratum is not close to any PSU in J-dimensional euclidean space. This leads to an empty stratum, to which PSUs are never allocated by Sparks' program. Thus, if we input the same 12 random strata to each, the comparison of final stratifications would be invalidated by the fact that one stratification might have only 10 or 11 strata instead of 12. It is, of course, easier to minimize $V^2$ with 12 strata than with 10 or 11.

We decided that the best solution was to use random sets of PSUs as initial stratum centers for Sparks' program and to use random stratifications for Rubin's program. (Rubin's program generates these internally.) The problem does not arise with the linkage algorithms since they do not require an initial stratification. Given a fixed number of strata, each of them usually generates a unique stratification. (The only exception is when two pairs of strata have identical similarity measures. This leads to ambiguity as to which pair to collapse first.)

## RESULTS AND ANALYSIS

### Comparison of Algorithms

We obtained 30 stratifications using Beale's algorithm and about 60 using Friedman and Rubin's algorithm in each of the three states. We discuss the forcing and reassignment passes in the next section. For purposes of comparison with Beale's algorithm, we were most interested in the hill-climbing pass. The method of reallocation is similar, but there is at least one difference. Friedman and Rubin's algorithm tries to reallocate each PSU to every other stratum. Beale's algorithm, as written by Sparks, tries to reallocate each PSU only to the stratum with the closest center in J-dimensional space (excepting the stratum to which the PSU is currently allocated). There may be other differences which we could not detect due to the size and complexity of Rubin's program and time constraints. We calculated the sum of scaled between-PSU variances, $V^2$, for all of the stratifications produced by Beale's algorithm and for 30 of the produced by the hill-climbing pass of Friedman and Rubin's algorithm. Table 1 shows the average $V^2$ produced by each algorithm in each of the three states.

Since we suspected that Friedman and Rubin's algorithm might be better than Beale's, we also calculated two-sample t-statistics to test this hypothesis in each state. In computing the t-statistics, we treated the values of $V^2$ as independent observations from two distributions in each state. We then followed standard procedures and appealed to the Central Limit Theorem to assume that the t-statistics are approximately normally distributed. (It is difficult to say how good the approximation is with thirty observations; it may be rather rough.) The t-statistics and the corresponding approximate tail areas are given in the fourth and fifth columns of Table 1. The null hypothesis that the two algorithms produce stratifications with equally small between-PSU variances can be rejected at the level of significance 0.01 in both Colorado and Pennsylvania. In Mississippi the difference is not large enough to conclude that Friedman and Rubin's algorithm is better.

Table 2 shows the sum of scaled between-PSU variances, $V^2$, for stratifications produced with the linkage algorithms for Colorado. These values of $V^2$ are far larger than those for stratifications produced by Beale's or Friedman and Rubin's algorithms, as can be seen by comparison with Table 1. For this reason, we dropped the linkage algorithms from consideration and did not test them on the other states. Though they may be excellent tools for exploratory data analysis, they seem to be ill-suited to the task of stratification.

### Forcing and Reassignment Passes

To assist us in deciding whether to modify the forcing and reassignment passes, we evaluated how useful the original versions are at reducing $S^2$, the sum of scaled within stratum sums of squared errors. We ran Friedman and Rubin's program a total of 201 times on the test states. We then calculated for each run how much of the total reduction in $S^2$ was due to the forcing and reassignment passes. The results are summarized in columns (1) through (3) of Table 3. Column (1) gives various potential reductions in $S^2$. Column (2) shows how many times each potential reduction was realized. It can be seen from column (3) that the two passes did not reduce $S^2$ at all 33.3% of the time. The reduction was less than 2% for 52% of the runs. The maximum reduction was 25%.

These results indicate how much reduction in $S^2$ could be achieved by using the forcing and reassignment passes if the program is run a single time. However, it is a good idea to run the program more than once since the quality of the stratifications produced by the program varies quite a bit from one run to the next and does not depend on the quality of the initial stratifications (see Table 4).

To simulate how useful the forcing and reassignment passes would be if the program was run several times, we randomly grouped the runs into groups of five, ten and twenty. From each group we compared the best stratification produced by just the hill-climbing pass with the best produced by all three passes. The percent reduction in $S^2$ between the two stratifications is the reduction due to the forcing and reassignment passes. Columns (4), (6) and (8) of Table 3 show how many times each potential reduction was realized when the group sizes were five, ten and twenty, respectively. The maximum reduction in $S^2$ that is achieved by using the two passes decreases as the number of runs in a group increases: 25% for one run, 9% for five runs, 5% for ten runs, and 3% for twenty runs.

Since the number of groups in the analysis is very small, the results are not conclusive. However, they do give an indication of how the usefulness of the forcing and reassignment passes decrease as the number of initial stratifications increases. Conversely, they provide some information on the potential design efficiency that might be lost if only the hill-climbing pass is used on a small number of initial stratifications instead of all three passes.

