TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. TX-84/239-5	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

Estimating Pavement Condition and Rehabilitation Costs Using Statistical Sampling Techniques

5. Report Date

June 1984

6. Performing Organization Code

7. Author(s)

Connie Jill Templeton and Robert L. Lytton

8. Performing Organization Report No.

Research Report 239-5

9. Performing Organization Name and Address

Texas Transportation Institute
The Texas A&M University System
College Station, Texas 77843

10. Work Unit No.

11. Contract or Grant No.

Study No. 2-18-79-239

12. Sponsoring Agency Name and Address

Texas State Department of Highways and Public Transportation; Transportation Planning Division
P. O. Box 5051, Austin, Texas 78763

13. Type of Report and Period Covered

Interim - September 1978
June 1984

14. Sponsoring Agency Code

15. Supplementary Notes

Research Study Title: Pavement Rehabilitation Fund Allocation

16. Abstract

The Texas Department of Highways and Public Transportation (SDHPT) is currently collecting large volumes of data with its Pavement Evaluation System. It is economically feasible only to conduct a sample survey of Texas 70,000 mile highway network each year.

This report describes a study performed by the Texas Transportation Institute to investigate the effectiveness of various sample sizes at predicting the pavement score distribution and the rehabilitation cost. A unique feature of this work is that the sample was used to estimate the distribution of pavement scores, rather than simply the mean score. By using the distribution of scores it is possibile to estimate the number of miles below a minimum acceptable score and the approximate cost to maintain and rehabilitate the entire pavement network.

The data used in this study was collected in 3 Texas districts in 1982. Within each District a 100% survey was made of the pavement condition. Therefore, the true pavement score distribution and estimated rehabilitation cost were known with certainty. In this study, several types of sampling plans with varying sample sizes were examined. It was concluded that sampling by county is the most efficient method, and that a 10-15 percent survey is adequate to determine the condition of the pavement network. It was also found that a 30-35 percent survey is required to achieve accurate estimates of the cost of rehabilitation.

17. Key Words Pavement Evaluation, Sampling, Rehabilitation.	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 245	22. Price

Form DOT F 1700.7 (8-69)

ESTIMATING PAVEMENT CONDITION AND REHABILITATION COSTS USING STATISTICAL SAMPLING TECHNIQUES

by

Connie Jill Templeton

and

Robert L. Lytton

Research Report 239-5
Research Study Number 2-18-79-239
Pavement Rehabilitation Fund Allocation

Sponsored by

Texas State Department of Highways and Public Transportation

June 1984

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas

ABSTRACT

The Texas Department of Highways and Public Transportation (SDHPT) is currently collecting large volumes of data with its Pavement Evaluation System. It is economically feasible only to conduct a sample survey of Texas 70,000 mile highway network each year.

This report describes a study performed by the Texas Transportation Institute to investigate the effectiveness of various sample sizes at predicting the pavement score distribution and the rehabilitation cost. A unique feature of this work is that the sample was used to estimate the distribution of pavement scores, rather than simply the mean score. By using the distribution of scores it is possible to estimate the number of miles below a minimum acceptable score and the approximate cost to maintain and rehabilitate the entire pavement network.

The data used in this study was collected in 3 Texas districts in 1982. Within each District a 100% survey was made of the pavement condition. Therefore, the true pavement score distribution and estimated rehabilitation cost were known with certainty. In this study, several types of sampling plans with varying sample sizes were examined. It was concluded that sampling by county is the most efficient method, and that a 10-15 percent survey is adequate to determine the condition of the pavement network. It was also found that a 30-35 percent survey is required to achieve accurate estimates of the cost of rehabilitation.

EXECUTIVE SUMMARY

This report is part of an on-going research project (239) between the Texas Transportation Institute and the State Department of Highways and Public Transportation. This study, in particular, was initiated following a request from the Department to evaluate the effectiveness of various sample sizes in predicting pavement condition and rehabilitation cost at a district level (see attached letter dated April 20, 1983).

The statistical analysis, which is described in detail in the following report, has been performed on the 1982 Pavement Evaluation System data collected in the three 100% survey districts (Districts 8, 11, and 15). The Pavement Evaluation System converts each section's pavement distress and serviceability data into a pavement score and for those sections below a minimum score it calculates a rehabilitation cost. Therefore at the onset of this statistical analysis, the pavement score and rehabilitation cost distribution was known. The analysis involved determining the effectiveness of different sampling schemes in being able to predict the true distribution of pavement scores and rehabilitation costs.

CRITERIA FOR JUDGING SAMPLING EFFECTIVENESS

Throughout this study it has been assumed that the sample taken would be used to estimate the distribution of pavement scores within a

COMMISSION
ROBERT H. DEDMAN, CHAIRMAN
A. SAM WALDROP
JOHN R. BUTLER, JR.

STATE DEPARTMENT OF HIGHWAYS
AND PUBLIC TRANSPORTATION
DEWITT C. GREER STATE HIGHWAY BLDG.
AUSTIN, TEXAS 78701

ENGINEER-DIRECTOR
MARK G. GOODE

April 20, 1983

IN REPLY REFER TO
FILE NO. D-18

Pavement Evaluation

Dr. Robert L. Lytton
Texas Transportation Institute
Texas A & M University
College Station, Texas 77843-3135

Dear Dr. Lytton:

We, the PES Action Team, have been requested by Mr. Goode to submit sample size data for evaluation of pavements to show various confidence levels or degrees of benefit for individual districts. Two tables are requested, one being for information concerning condition of the highways and the other should be for rehabilitation cost information.

It is suggested that this information be developed as tables which indicate sample size verses confidence level and should begin with a sample size of 5%. Several increments should be shown between the 5% and the optimum sample size. Reasons for the specified optimum sample size should be provided.

Since any changes in the present 5% sample size would most likely change the computer programs, the development of this information will have to consider the programming time required to change the programs used. Therefore, tentative date for completion of this task is July 1, 1983.

A meeting with the PES Action Team could be arranged in conjunction with the quarterly meeting scheduled for Project No. 284, if you feel it would be beneficial.

Your comments concerning the requested work, schedule and meeting are requested.

Sincerely yours,

Robert R Guinn

Robert R. Guinn
Engineer of Maintenance

RRG:mlk
cc: Joanne Walsh, MIPR
Joel Young, D-19

district. An example of a typical cumulative distribution is shown below in Figure i.

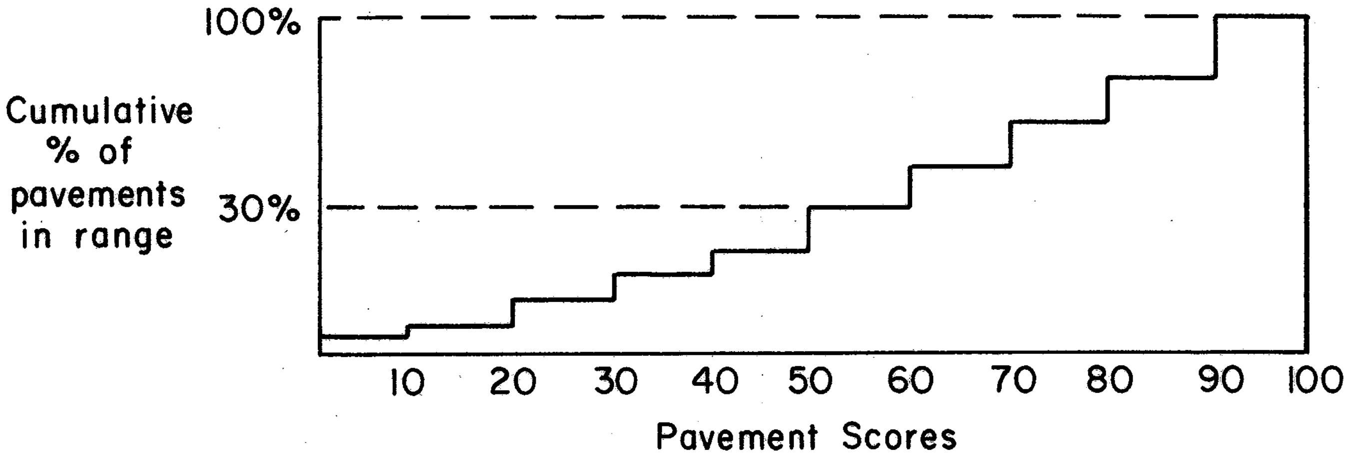


Figure i. Example Distribution of Pavement Scores

In the example given above, the percentage of pavements with a score of 50 or less is 30%. Knowing the distribution of scores, it is possible to calculate the number of miles of a particular pavement type that is below minimum acceptable condition. It is clear that being able to predict this distribution yields considerably more information than does the mean pavement score. Knowledge of the mean score does not indicate how many miles of highway are in poor condition and this is the factor that the district and state personnel are interested in.

Therefore in the statistical analysis, samples of various sizes were extracted from the 100% survey districts. The sampled data distribution was then compared to the 100% survey data as shown below in Figure ii.

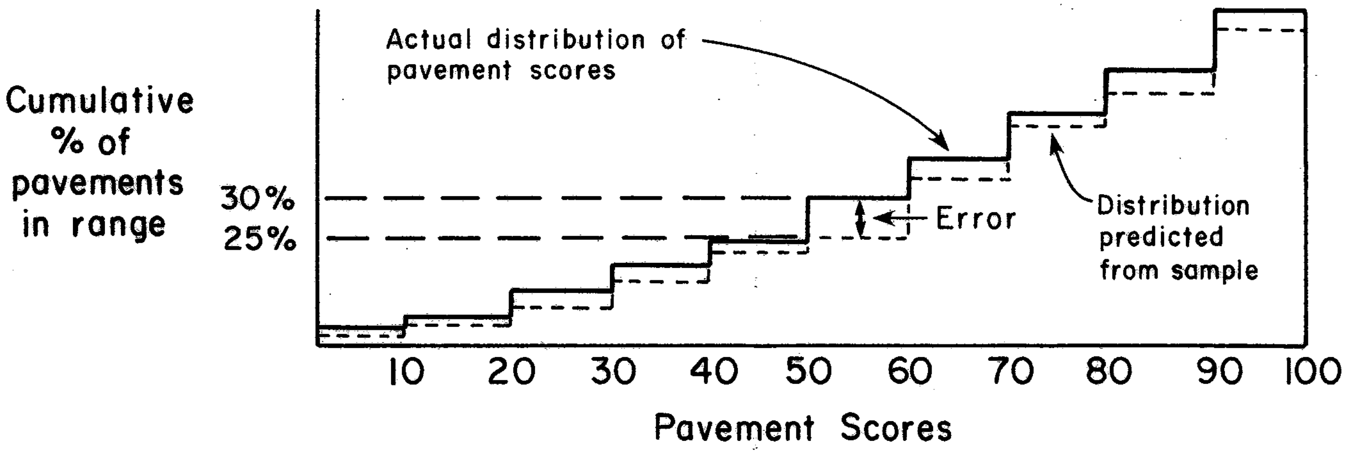


Figure ii. Comparing True with Predicted Pavement Score Distribution to Define Maximum Error

In this example, the true percentage of pavements with a score below 50 was 30%, whereas the sample data predicted that only 25% were below a score of 50. This is an error of 5%. The Maximum Error was thus defined as the greatest difference between the true and the sample distribution. It was this maximum error that was used to judge the effectiveness of the various sampling schemes.

ANALYSIS PERFORMED

The analysis performed in this study involved the following:

1. Taking 300 different random 5% samples from each of the 100% sample Districts.
2. For each sample calculate the Maximum Error.
3. Find the mean Maximum Error for that sample size.
4. Repeat steps (1), (2), and (3) for each of the following sample sizes - 10%, 15%, 20%, ..., 50%.

This proce~~dure was repeated fo~~r each of the 4 pavement classes (IH, US, SH, FM) for each of the three 100% survey districts. Complete details of this analysis are described in the following TTI Report.

TYPICAL RESULTS

The ability of a sample to predict the pavement score distribution is strongly dependent upon the total number of sections to be sampled from. For instance, District 8 had 154 two mile sections of Interstate and 938 two mile sections of Farm-to-Market highway. The accuracy obtained from different sample sizes is shown in Figure iii. Therefore if 5% error is the maximum permissible, then a 15% sample of FM, 25% sample of US, and 50% sample of IH pavements would be required.

To estimate rehabilitation cost for a district from a sample, it must be remembered that only a small percentage of the pavements

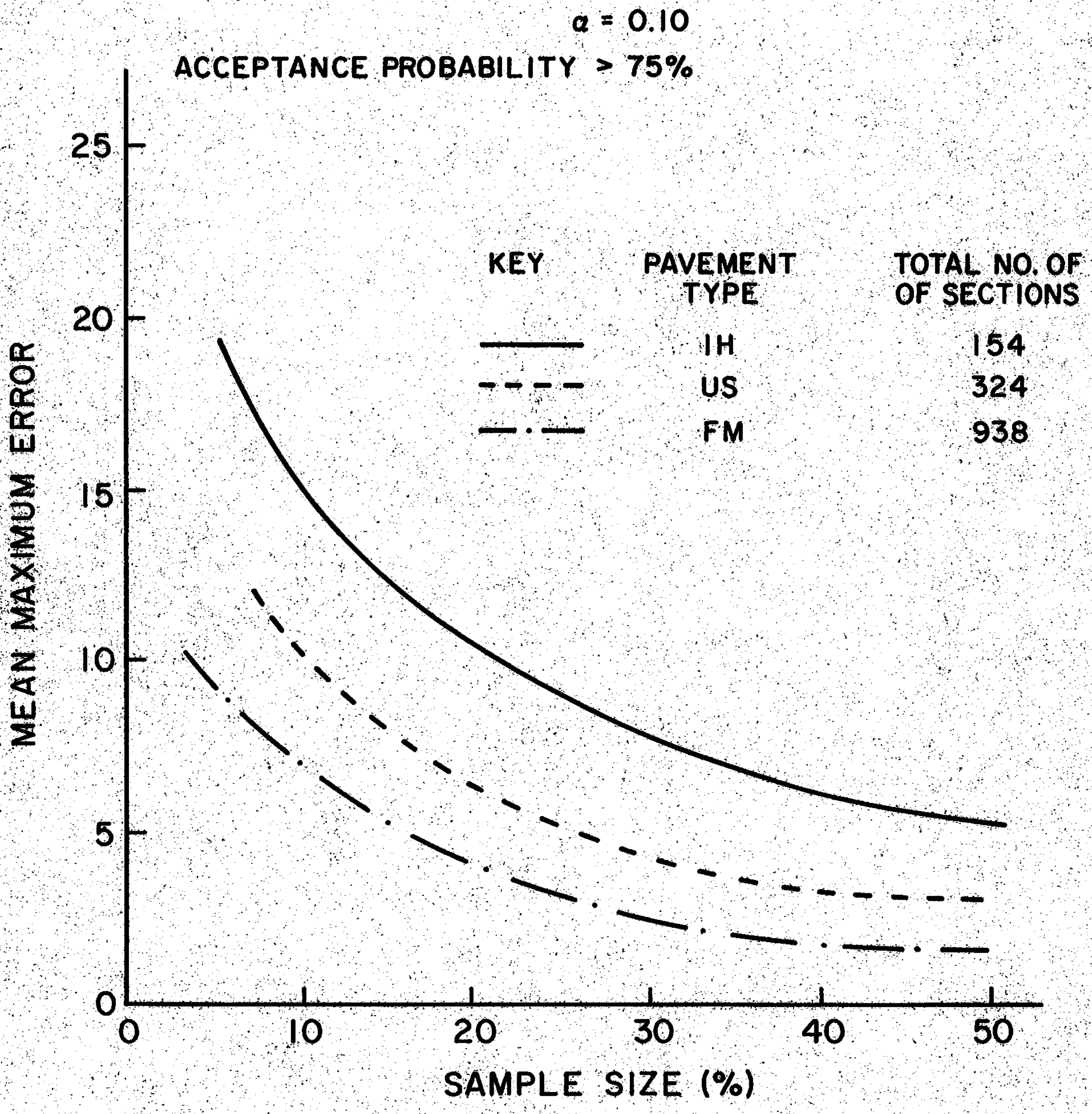


Figure iii. Mean Maximum Error vs. Sample Size for Different Pavement Types require rehabilitation each year. This is typically in the range of 5 to 15%. Therefore, for any single run of PES, only 10% (for example) of the district's pavement may require rehabilitation funds. The remaining 90% have zero rehabilitation funds associated with them. This makes prediction of a district's total rehabilitation costs, from a sample of the district's highways, a very difficult task. Figure iv shows the expected cost prediction accuracy associated with the various sample sizes. This curve was generated from simulations run on District 8, State Highway data. As shown, at least a 26% sample would be required to predict total rehabilitation cost to ± 20% of the actual figure.

CONCLUSIONS

From this study, it is concluded that:

1. A 10-15% survey is adequate for predicting network pavement condition.
2. A 30-35% survey is required to estimate total rehabilitation cost requirements.

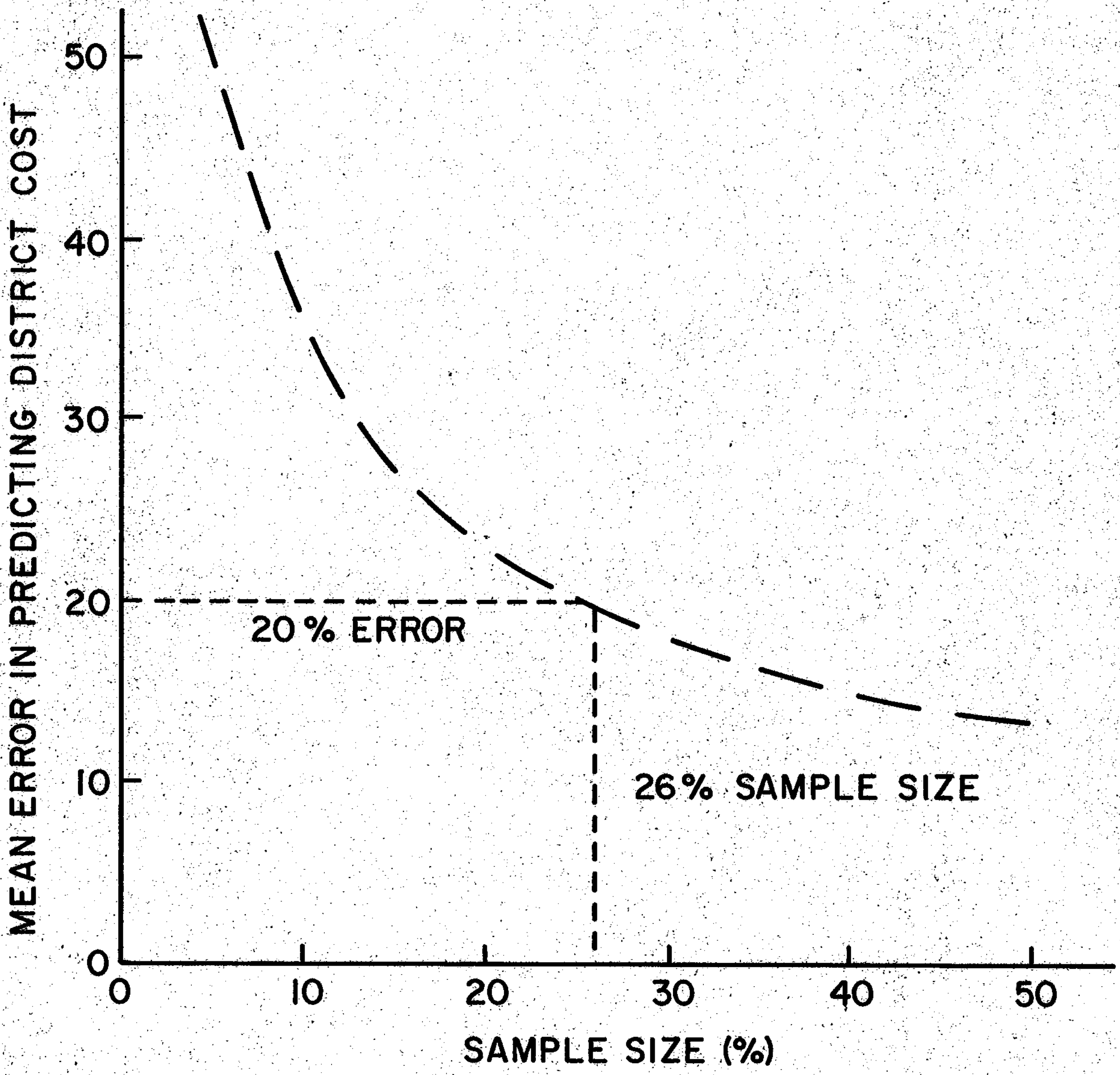


Figure iv. Mean Error in Predicting District Rehabilitation Cost vs. Sample Size

ACKNOWLEDGEMENT

This report is part of an on-going research project between the Texas Transportation Institute and the State Department of Highways and Public Transportation. The support and assistance given by Robert R. Guinn (D-18), Joel F. Young (D-19), and Joann M. Walsh (MIPR), is greatly appreciated.

DISCLAIMER

The views, interpretations, and conclusions expressed or compiled in this report are those of the research group. They are not necessarily those of the Texas Department of Highways and Public Transportation.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

LIST OF TABLES

LIST OF TABLES (cont'd)

APPENDIX TABLES

LIST OF TABLES (cont'd)

LIST OF TABLES (cont'd)

LIST OF TABLES (cont'd)

LIST OF FIGURES

LIST OF FIGURES (cont'd)

APPENDIX FIGURES

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

LIST OF FIGURES (cont'd)

CHAPTER I

INTRODUCTION

Statistics suggest that the national highway network is gradually deteriorating. In his 1977 Report to Congress the Secretary of Transportation indicated that from 1970 to 1975 there was a small but significant shift in pavement condition from the "good" category to the "fair" category (1). During the period from 1962 through 1979, construction funding decreased from 60 percent to 42 percent, while maintenance and rehabilitation funding increased from 23 percent to 33 percent of the total U.S. highway disbursements (2). These increasing shifts from new construction of highways to maintenance or rehabilitation of existing systems enhance the importance of proper procedures to accurately estimate average maintenance and rehabilitation costs per square yard. These costs combined with existing paved area will result in average maintenance and rehabilitation budgets needed for each district of the state system.

While the cost of operating, maintaining, and improving the major highway systems escalates yearly, the rate of increase of the revenues to cover these costs is diminishing. The high cost of fuel has motivated individuals to buy more fuel-efficient vehicles and to do less recreational driving. The consequent decrease in fuel use has greatly reduced the much-needed gas tax revenues. This reality of declining revenues will accelerate the trend toward preserving the current highway network now in place.

In order to assist the Texas Department of Highways and Public Transportation (SDHPT) with the management of its 70,000 mile highway network, the Department has recently implemented its Pavement Evaluation System. This system includes extensive data collection in the form of pavement inspection procedures. These data are used to generate for each pavement section firstly, its current condition (pavement score) and secondly, its rehabilitation cost (for those sections below a minimum pavement score). One of the aims of this system is to provide administrators with accurate estimates of current pavement condition and rehabilitation cost estimates.

Because of limitations on funds, time, equipment, and manpower, it is not anticipated that the Department will be able to inspect every pavement each year. It will be necessary to develop an efficient sampling scheme to provide the required information. The objective of this study has been to investigate the effectiveness of various sampling schemes, including sample survey design and sample size. The analysis to be described in this report is based on an analysis of data collected in three Districts, these being District 8 (Abilene), District 11 (Lufkin), and District 15 (San Antonio). In the 1982 pavement inspections, each of these Districts collected condition data on every flexible pavement within their district. Therefore, it was known with statistical certainty what was the distribution of pavement scores and rehabilitation costs in each of these districts. Using these data, it was possible, using simulation techniques, to estimate the effectiveness of various sampling schemes and sample sizes.

This report is divided into seven chapters and six appendices. The first chapter serves as an introduction. The second chapter gives a brief history of sampling strategies used in predicting certain characteristics related to highways. It also contains an explanation of the data gathering techniques used by the State Department of Highways and Public Transportation (SDHPT) in the collection of the data analyzed in this report. The third chapter examines the accuracy of a five percent sample in predicting pavement condition. The next chapter delves into pavement condition methodology further. Density functions are fitted to the Pavement Score distributions, and two methodologies are presented for predicting the percentage of roads with a particular Pavement Score. The fifth chapter determines the best estimator to use for predicting costs of maintenance and rehabilitation. The sixth chapter looks at various sampling designs and sample sizes to ascertain the best design and its accuracy in predicting costs as well as pavement condition. The final chapter includes the summary, conclusions and recommendations to the SDHPT and recommendations for further work.

Appendices A and B relate to the third chapter. Appendix C describes the Kolmogorov-Smirnov Goodness of Fit Test, while in Appendix D the results of procedures used in the fifth chapter are shown in tabular as well as graphical form. Appendix E contains tables showing the results of procedures described in the sixth chapter. Appendix F contains the computer programs used in the analysis of the data.