## SUMMARY

Testing in the concurrent paper by Kostanich, Judkins, Singh, and Schautz as well as testing not published has shown us that there are at least two methods of stratification superior to those considered in this paper. These methods are to use the algorithm developed by Hagnell and Dahmström or to use the modified hill-climbing pass of Friedman and Rubin's algorithm developed here at the Bureau to directly minimize $V^2$. If neither of these methods is available, then we would rank the alternatives presented in this paper in the following order: (1) Friedman and Rubin's original algorithm, (2) Beale's algorithm, and (3) the various linkage algorithms. The difference between the first two is slight but seems to be nonetheless present. We have not ranked the rectangular methods since they were not easily applicable to the problem of stratifying small numbers of PSUs using several stratification variables. A more appropriate problem for comparison of the rectangular methods with the cluster analysis methods would be one involving large numbers of sampling units.

Our testing of Friedman and Rubin's algorithm showed that most or all of the reduction in $S^2$ achieved by using the forcing and reassignment passes could also be achieved by using the hill-climbing pass on a number of initial stratification. The Bureau chose, therefore, not to modify or use the forcing and reassignment passes. It is our feeling that under most conditions it would not be worthwhile to modify and use them in conjunction with the modified hill-climbing pass. The only conditions under which it might be worthwhile to modify the two passes would be when the numbers of PSUs, strata, and characteristics are so large as to make prohibitive the cost of using the modified hill-climbing with a number of initial stratifications.

ACKNOWLEDGMENTS

The authors wish to thank Donna Kostanich for her comments and Frances Gunther, Margaret Poole, and Edith Oechsler for their typing.

FOOTNOTES

[1]Hierarchical methods may be of the greatest interest to biologists since not only species, but also genera, classes, families, etcetera may be identified.

[2]Note that this association between vectors can be used in the BMDP package by specifying the correlation option with a transposed data matrix.

BIBLIOGRAPHY

Anderberg, M. R. (1973). Cluster Analysis for Applications. Academic Press, New York.

Beale, E. M. L. (1969). "Euclidean Cluster Analysis." Bulletin of International Statistical Institute, Proceedings of the 37th Session. (Contributed paper).

Cochran, W. G. (1977). Sampling Techniques. 3rd edition. John Wiley & Sons, New York.

Dahmström, P. and Hagnell, M. "The Formation of Strata Using Cluster Analysis." Statistisk Tidskrift 1974: 6, 477-485.

Dahmström, P. and Hagnell, M. "Multivariate Stratification of Primary Sampling Units in a General Purpose Survey." Scand, J. Statist. 1978: 5, 53-56.

Dalenius, T. (1950). "The Problem of Optimum Stratification." Scandinavisk Aktuarietidskrift 33, 203-213.

Day, G. S. and Heeler, R. M. (1971). "Using Cluster Analysis to Improve Marketing Experiments." Journal Marketing Research 8, 340-347.

Friedman, H. P. and Rubin, J. (1967). "On Some Invariant Criteria for Grouping Data." Journal of the American Statistical Association 62, 1159-1178.

Ghosh, S. P. (1963). "Optimum Stratification with Two Characters." Annals Mathematical Statistics 34, 866-872.

Golder, P. A. and Yeomans, K. A. (1973). "The Use of Cluster Analysis for Stratification." Applied Statistics 22, 213-219.

Green, P. E., Frank, R. F. and Robinson, P. J. (1967). "Cluster Analysis in Test Market Selection." Management Science 13, B-387-400.

Hagood, M. J. and Bernert, E. H. (1945). "Component Indexes as a Basis for Stratification in Sampling." J. American Statistical Association 40, 330-341.

Hartigan, J. A. (1975). Clustering Algorithms. Wiley, New York.

Johnson, S. C. (1967). "Hierarchical Clustering Schemes." Psychometrika 32, 241-254.

Kish, L. (1965). Survey Sampling. John Wiley & Sons, New York.

Kish, L. and Anderson, D. W. (1978). "Multivariate and Multi-purpose Stratification." Journal American Statistical Association 73, 24-34.

Kostanich, D. L., et al. "Modification of Friedman-Rubin's Clustering Algorithm for Use in Stratified PPS Sampling." Paper presented at the 1981 American Statistical Association in Detroit.

MacQueen, J. B. (1967). "Some Methods for Classification and Analysis of Multivariate Observations." Proceedings Symp. Math. Statist. and Probability, 5th, Berkeley, 1, 281-297, AD 669871, Univ. of California Press, Berkeley.

Marriott, F. H. C. (1971). "Practical Problems in a Method of Cluster Analysis." Biometrics 27, No. 3, 501-514.

Mezzich, Juan E. "An Evaluation of Quantitative Taxonomic Methods," PhD. dissertation, Ohio State University, 1975. (Library of Congress has it on microfilm. 7526626)

Rubin, J. (1967). "Optimal Classification into Groups: An Approach for Solving the Taxonomy Problem." Journal of Theoretical Biology 15, 103-144.

Schneeberger, H. (1970). "Optimierung in der Stichprobentheorie durch Schichtung." Statistische Hefte 11, 242-253.

Scott, A. J. and Symons, M. J. (1971). "Clustering Methods Based on Likelihood Ratio Criteria." Biometrics 27, No. 2, 387-398.