CHAPTER II
LITERATURE SURVEY

Sample Survey Designs

The purpose of a sample survey is to make inferences about the parameters and/or type of probability distribution of the sampled population. The design of the sample survey should be carefully chosen in order to minimize costs while maximizing the information gained in the survey. In general the sampling costs are either minimized or significantly reduced when the sample size is chosen as small as possible. Some of the sample survey designs available and a brief description of each follow (3, 4, 5, 6):

1. Simple random sampling. This method assigns each possible sample an equal and independent chance of being selected from the population.
2. Stratified random sampling. This method divides the population into nonoverlapping groups or strata. When the strata have been determined, a sample is drawn from each, the drawings being made independently in different strata. If a simple random sample is taken in each stratum, the whole procedure is described as stratified random sampling.
3. One-stage cluster sampling. In this method the population is divided into subpopulations and some but not all of these subpopulations are represented in the sample. Those represented may be included in the sample in their entirety or they may be subsampled. In stratified sampling the population is also divided into

subpopulations, but all subpopulations are represented and they are always subsampled.

4. Systematic sampling. This method obtains a sample by randomly selecting one element from the first k elements in the population and every k-th element thereafter.

Combinations of the above five methods can also be defined as additional sampling schemes. Several different survey designs have been used by the SDHPT for estimating various highway data elements.

In 1975, data were collected on all flexible highways in District 21 after dividing the network into two-mile segments. The following list briefly describes the data collected (7):

1. Construction information: Includes layer thickness, widths and available material properties.
2. Traffic history: Includes average daily traffic and 18,000 lb. equivalent axle loads applied with time.
3. Climate data: Monthly rainfall and temperatures, freeze-thaw cycles, Thornwaite indices (moisture balance indices which measure the difference between rainfall and evaporation/transpiration).
4. Roughness: Serviceability indices obtained with the Mays Ride Meter.
5. Visual condition: Distress manifestations obtained primarily by use of a visual process.
6. Deflection: Obtained using the Dynaflect.
7. Rut depth measurements.
8. Skid number (measurement of resistance between the tires of a

vehicle and the pavement when the vehicle is braking) at 40 mph.

Three elements of pavement performance calculated from these data were Serviceability Index, Pavement Rating Score, and Surface Curvature Index. These elements are briefly discussed as follows.

The Serviceability Index is an indicator of road roughness and is based on a scale which ranges from 0 to 5. A value of 5 represents a road which is perfectly smooth and 0 indicates a road which is virtually impassable. For this survey, the car-mounted Mays Ride Meter was used to determine the Serviceability Index (8). This instrument accumulates roughness over a 0.2 mile distance, thus ten Serviceability Index values were obtained in each of the two-mile highway segments. The instrument provides a raw value which is reduced to the 0 to 5 scale by using SDHPT calibration procedures.

The Pavement Rating Score is an indicator of visually determined distress manifestations present on the pavement surface (9). The evaluation procedure yields a score which ranges from 100 (perfect pavement - no observable distress) to 0 (extreme amount of distress present on the pavement surface). The nine different distress types included are rutting, raveling, flushing, corrugations, alligator cracking, longitudinal cracking, transverse cracking, patching and failures. Each distress type is evaluated by determining the "area" and "severity" for each. The Pavement Rating Score is determined by subtracting points from 100 for each area-severity combination for each of the nine distress types.

The Surface Curvature Index, a measure of the structural adequacy of a pavement, is obtained by use of the Dynaflect. This instrument is a small, two-wheel trailer which applies a peak-to-peak dynamic force of 1,000 lbs. at a fixed frequency of 8 Hz. The resulting deflections (in milli-inches) are measured at five locations spaced at one-foot intervals on the axis of symmetry which passes between the load wheels. The Surface Curvature Index is the difference in measured deflections between the first and second deflection sensors.

Using these data and simulation techniques, Mahoney and Lytton (7) developed a stratified two-stage sampling scheme to estimate the three previously described pavement performance elements for the Texas highway network. It was found that the optimal sample size is a function of the weights placed on the sampling costs and sampling variability. The results indicated that on the average the optimal sample size lies between 1.5 and 6.6 percent of the total centerline mileage of the highway network.

In a subsequent study aimed at estimating maintenance and rehabilitation costs, pavement condition survey data from District 17 were used. It was found that a simple random sample of 10 percent of the centerline mileage yielded a coefficient of variation of total maintenance and rehabilitation costs of 31 percent, while a 40 percent sample gave a coefficient of variation of 14 percent.

In an effort to assess the national highway systems, the Federal Highway Administration (10) has implemented the Highway Performance Monitoring System, a program management tool which assesses the extent and condition of the highway systems and attempts to detect changes from one year to the next strictly by sampling. In this system cluster type sampling was applied to local road sections for the estimation of selected highway data elements. The sample design was based on a random selection of a fixed number of sections or milepoints within geographic subareas (subclusters) contained in counties or urban subdivisions (clusters) of a state (11). Each subcluster and cluster were also randomly selected.

Alternatively, stratified random sampling has been used for interstates, freeways, arterials and collectors. Each of these functional classes within each of the geographic areas (rural, small urban, and urban) was considered as a separate population. Within each population, road sections were stratified by average annual daily traffic (AADT)(12). In computing sample size, it was suggested that a ten percent change in the mean value should be detectable at a 90 percent level of confidence for important functional classes of pavement and at an 80 percent level of confidence for other classes (13).

Morgan and Burati (14) compared simple random sampling of stratified random sampling when testing small rectangular sections of bituminous pavement. Assuming that the sample results are positively correlated and that the correlation between two sampling locations decreases exponentially with distance, it was found that stratified random sampling achieved a lower sampling variance than simple random sampling.

Related Background

In 1982 all highways within each Texas District were divided into segments approximately two miles in length. Five percent of the total number of segments in each of 21 districts and each of the four roadway systems (Interstate, Farm-to-Market, State, and U.S.) were selected at random for sampling. A segment included all paved areas between two designated mileposts. Hence an interstate highway segment could include four roadways (two main roadways and two frontage roads). However, for the purpose of analysis presented in this report, all frontage roads were deleted and only main roadways were considered. One lane of each roadway within the selected segment was sampled and each of these observations was considered a sampling unit. Figure 1(a) shows a divided highway segment with main roadways only. The shaded area depicts the two observations associated with the segment. The outer lanes (R and L) were not necessarily always chosen. Other possible combinations could have been R and M; S and M; or S and L. Figure 1(b) shows a two lane highway segment. Only one observation is chosen from this segment. Either lane, L or R, could have been chosen.

In the remaining three districts [Districts 8 (Abilene), 11 (Lufkin), and 15 (San Antonio)] a 100 percent sample in each roadway system was taken. Figure 2 depicts the location of each of these districts.

For each observation in both samples (five percent and 100 percent) the Serviceability Index was determined with the Mays Ride Meter. A visual defects rating was computed in much the same manner as the Pavement Rating Score discussed in the above section. However, patching and corrugation were not included in the distress types. Skid number and deflection data were also not collected. The Pavement

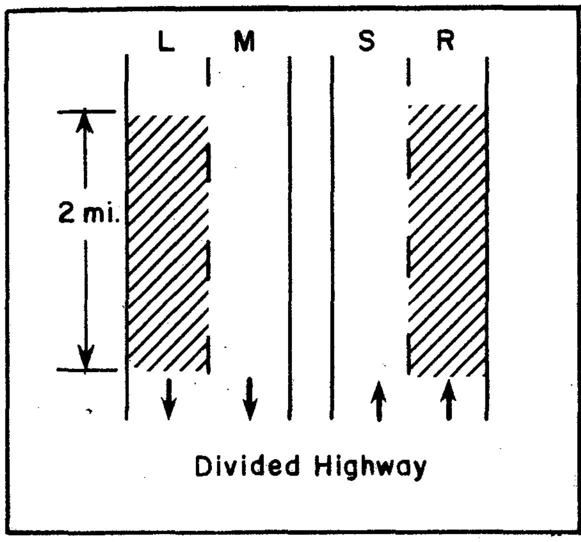


FIGURE 1(a). Observation from a two-mile segment of a divided highway.

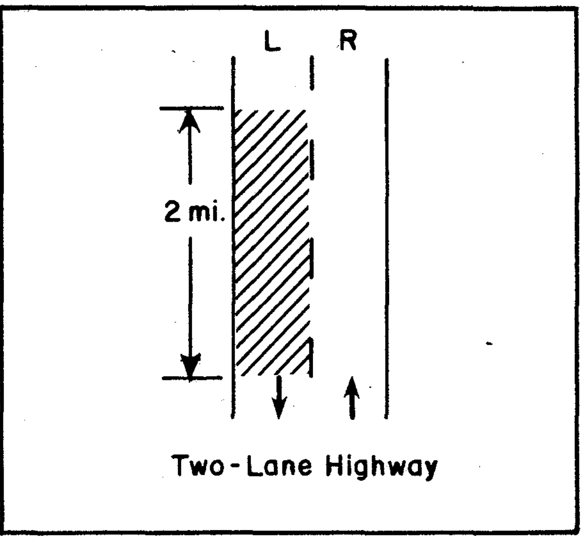


FIGURE 1(b). Observation from a two-mile segment of a two-lane highway.

Score, a measure of the overall pavement condition, is a function of the utilities associated with the Serviceability Index and the visual defects ratings. These utilities are measures of the seriousness of the distress. High values of the utility associated with the visual defects ratings correspond to low values of the extent of the distress, while high values of the utility associated with the Serviceability Index correspond to a high level of ride quality.

If the Pavement Score fell below a specified minimum level, a funding strategy for that particular observation (a two-mile single lane highway strip) was chosen. A funding strategy is an attempt to make available the proper amount of money required to maintain a highway segment. A cost per square yard was associated with each strategy;

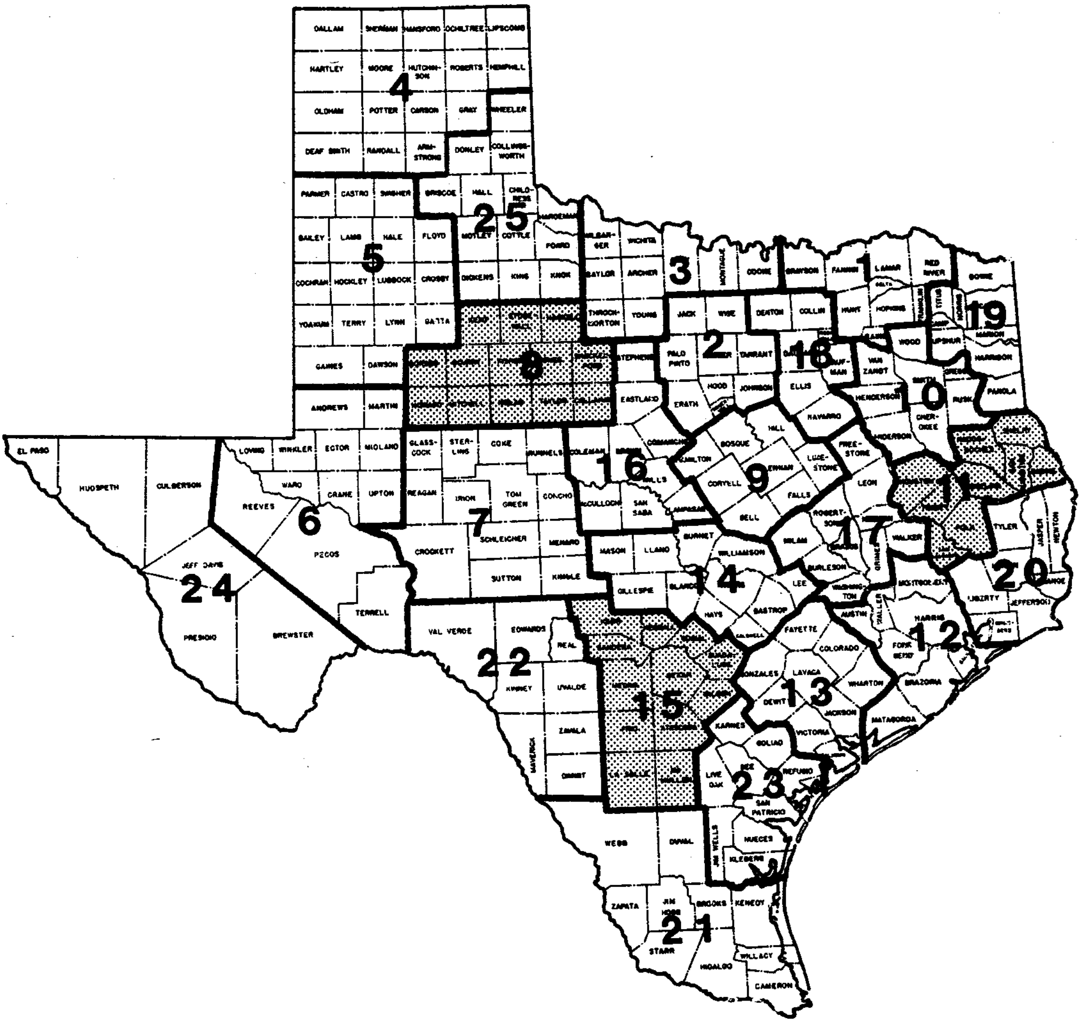


FIGURE 2. Location of 100 percent sampled districts

hence, a total cost for each observation could be computed. The five funding strategies, along with their associated costs, from which one was selected were as follows:

1. Seal coat, or fog seal, or extensive patching plus seal ($0.36/sq.yd.).
2. One inch asphaltic concrete pavement (ACP) overlay or seal plus level-up ($1.58/sq.yd.).
3. Two and one-half inch ACP overlay ($3.41/sq.yd.).
4. Four inch ACP overlay ($6.05/sq.yd.).
5. Seven and one-half inch ACP overlay ($11.93/sq.yd.).

For each of the above strategies the estimated rehabilitated Pavement Score was computed and run through deterioration calculations to determine the life expectancy. This expected life was compared to the minimum allowable expected life to determine which of the five strategies had the smallest allowable expected life, and that one was chosen as the strategy to be implemented. Given the chosen strategy, the cost of rehabilitation was finally computed.

CHAPTER III

ACCURACY OF FIVE PERCENT SAMPLE IN PREDICTING PAVEMENT CONDITION

Histograms of Pavement Scores

In order to determine the accuracy resulting from a five percent sample in predicting pavement condition, the data from the 100 percent sampled districts were first divided into fifteen groups, and then a five percent random sample was taken from each group. Observations within each of the three districts were classified by roadway system; similar classification was performed on the observations from the three districts combined into one single population. The fifteen data groups will be represented by the following notation: IH08; FM08; SH08; US08; FM11; SH11; US11; IH15; FM15; SH15; US15; IH08,11; FM08, 11,15; SH08,11,15; and US08,11,15. The first two letters indicate the roadway system and the following numbers indicate the district. Thus IH08 specifies those observations which come from Interstate Highways in District 8. (Note that District 11 contains no Interstate Highways.) Table 1 sets out the number of observations contained in the 100 percent sample and the five percent sample for each of the above named groups.

The determination of the Pavement Score distribution is of fundamental importance in this report. Appendix A shows the histograms of the Pavement Score for the fifteen groups previously described. Each group has two histograms, one for the 5 percent sample and one for the 100 percent sample. All histograms are given in Figures A-1 through A-30.

Table 1. Number of observations for 100 percent and five percent sample for each data group.

Group	100% Sample	5% Sample
IH08	154	8
FM08	938	48
SH08	270	12
US08	324	15
FM11	832	41
SH11	300	14
US11	241	10
IH15	323	15
FM15	1101	55
SH15	499	23
US15	331	17
IH08,15	477	23
FM08,11,15	1871	144
SH08,11,15	1069	49
US08,11,15	896	42

In these histograms, the Pavement Scores have been classified into twenty class intervals, each with a width of five.

A visual comparison between the 100 percent sample histogram and the five percent histogram shows a similarity between the two distributions in some data groups and a dissimilarity between the two distributions in other data groups. For example the five percent sample histogram for FM15 resembles the 100 percent sample histogram; but the IH15 five percent sample histogram is very dissimilar to its 100 percent sample counterpart. One hundred percent sample histograms for the three districts combined are similar to the five percent sample histograms for all roadway systems. An examination of this histograms shows that in those data groups with a larger number of observations, the five percent sample histogram more closely resembles the 100 percent sample histogram. Tables A-1 through A-15 of Appendix A set out the cumulative percentages of both the 100 percent sample distribution and the five percent sample distribution at each division for every data group. The absolute difference between the percentages of each of these distributions is also shown. The maximum absolute difference is denoted by an asterisk.

Confidence Bands on F(X)

In order to statistically compare the Pavement Score distribution based on the five percent sample with the Pavement Score distribution based on the 100 percent sample, a percentile of the Kolmogorov-Smirnov test statistic (15) is used. This percentile along with an empirical cumulative distribution, $S_n(X)$, can be utilized to form a $(1-\alpha)$ percent confidence band for a true cumulative distribution F(X). This means that one can be $(1-\alpha)$ percent confident that the true cumulative distribution lies within this band. As the percent of confidence is increased, the band becomes wider. Further discussion of this procedure can be found in Appendix B. Appendix B also contains Figures B-1 through B-15 which depict 80 percent confidence limits, U(X) and L(X), as well as the 100 percent sample cumulative distribution, F(X) for each data group. These limits are constructed from the five percent sample distribution.

In every data group the 100 percent sample cumulative distribution is contained within the 80 percent confidence bands. As can be seen from the figures, in data groups with a larger number of observations, the band at a given level of confidence is narrower. For example, the band for IH08, with a five percent sample size of eight, is much wider than the band for FM08, with a five percent sample size of 48. The bands for all three districts combined are all narrow. Combining all three districts leads to conclusions concerning stateside estimates on pavement condition.

Because the Kolmogorov-Smirnov procedure requires random sampling, it is assumed that the observations are independent. The actual departures from this assumption, however, are of only minor consequence. According to Conover (15) if only discrete values of the Pavement Scores are used, the confidence band is conservative. That is, the true but unknown confidence coefficient is greater than $(1-\alpha)$ percent.

Confidence Intervals on p

The binomial distribution can be used to construct confidence intervals on p, the percentage of roads falling within a particular pavement score division. Because p is not close to 1/2, use of the F distribution is made (16). Employing the distribution and the five percent sample data, 100 γ percent confidence limits for p are:

$$L = \frac{XF_{[(1-\gamma)/2],[2X,2(n-X+1)]}}{(n-X+1) + XF_{[(1-\gamma)/2],[2X,2(n-X+1)]}} \qquad (1)$$

$$U = \frac{(X+1)\ F_{[(1+\gamma)/2],[2(X+1),2(n-X)]}}{(n-X) + (X+1)\ F_{[(1+\gamma)/2],[2(X+1),2(n-X)]}} \qquad (2)$$

Tables 2, 3, 4, and 5 set out the confidence intervals for each pavement division at 50, 80 and 95 percent levels of confidence. These confidence intervals were constructed only for the three 100 percent sampled districts combined. Although these confidence intervals are dependent, each individual confidence interval will have its prescribed level of confidence.

TABLE 2. Computed Confidence Intervals on p, District 8,15, System IH

Pavement Score Division	Level of Confidence That p Lies Within Score Division		
	50%	80%	95%
0-5	(0.0125, 0.1129)	(0.0046, 0.1597)	(0.0011, 0.2215)
6-10	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
11-15	(0.0125, 0.1129)	(0.0046, 0.1597)	(0.0011, 0.2215)
16-20	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
21-25	(0.0125, 0.1129)	(0.0046, 0.1597)	(0.0011, 0.2215)
26-30	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
31-35	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
36-40	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
41-45	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
46-50	(0.0418, 0.1637)	(0.0234, 0.2161)	(0.0107, 0.2813)
51-55	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
56-60	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
61-65	(0.0418, 0.1637)	(0.0234, 0.2161)	(0.0107, 0.2813)
66-70	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
71-75	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
76-80	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
81-85	(0.0000, 0.0589)	(0.0000, 0.0959)	(0.0000, 0.1497)
86-90	(0.0125, 0.1129)	(0.0046, 0.1597)	(0.0011, 0.2215)
91-95	(0.0125, 0.1129)	(0.0046, 0.1597)	(0.0011, 0.2215)
96-100	(0.5178, 0.6907)	(0.4565, 0.7435)	(0.3896, 0.7966)

TABLE 3. Computed Confidence Intervals on p, District 8,11,15, System FM

Pavement Score Division	Level of Confidence That p Lies Within Score Division		
	50%	80%	95%
0-5	(0.0020, 0.0185)	(0.0007, 0.0264)	(0.0002, 0.0376)
6-10	(0.0067, 0.0269)	(0.0037, 0.0360)	(0.0017, 0.0484)
11-15	(0.0120, 0.0350)	(0.0077, 0.0452)	(0.0043, 0.0585)
16-20	(0.0067, 0.0269)	(0.0037, 0.0360)	(0.0017, 0.0484)
21-25	(0.0177, 0.0427)	(0.0122, 0.0541)	(0.0077, 0.0682)
26-30	(0.0020, 0.0185)	(0.0007, 0.0264)	(0.0002, 0.0376)
31-35	(0.0067, 0.0269)	(0.0037, 0.0360)	(0.0017, 0.0484)
36-40	(0.0120, 0.0350)	(0.0077, 0.0452)	(0.0043, 0.0585)
41-45	(0.0177, 0.0427)	(0.0122, 0.0541)	(0.0077, 0.0682)
46-50	(0.0177, 0.0427)	(0.0122, 0.0541)	(0.0077, 0.0682)
51-55	(0.0359, 0.0665)	(0.0281, 0.0800)	(0.0207, 0.0965)
56-60	(0.0120, 0.0350)	(0.0077, 0.0452)	(0.0043, 0.0585)
61-65	(0.0359, 0.0665)	(0.0281, 0.0800)	(0.0207, 0.0965)
66-70	(0.0596, 0.0962)	(0.0485, 0.0906)	(0.0379, 0.1288)
71-75	(0.0486, 0.0810)	(0.0395, 0.0951)	(0.0307, 0.1124)
76-80	(0.0295, 0.0583)	(0.0222, 0.0703)	(0.0156, 0.0849)
81-85	(0.0120, 0.0350)	(0.0077, 0.0452)	(0.0043, 0.0585)
86-90	(0.0412, 0.0730)	(0.0322, 0.0858)	(0.0238, 0.1016)
91-95	(0.0596, 0.0962)	(0.0485, 0.0906)	(0.0379, 0.1288)
96-100	(0.3416, 0.3927)	(0.3224, 0.4118)	(0.2991, 0.4354)

TABLE 4. Computed Confidence Intervals on p, District 8,11,15, System SH

Pavement Score Division	Level of Confidence That p Lies Within Score Division		
	50%	80%	95%
0-5	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
6-10	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
11-15	(0.0000, 0.0277)	(0.0000, 0.0457)	(0.0000, 0.0719)
16-20	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
21-25	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
26-30	(0.0353, 0.1016)	(0.0228, 0.1301)	(0.0129, 0.1667)
31-35	(0.0000, 0.0277)	(0.0000, 0.0457)	(0.0000, 0.0719)
36-40	(0.0353, 0.1016)	(0.0228, 0.1301)	(0.0129, 0.1667)
41-45	(0.0000, 0.0277)	(0.0000, 0.0457)	(0.0000, 0.0719)
46-50	(0.0196, 0.0353)	(0.0109, 0.0227)	(0.0049, 0.0128)
51-55	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
56-60	(0.0000, 0.0277)	(0.0000, 0.0457)	(0.0000, 0.0719)
61-65	(0.0693, 0.1496)	(0.0506, 0.1846)	(0.0341, 0.2283)
66-70	(0.0196, 0.0353)	(0.0109, 0.0227)	(0.0049, 0.0128)
71-75	(0.0196, 0.0353)	(0.0109, 0.0227)	(0.0049, 0.0128)
76-80	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
81-85	(0.0196, 0.0353)	(0.0109, 0.0227)	(0.0049, 0.0128)
86-90	(0.0196, 0.0353)	(0.0109, 0.0227)	(0.0049, 0.0128)
91-95	(0.0058, 0.0540)	(0.0021, 0.0765)	(0.0005, 0.1074)
96-100	(0.3714, 0.4832)	(0.3306, 0.5238)	(0.2867, 0.5676)

TABLE 5. Computed Confidence Intervals on p, District 8,11,15, System US

Pavement Score Division	Level of Confidence That p Lies Within Score Division		
	50%	80%	95%
0-5	(0.0228, 0.0919)	(0.0127, 0.1229)	(0.0058, 0.1647)
6-10	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
11-15	(0.0228, 0.0919)	(0.0127, 0.1229)	(0.0058, 0.1647)
16-20	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
21-25	(0.0000, 0.0327)	(0.0000, 0.0538)	(0.0000, 0.0855)
26-30	(0.0000, 0.0327)	(0.0000, 0.0538)	(0.0000, 0.0855)
31-35	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
36-40	(0.0000, 0.0327)	(0.0000, 0.0538)	(0.0000, 0.0855)
41-45	(0.0228, 0.0919)	(0.0127, 0.1229)	(0.0058, 0.1647)
46-50	(0.0000, 0.0327)	(0.0000, 0.0538)	(0.0000, 0.0855)
51-55	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
56-60	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
61-65	(0.0228, 0.0919)	(0.0127, 0.1229)	(0.0058, 0.1647)
66-70	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
71-75	(0.0000, 0.0327)	(0.0000, 0.0538)	(0.0000, 0.0855)
76-80	(0.0412, 0.1192)	(0.0264, 0.1536)	(0.0149, 0.1981)
81-85	(0.0000, 0.0327)	(0.0000, 0.0538)	(0.0000, 0.0855)
86-90	(0.0068, 0.0630)	(0.0025, 0.0905)	(0.0005, 0.1280)
91-95	(0.0228, 0.0919)	(0.0127, 0.1229)	(0.0058, 0.1647)
96-100	(0.4578, 0.5877)	(0.4102, 0.6345)	(0.3583, 0.6837)

CHAPTER IV

PAVEMENT SCORE METHODOLOGY

Fitting Density Functions to Data

In order to gain greater insight into pavement condition, a probability density function is fitted to the Pavement Score distribution of each data group. The beta distribution is chosen as the postulated family of density functions. The probability density function is defined as follows:

$$f(X) = \frac{1}{B(a,b)} X^{a-1} (1-X)^{b-1} \text{ for } a,b>0;\ 0<X<1 \qquad (3)$$

In Equation 3, B(a,b) is defined as

$$B(a,b) = \int_0^1 X^{a-1} (1-X)^{b-1}\, dx \qquad (4)$$

Since the random variable, X, must be in the interval zero to one, all the Pavement Scores are divided by 100 to satisfy this condition.