Sparks, D. N. (1973). "Euclidean Cluster Analysis." Applied Statistics 22, No. 1, 126-130.

U.S. Dept. of Commerce, Bureau of the Census. The Current Population Survey: A Report on Methodology, by J. Steinberg, T. B. Jabine, and L. Pritzker, Technical Paper 7, Washington, D.C.: GPO, 1963.

U.S. Dept. of Commerce, Bureau of the Census. The Current Population Survey: Design and Methodology, by R. Hanson. Technical Paper 40, Washington, D.C.: GPO, 1978.

TABLE 1. Comparison of Beale's and Friedman and Rubin's Algorithms

| State | Average $V^2$ Obtained from 30 starts | | Test the hypothesis that both perform equally well against the hypothesis that Friedman & Rubin's performs better | |
|---|---|---|---|---|
| | Beale's Algorithm | Friedman & Rubin's Algorithm (Hill-climbing Pass only) | Two-Sample t-Statistic | Approximate Tail Area |
| Colorado | $8.351 \times 10^7$ | $6.814 \times 10^7$ | -3.62 | 0.0002 |
| Pennsylvania | $4.095 \times 10^8$ | $3.669 \times 10^8$ | -2.48 | 0.0065 |
| Mississippi | $4.074 \times 10^8$ | $4.051 \times 10^8$ | -0.15 | 0.44 |

Table 2. $V^2$ Obtained with Linkage Algorithms in Colorado

| Type of Stratum Similarity \ Measure of PSU Similarity | Euclidean Distance | Association Measure |
|---|---|---|
| Complete Linkage | 1.5353x10^8 | 7.2328x10^8 |
| Average Linkage | 5.3534x10^8 | 7.4116x10^8 |
| Single Linkage | 2.7633x10^9 | 2.6188x10^9 |

TABLE 3. REDUCTION IN $S^2$ DUE TO FORCING AND REASSIGNMENT PASSES OVER HILL-CLIMBING PASS
(Colorado, Mississippi and Pennsylvania Combined)

| Percent Reduction in $S^2$ | Single Runs | | Groups of 5 Runs | | Groups of 10 Runs | | Groups of 20 Runs | |
|---|---|---|---|---|---|---|---|---|
| | Number | Cumulative Percent | Number | Cumulative Percent | Number | Cumulative Percent | Number | Cumulative Percent |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 0.0 | 67 | 33.3 | 8 | 20.5 | 2 | 11.8 | 1 | 11.1 |
| 0.0-0.5 | 16 | 41.3 | 7 | 38.4 | 5 | 41.2 | 3 | 44.4 |
| 0.5-1.0 | 8 | 45.3 | 2 | 43.5 | 0 | 41.2 | 0 | 44.4 |
| 1.0-2.0 | 13 | 51.8 | 8 | 64.0 | 7 | 82.4 | 3 | 77.7 |
| 2.0-3.0 | 6 | 54.8 | 6 | 79.4 | 2 | 94.2 | 2 | 99.9 |
| 3.0-4.0 | 13 | 61.3 | 0 | 79.4 | 0 | 94.2 | | |
| 4.0-5.0 | 10 | 66.3 | 6 | 94.8 | 1 | 100.1 | | |
| 5.0-6.0 | 16 | 74.3 | 1 | 97.4 | | | | |
| 6.0-7.0 | 8 | 78.3 | 0 | 97.4 | | | | |
| 7.0-8.0 | 11 | 83.8 | 0 | 97.4 | | | | |
| 8.0-9.0 | 6 | 86.8 | 1 | 100.0 | | | | |
| 9.0-10.0 | 10 | 91.8 | | | | | | |
| 10.0-14.0 | 9 | 96.3 | | | | | | |
| 14.0-18.0 | 4 | 98.3 | | | | | | |
| 18.0-22.0 | 3 | 99.8 | | | | | | |
| 22.0-26.0 | 1 | 100.3 | | | | | | |
| Total | 201 | 100.3 | 39 | 100.0 | 17 | 100.1 | 9 | 99.9 |

TABLE 4. SUM OF SCALED WITHIN STRATUM SUMS OF SQUARED ERRORS ($S^2$) FOR VARIOUS STAGES OF STRATIFICATION FOR COLORADO DATA

| S. No. | Value of $S^2$ | | |
|---|---|---|---|
| | For Initial Stratification | After Hill-Climbing Pass Only | After All Three Passes |
| 1 | 3.2590 | .44363 | .44363 |
| 2 | 3.4349 | .45302 | .45302 |
| 3 | 2.9385 | .45807 | .45807 |
| 4 | 3.3734 | .45480 | .43535* |
| 5 | 2.6566* | .47285 | .45664 |
| 6 | 2.9644 | .44586 | .43535* |
| 7 | 3.3550 | .43539* | .43539 |
| 8 | 3.1575 | .48545 | .45775 |
| 9 | 3.1466 | .46615 | .45199 |
| 10 | 3.0848 | .47729 | .47729 |

* Indicates the lowest $S^2$ for the column.