The parameters, a and b, are estimated by the method of moments. This procedure equates

$$\mu = E[X] \text{ to } \bar{X} = \sum_{i=1}^{n} X_i/n \qquad (5)$$

and

$$\sigma^2 = E[(X-\mu)^2] \text{ to } S^2 = \sum_{i=1}^{n} (X_i-\bar{X})^2/n. \qquad (6)$$

In the beta distribution,

$$\mu = \frac{a}{a+b} \qquad (7)$$

and

$$\sigma^2 = \frac{ab}{(a+b+1)(a+b)^2} \quad . \tag{8}$$

The mean of the 100 percent sample distribution is set equal to and the variance to σ^2. Upon solving for a and b, the following two equations result:

$$\hat{a} = \frac{\hat{\mu}^2(1-\hat{\mu})}{\hat{\sigma}^2} - \hat{\mu} \tag{9}$$

$$\hat{b} = \frac{\hat{a} - \hat{a}\hat{\mu}}{\hat{\mu}} \tag{10}$$

Table 6 sets out the estimated values of the parameters a and b for each of the data groups calculated from the 100 percent sample distribution. According to Hogg and Craig (<u>17</u>) $\hat{a}$ and $\hat{b}$ are consistent estimators of a and b.

To determine if the 100 percent sample does actually come from a beta distribution with parameters as specified in Table 6, the Kolmogorov-Smirnov goodness of fit test is used (<u>15</u>). Because of the results of this procedure, explained in Appendix C, it is assumed that the 100 percent sample does come from a beta distribution.

Comparison of Estimation Methods

Because the percentage of roads with a Pavement Score at or below 40 is of special importance to the SDHPT, two methods for predicting this percentage from random samples are compared. In the first method the percentage of roads falling at or below 40 was computed directly

TABLE 6. Estimated values of the parameters a and b of the beta distribution.

Data Group	$\hat{a}$	$\hat{b}$
IH08	0.434	0.291
FM08	1.522	0.564
SH08	0.810	0.351
US08	0.843	0.220
FM11	1.048	0.612
SH11	0.990	0.476
US11	0.467	0.369
IH15	1.226	0.161
FM15	1.844	0.304
SH15	1.038	0.217
US15	1.125	0.218

from the sample. The number of observations with a score of 40 or below in the sample was divided by the total observations in the sample. The second method made use of the cumulative beta distribution. μ and σ^2 were estimated from the sample. Making use of Equation 9 and Equation 10, a and b are then estimated. Utilizing the IMSL subroutine, MDBETA (described in Appendix C), the percentage of observations falling at or below 40 is then estimated.

These two methods are then used in a simulation computer program to determine which one would more accurately predict the percentage at or below 40 in the 21 other districts sampled. Because a five percent random sample was taken in these 21 districts, a five percent random sample is used in the simulation program.

In the simulation, 300 random samples are drawn using the IMSL subroutine GGPER. GGPER is a subroutine which generates a random permutation of a specified number of integers. (The justification of 300 samples will be discussed in the next chapter.) With each sample the estimated percentage was calculated using each methodology described above. The following equation was used to calculate the mean error (μ_{error}) for each of the two estimating methodologies:

$$\mu_{error} = \sum_{i=1}^{300} \frac{|p - \hat{p}_i|}{300} \qquad (11)$$

where

p = true percentage with a score at or below 40 calculated from the 100 percent sample, and

$\hat{p}_i$ = the estimated percentage with a score at or below 40 in the i-th iteration.

Table 7 shows the mean error for each of the estimating methodologies in each data group. As can be seen, the mean error is less for every data group when the percentage is calculated using the beta distribution. It is thus recommended that in the 21 five percent sampled districts the percentage of roads with a Pavement Score at or below 40 be estimated by use of the cumulative beta distribution, with its parameters being estimated from the sample by the method of moments.

Table 8 sets out these estimated percentages for each roadway system in the 21 five percent sampled districts. Estimates using both procedures of estimation are shown. However, because of the results of the simulation studies discussed above, it is felt that the estimate obtained from the cumulative beta distribution would be a better point estimator than the one obtained directly from the sample. Because of the very small sample sizes in the IH systems, it was often impossible to get a reasonable estimate.

Using Equation 1 and Equation 2, 95% confidence intervals were constructed for the percentage of roads with a Pavement Score of 40 or less. Table 9 sets out the upper and lower limits of these intervals for each system (except IH) in each of the 21 five percent sample districts. For example, in District 1, System FM, we can be 95% confident that the percentage of roads with a score of 40 or less lies within the interval 9.7% and 32.4%. The length of these intervals are quite large. If a shorter interval is needed with the same level of confidence (95%), a larger sample must be taken.

TABLE 7. Mean Error in Predicting Percentage of Roads Below a Pavement Score of 40.

Data Group	Mean Error Calculated Directly from Sample	Mean Error Calculated from Beta Distribution
IH08	0.175	0.154
FM08	0.036	0.031
SH08	0.098	0.080
US08	0.071	0.058
FM11	0.051	0.045
SH11	0.085	0.069
US11	0.125	0.109
IH15	0.051	0.039
FM15	0.022	0.017
SH15	0.051	0.044
US15	0.057	0.039

TABLE 8. Estimated Percentage of Roads with a Pavement Score of 40 or less in 5% Sampled Districts

District	System	Estimation Directly from Sample	Estimation Using Cumulative Beta Distribution
1	IH	--	0.27
	FM	0.19	0.21
	SH	0.31	0.20
	US	0.29	0.25
2	IH	--	0.76
	FM	0.09	0.14
	SH	0.24	0.19
	US	0.33	0.30
3	IH	--	0.92
	FM	0.08	0.06
	SH	0.00	0.07
	US	0.06	0.06
4	IH	--	--
	FM	0.00	0.01
	SH	0.07	0.06
	US	0.06	0.06
6	IH	--	0.04
	FM	0.00	0.02
	SH	0.00	0.02
	US	0.00	0.01

TABLE 8. Estimated Percentage of Roads with a Pavement Score of 40 or less in 5% Sampled Districts (Cont'd)

District	System	Estimation Directly from Sample	Estimation Using Cumulative Beta Distribution
7	IH	--	0.17
	FM	0.03	0.03
	SH	0.20	0.25
	US	0.14	0.06
9	IH	--	--
	FM	0.02	0.02
	SH	0.13	0.14
	US	0.00	0.01
10	IH	--	--
	FM	0.08	0.08
	SH	0.27	0.31
	US	0.27	0.31
12	IH	--	0.05
	FM	0.12	0.11
	SH	0.33	0.29
	US	0.00	0.14
13	IH	--	--
	FM	0.18	0.16
	SH	0.19	0.21
	US	0.25	0.28

TABLE 8. Estimated Percentage of Roads with a Pavement Score of 40 or less in 5% Sampled Districts (Cont'd)

District	System	Estimation Directly from Sample	Estimation Using Cumulative Beta Distribution
14	IH	--	0.00
	FM	0.10	0.08
	SH	0.17	0.16
	US	0.27	0.29
17	IH	--	0.72
	FM	0.32	0.31
	SH	0.08	0.08
	US	0.00	0.06
18	IH	--	--
	FM	0.27	0.21
	SH	0.60	0.42
	US	1.00	--
19	IH	--	0.08
	FM	0.14	0.12
	SH	0.30	0.29
	US	0.36	0.29
20	IH	--	--
	FM	0.08	0.07
	SH	0.42	0.35
	US	0.17	0.13

TABLE 8. Estimated Percentage of Roads with a Pavement Score of 40 or less in 5% Sampled Districts (Cont'd)

District	System	Estimation Directly from Sample	Estimation Using Cumulative Beta Distribution
21	IH	--	0.08
	FM	0.10	0.09
	SH	0.30	0.28
	US	0.10	0.13
23	IH	--	--
	FM	0.08	0.13
	SH	0.20	0.25
	US	0.21	0.26
24	IH	--	0.17
	FM	0.00	0.09
	SH	0.00	0.02
	US	0.00	0.07
25	IH	--	--
	FM	0.22	0.17
	SH	0.15	0.11
	US	0.14	0.09

TABLE 9. 95% Confidence Intervals on Percentage of Roads with a Pavement Score of 40 or less

District	System	Lower Limit	Upper Limit
1	FM	0.097	0.324
	SH	0.122	0.606
	US	0.037	0.709
2	FM	0.017	
	SH	0.068	0.504
	US	0.100	0.657
3	FM	0.018	0.243
	SH	0.000	0.265
	US	0.001	0.277
4	FM	0.000	0.086
	SH	0.002	0.339
	US	0.001	0.277
6	FM	0.000	0.089
	SH	0.000	0.160
	US	0.000	0.207
7	FM	0.001	0.138
	SH	0.043	0.481
	US	0.018	0.428

TABLE 9. 95% Confidence Intervals on Percentage of Roads with a Pavement Score of 40 or less (Cont'd)

District	System	Lower Limit	Upper Limit
9	FM	0.001	0.118
	SH	0.015	0.383
	US	0.000	0.410
10	FM	0.021	0.191
	SH	0.106	0.503
	US	0.079	0.119
12	FM	0.025	0.346
	SH	0.075	0.701
	US	0.000	0.708
13	FM	0.079	0.335
	SH	0.041	0.457
	US	0.055	0.572
14	FM	0.028	0.249
	SH	0.021	0.484
	US	0.060	0.609
17	FM	0.189	0.476
	SH	0.002	0.385
	US	0.000	0.285

TABLE 9. 95% Confidence Intervals on Percentage of Roads with a Pavement Score of 40 or less (Cont'd)

District	System	Lower Limit	Upper Limit
18	FM	0.107	0.556
	SH	0.147	0.947
	US	0.025	1.000
19	FM	0.047	0.318
	SH	0.067	0.653
	US	0.109	0.686
20	FM	0.010	0.272
	SH	0.151	0.723
	US	0.004	0.641
21	FM	0.028	0.249
	SH	0.135	0.536
	US	0.020	0.256
23	FM	0.017	0.234
	SH	0.025	0.583
	US	0.046	0.508
24	FM	0.000	0.241
	SH	0.000	0.285
	US	0.000	0.318

TABLE 9. 95% Confidence Intervals on Percentage of Roads with a Pavement Score of 40 or less (Cont'd)

District	System	Lower Limit	Upper Limit
25	FM	0.095	0.438
	SH	0.019	0.484
	US	0.018	0.428

CHAPTER V

EVALUATION OF THE ESTIMATOR OF THE MEAN COST PER SQUARE YARD

Description of Estimators

After a sample is taken a decision must be made as to what statistic (estimator) will be calculated from the data. In this thesis that statistic is defined as the best estimator of the mean cost per square yard. The estimator which minimizes the mean squared error most frequently over the varying sample sizes will be chosen.

Although the average length of segments is two miles, it varies from 0.3 to 3 miles. Widths also vary substantially among road systems. Because of this variability of size among the observations, an auxiliary variable, the area of observation, was collected. The following three estimators are investigated to predict mean cost per square yard:

(1) Mean of Ratio: $$\hat{R} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i} \tag{12}$$

(2) Ratio of Means: $$\hat{r} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \tag{13}$$

(3) Ratio of Means: $\hat{r}_s = \frac{n_c}{n} \frac{\sum_{i=1}^{n_c} Y_{ci}}{\sum_{i=1}^{n_c} X_{ci}}$ (14)

(Separate Estimator-
Post-Stratified)

where

Y_i = total cost of maintenance or rehabilitation for observation i,

X_i = total square yards in observation i,

Y_{ci} = total non-zero cost of observation i,

X_{ci} = total square yards in i-th observation with non-zero cost,

n_c = number of observations in sample with non-zero cost, and

n = number of observations sampled.

Simulation Procedures

In order to determine which of the above named estimators most accurately predict mean cost per square yard, simulation procedures will be employed on each of the data groups. The data groups with all three 100 percent sample districts combined will not be examined.

Working with one data group at a time, a computer program was prepared to randomly select observations and compute the squared difference between the computed value of each of the estimators and the true mean cost per square yard. Table 10 shows the true mean cost per square yard for each of the data groups.

TABLE 10. True Mean Cost per Square Yard for Maintenance and Rehabilitation

Data Group	Mean Cost
IH08	$1.35
FM08	0.14
SH08	0.40
US08	0.30
FM11	0.37
SH11	0.62
US11	1.25
IH15	0.46
FM15	0.12
SH15	0.21
US15	0.40

This selection process was computerized because 300 samples are to be randomly selected and the mean squared error (MSE) computed for each of the three estimators. Three hundred iterations were used because this was the point where the estimated MSE seemed to stabilize with a five percent sample from all data groups combined. Since all three estimators are biased, the MSE was chosen as the selection criterion since it includes the variance of the estimator as well as its bias. The MSE is defined as follows:

$$\text{MSE} = E[(\hat{\mu}-\mu)^2] \tag{15}$$

where

μ = true mean cost per square yard, and

$\hat{\mu}$ = estimated mean cost per square yard.

Equation 15 can be estimated for each estimator by:

$$\widehat{MSE} = \frac{\sum_{j=1}^{k} (\hat{\mu}_j - \mu)^2}{K} \qquad (16)$$

where

K = number of iterations (300),

μ = true mean cost per square yard, and

$\hat{\mu}_j$ = estimated mean cost per square yard in iteration j.

In each iteration j the value of μ_j is computed using the three estimators defined in Equation 12, Equation 13, and Equation 14; namely mean of ratios, ratio of means and ratio of means (post-stratified, separate estimator). This entire procedure is repeated for sample sizes of 5, 10, 20 and 30 percent of the total observations in each data group.

Tables D-1 through D-3 in Appendix D set out the estimated MSE for each of the estimators at the various sample sizes. Figures D-1 through D-11 show graphicaly the relationship between sample size and estimated MSE for each of the three estimators for each data group. As can be seen from the tables and figures the mean of ratios minimizes the mean squared error most frequently over the varying sample sizes for every data group except US08 (Figure D-4) and FM11 (Figure D-5). In these two groups, however, the MSE of the mean of ratios is only slightly larger than the MSE computed using ratio of means (separate estimator). Because of these findings the mean of ratios is chosen as the estimator of the mean cost per square yard for maintenance and rehabilitation.

CHAPTER VI

SAMPLE SURVEY DESIGN

The accuracy of any estimate made from a sample depends not only on the method (statistic) by which the estimate was calculated but also on the sample survey design. Three designs are investigated in this research:

(1) Simple random sampling with district stratification,

(2) Simple random sampling with county stratification, and

(3) Systematic sampling with district stratification.

The final objective in each survey design is to find estimates for each district and the entire state. A methodology combining district results to produce state-wide results will be discussed in the next chapter. Each of the roadway systems (IH, FM, SH and US) represents a separate population in the proposed analysis. According to the results presented in the previous chapter, the mean cost per square yard will be estimated by the mean of the ratios in every sample design.

Description of the Sample Designs

A. Simple Random Sampling with District Stratification

A sample of n "two-mile" observations is selected from each of the eleven data groups. For the Interstate system, and in some instances for the SH and US systems, the selection process is not entirely random. In the case of two-roadway segments, when one observation is chosen at random on one roadway, a second observation must be chosen on the other roadway (See Figure 1a). This process was done in order to simulate the process as it is actually carried out by the SDHPT surveyors.

B. Simple Random Sampling with County Stratification

In simple random sampling with county stratification the district population composed of N observations is first divided into L nonoverlapping county populations with N_h observations in the h-th population, h=1, 2,...,L. Each county population is considered to be a stratum. Random samples of size $n_1, n_2, \ldots, n_L$ are then drawn from these strata. A process similar to that described in simple random sampling with district stratification is followed for two-roadway segments.

Stratification may produce a gain in accuracy in the estimates of district mean cost per square yard for maintenance and rehabilitation. By dividing a heterogeneous district into internally homogeneous counties, a more accurate estimate can be obtained.

The estimate for the district mean cost can be computed according to two different methods:

Method A: $$\hat{R}_A = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i} \tag{17}$$

Method B: $$R_B = \frac{1}{N} \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{Y_{hi}}{X_{hi}} \tag{18}$$

where

Y_i = total cost of maintenance or rehabilitation for observation i in a given district,

X_i = total square yards in observation i in the district,

Y_{hi} = total cost of maintenance or rehabilitation for observation i, county h, in the district,

X_{hi} = total square yards in observation i, county h, in the district,

n = total number of sampled observations in the district,

n_h = total number of sampled observations in county h,

N_h = total number of observations in county h,

N = total number of observations in the district, and

L = number of counties in the district.

C. Systematic Sampling with District Stratification

In systematic sampling an observation is taken at random from the first k observations and every k-th observation is sampled thereafter. In this sample design when an observation in one roadway is selected, the observation in the adjacent roadway is not chosen in the case of two-roadway segments.

Intuitively, systematic sampling seems likely to be more precise than simple random sampling. In effect, it stratifies the population into n strata, which consist of the first k units, the second k units, and so on. It might therefore be expected that the systematic sample would be about as precise as the corresponding stratified random sample with one unit per stratum. The difference is that with the systematic sample the observations occur at the same relative position in the stratum, whereas with the stratified random sample the position in the stratum is determined separately by randomization within each stratum. The systematic sample is spread more evenly over the population, and this fact has sometimes made systematic sampling considerably more accurate than stratified random sampling. Systematic sampling is precise when observations within the same sample are heterogeneous and it is imprecise when they are homogeneous.

Simulation Process for Cost Estimation

Simulation models are employed for each of the three sample designs, previously discussed in order to determine the most efficient procedure for estimating rehabilitation or maintenance costs. In the simulation methodology each district and each system (eleven data groups) will be examined separately. Sample sizes ranging from five percent of the total number of observations in each data group to 50 percent, considering five percent increments, are examined. However, in the sampling plans using stratification by counties, a minimum of two observations per county was specified. In some cases this caused the sample size to be slightly higher than that corresponding to the five percent increments. After the selection of each sample, the mean of ratio estimator is used to estimate the mean cost per square yard for maintenance and rehabilitation.

In simple random sampling with district stratification and both methods of simple random sampling with county stratification [Method A, Equation 17 and Method B, Equation 18] 300 samples are selected to compute the various statistics. Three hundred runs of the simulation model are needed to reach the point at which the mean squared error stabilizes, as explained in the previous chapter. In the case of systematic sampling, a reasonable finite number of possible samples exists, and the actual expected values can be directly computed.

The statistics to be computed by the simulation model are:

(1) Mean squared error ($\widehat{MSE}$)

(2) Mean of $\hat{R}$, estimator of cost per square yard, ($\hat{\mu}_{\hat{R}}$),

(3) Variance of $\hat{R}$ ($\hat{\sigma}^2\hat{R}$),

(4) Mean of the error ($\hat{\mu}_{error}$), and

(5) Variance of the error ($\hat{\sigma}^2_{error}$).

Equations 19 through 24 show how these statistics were estimated from the simulation procedures with simple random sampling with district stratification and simple random sampling with county stratification (methods A and B). When using systematic sampling, exact values for the statistics can be calculated. Instead of 300 iterations the sampling process is continued until every possible sample has been chosen. The number of possible samples is given by k=N/n. Therefore, to find the formulas for systematic sampling, simply replace 300 in Equations 19 through 24 with k.

The mean squared error is estimated as in the previous chapter. That is,

$$\widehat{MSE} = \frac{\sum_{j=1}^{300} (\hat{R}_j - \mu)^2}{300} \qquad (19)$$

where

μ = true mean cost per square yard, and

$\hat{R}_j$ = estimated mean cost per square yard in iteration j.

The mean of the estimator (R) was computed as follows:

$$\hat{\mu}_{\hat{R}} = \frac{\sum_{j=1}^{300} \hat{R}_j}{300} \tag{20}$$

where

$\hat{R}_j$ = estimated mean cost per square yard in iteration j.

The variance of R was computed thus:

$$\hat{\sigma}_R^2 = \frac{\sum_{j=1}^{300} (\hat{R}_j - \hat{\mu}_{\hat{R}})^2}{300} \tag{21}$$

where $\hat{R}_j$ and $\hat{\mu}_{\hat{R}}$ are defined as above.

The error of the estimate in iteration j is defined as follows:

$$\text{ERROR}_j = \frac{|\hat{R}_j - \mu|}{\mu} \tag{22}$$

where R_j and μ are defined as above.

The mean of the error was calculated thus:

$$\hat{\mu}_{error} = \frac{\sum_{j=1}^{300} \text{ERROR}_j}{300} \tag{23}$$

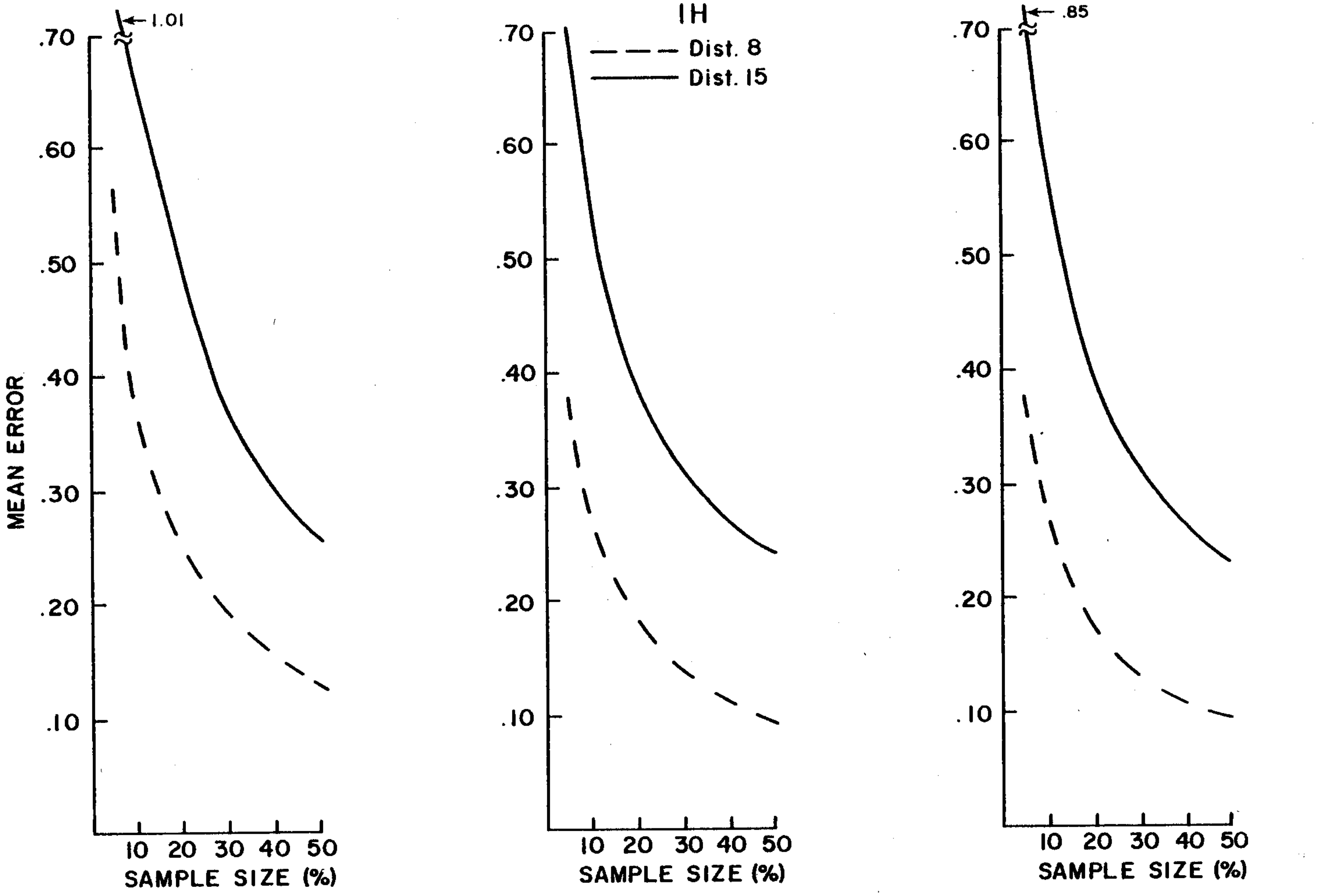
The variance of the error was calculated as follows:

$$\sigma^2_{error} = \frac{\sum_{j=1}^{300} (\text{ERROR}_j - \hat{\mu}_{error})^2}{300} \tag{24}$$

Tables E-1 through E-12 of Appendix E set out these computed statistics for various sample sizes. Each table represents a particular sample design for a particular 100 percent sampled district.

As can be seen in Tables E-1 through E-9, all of the statistics (excluding $\hat{\mu}_{\hat{R}}$) decrease as the sample size increases, as expected. However, in Tables E-10 through E-12 this is not the case. The statistics fluctuate even as the sample size is increased. Cochran (3) states that the performance of systematic sampling in relation to that of stratified or simple random sampling is greatly dependent on the properties of the population. There are population for which systematic sampling is extremely accurate and others for which it is less accurate than simple random sampling. For some populations and some values of n, the variance of the estimator may even increase when a larger sample is taken. Results of this study indicate that this population does not lend itself to this particular type of systematic sampling. In each data group there exist a few large-valued outliers. When two or more of these outliers fall within one sample, the variance of the units within the sample is inflated, and a larger than expected mean squared error results.

In Figures 3, 4, 5, and 6 the results of the different sampling plans (excluding systematic sampling) are plotted for each district. Each sampling plan is plotted as a separate graph. Hence the results as they differ from district to district can be seen. The estimated mean error ($\hat{\mu}_{error}$) is plotted against the various sample sizes. As stated above, as the sample size increases, the mean error decreases.

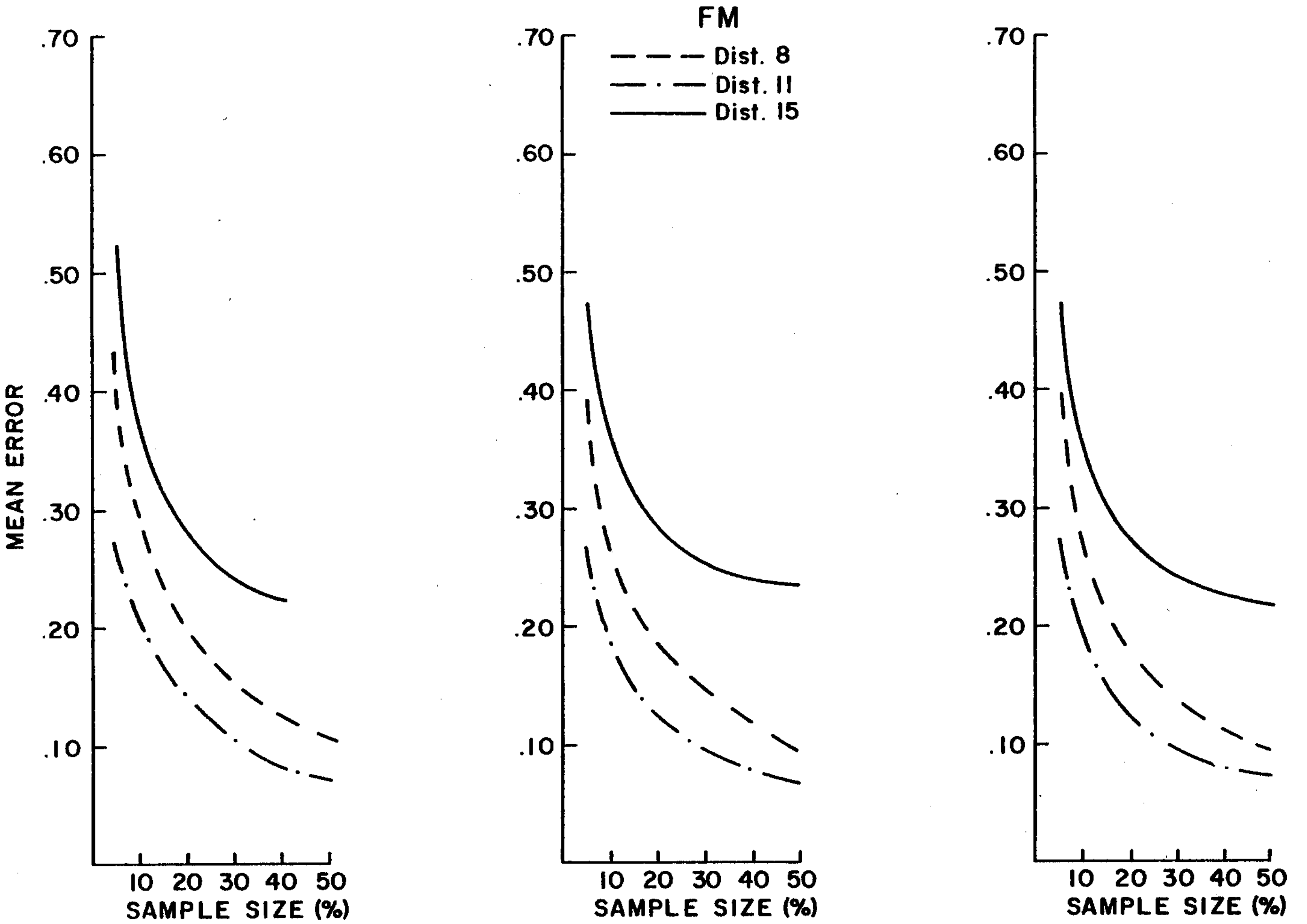


District Stratification

County Stratification (Method A)

County Stratification (Method B)

FIGURE 3. Comparison of district cost estimation results using various sample designs, System IH

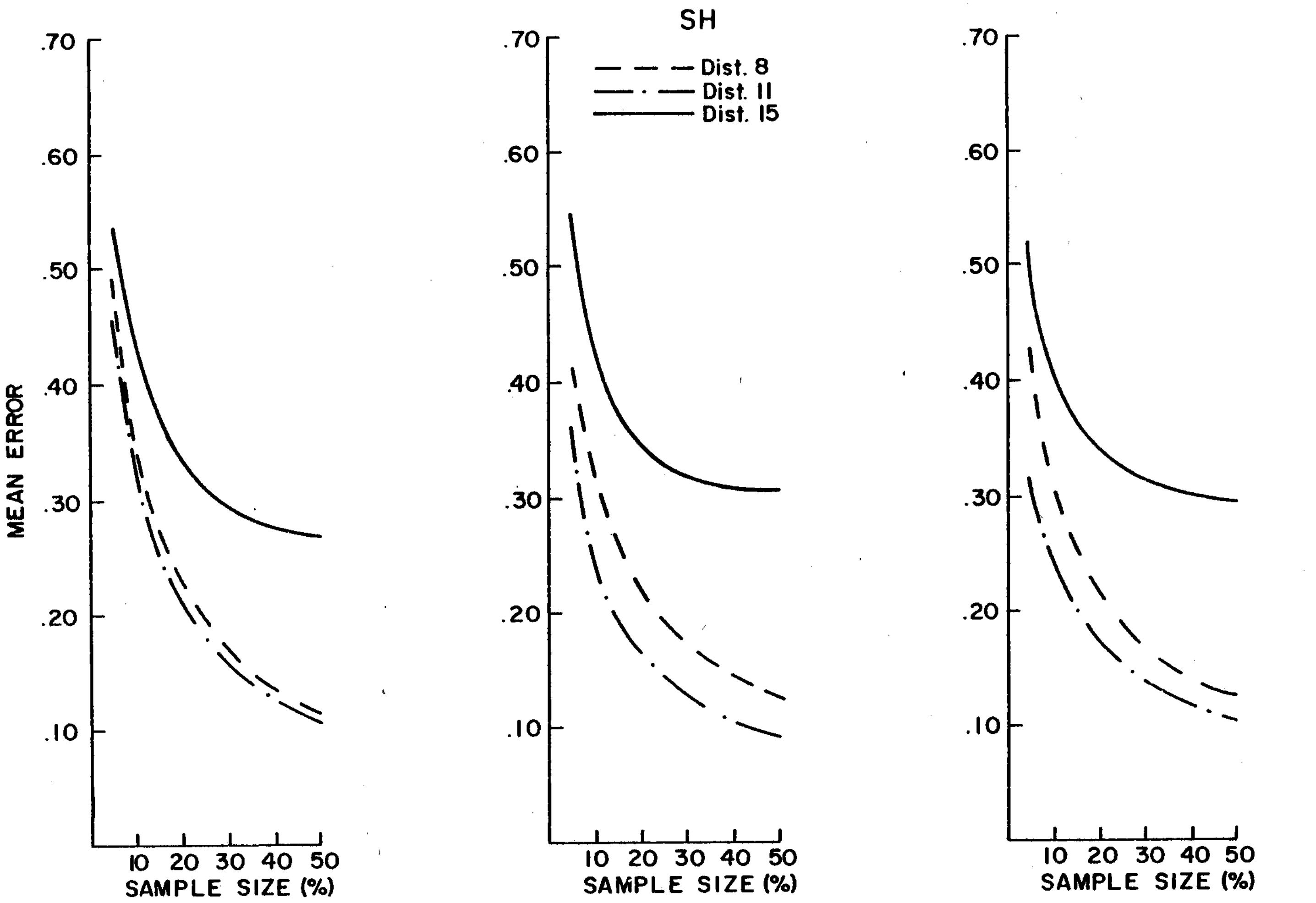


District Stratification

County Stratification (Method A)

County Stratification (Method B)

FIGURE 4. Comparison of district cost estimation results using various sample designs, System FM

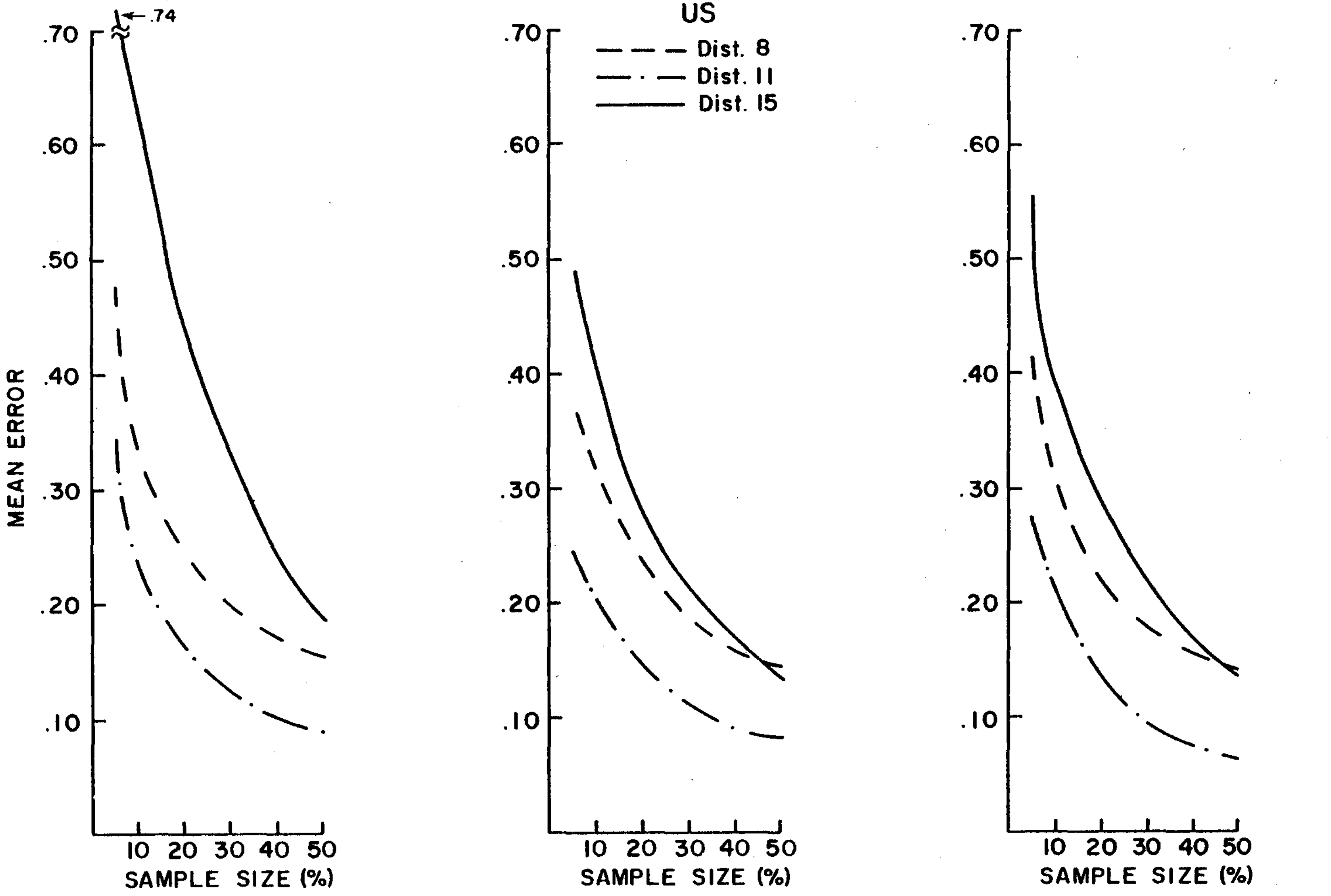


District Stratification

County Stratification (Method A)

County Stratification (Method B)

FIGURE 5. Comparison of district cost estimation results using various sample designs, System SH



District Stratification

County Stratification (Method A)

County Stratification (Method B)

FIGURE 6. Comparison of district cost estimation results using various sample designs, System US

Also note that in every case District 15 has the highest mean error over all sample sizes, followed by District 11. The mean error for District 8 is least over all sample sizes. These results seem to be related to the percentage of observations with zero cost associated with them. Table 11 sets out these percentages for each of the eleven data groups.

TABLE 11. Percentage of observations with zero costs.

System District	IH	FM	SH	US
8	61.69	88.38	75.19	85.49
11	--	78.13	78.33	58.51
15	92.26	96.19	89.58	88.52

As can be seen in Table 11, District 15 has the highest percentage with zero costs, followed by District 8 and District 11 (except SH). The data tends to suggest that the greater the percentage of roads with zero costs, the higher the estimated mean error of prediction. Hence the better the condition of roads within a district, the more difficult it will be to predict a mean cost per square yard.

From Figures 3, 4, 5, and 6, it can be noted that substantial reductions in the mean error correspond to low sample size, but after a certain state (35 to 40 percent) the decrease in the sampling error becomes marginal and incommensurate with the increase in the effort required for having a larger sample.

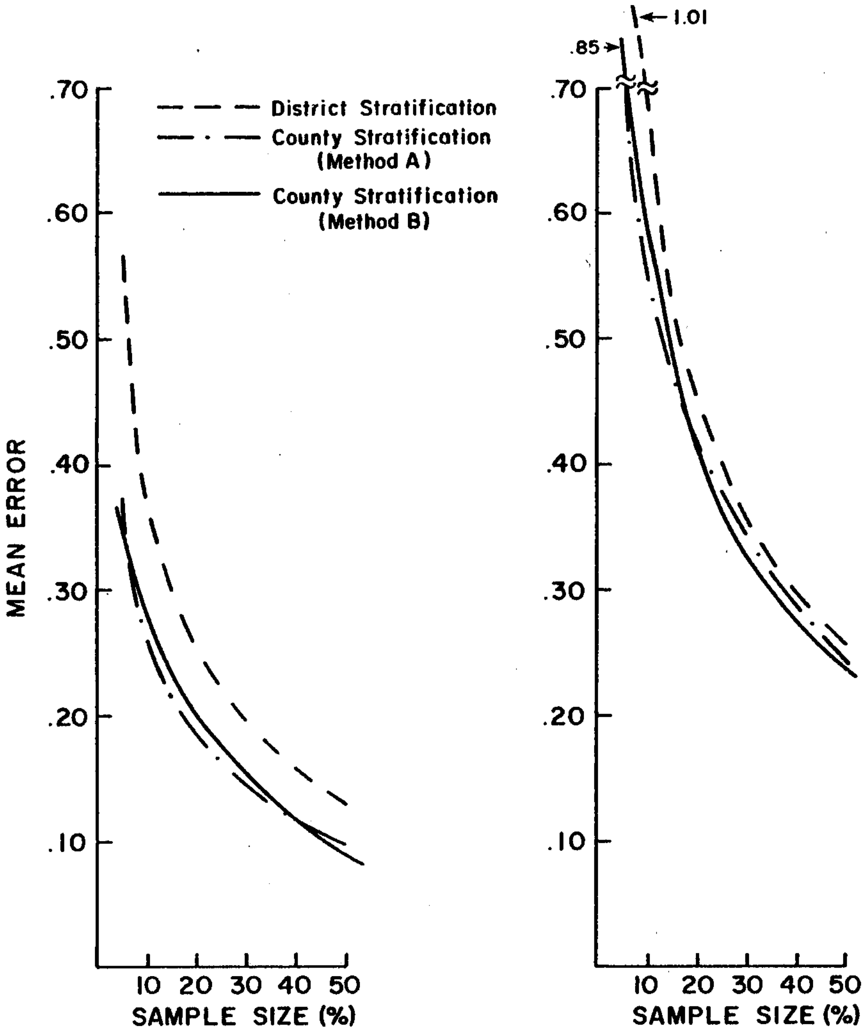
Figures 7, 8, 9, and 10 show the relationship between the error and sample size for the three sampling designs within each of the data groups. In these figures, a comparison of sample design can be made readily. Consistently, both Methods A and B (Equations 17 and 18) for design with county stratification produce a smaller mean error over all sample sizes than the design with district stratification (except SH15). Methods A and B of the design with county stratification produce very similar results. Cochran (3) states that the estimator resulting from Method A coincides with that resulting from Method B if proportional allocation is used. Proportional allocation results when

$$\frac{n_h}{n} = \frac{N_h}{N} \quad \text{or} \quad \frac{n_h}{N_h} = \frac{n}{N} .$$

Proportional allocation is used in all counties, except those with a very small number of observations. Hence it would be expected that Methods A and B would produce very similar results.

Simulation Procedure for Estimation of Pavement Condition

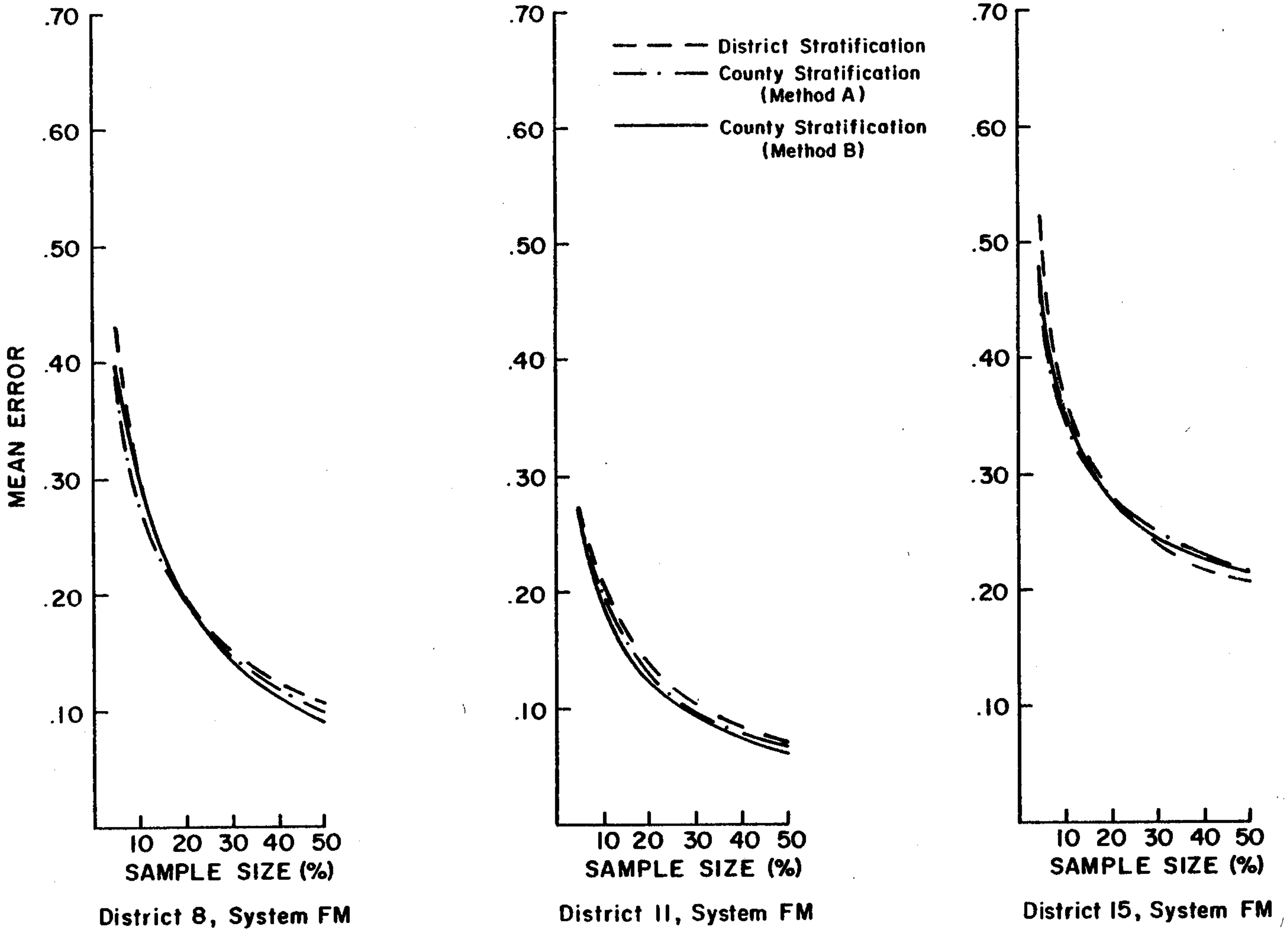
Simulation procedures were also employed to ascertain the accuracy of predicting pavement condition using simple random sampling with district stratification and both Methods A and B of simple random sampling with county stratification. Since the mean Pavement Score of a data group entails little information about the overall pavement condition of a data group, some other statistic must be used to determine the accuracy of the sample.



District 8, System IH

District 15, System IH

FIGURE 7. Comparison of sample designs used to estimate costs, System IH

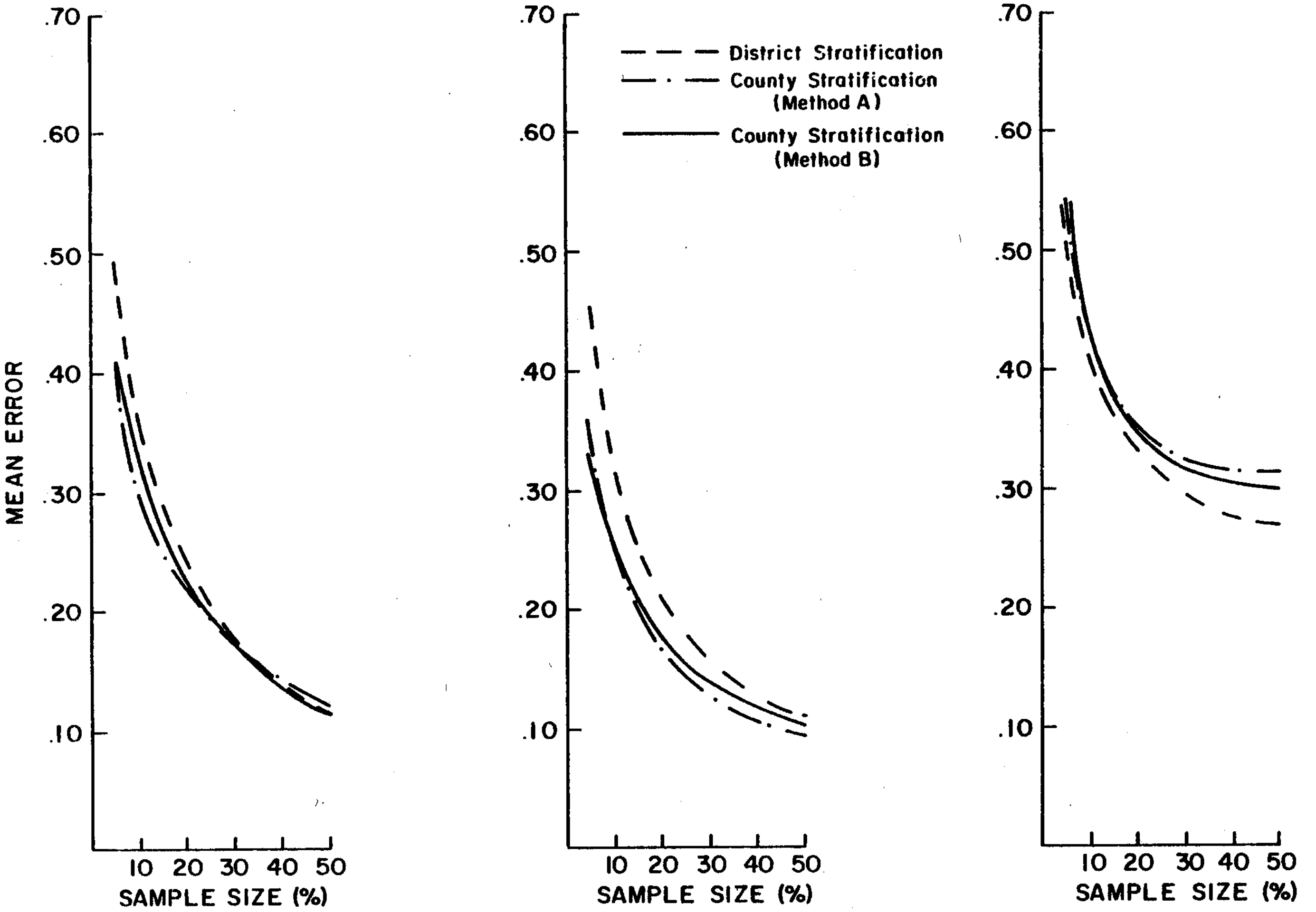


District 8, System FM

District 11, System FM

District 15, System FM

FIGURE 8. Comparison of sample designs used to estimate costs, System FM

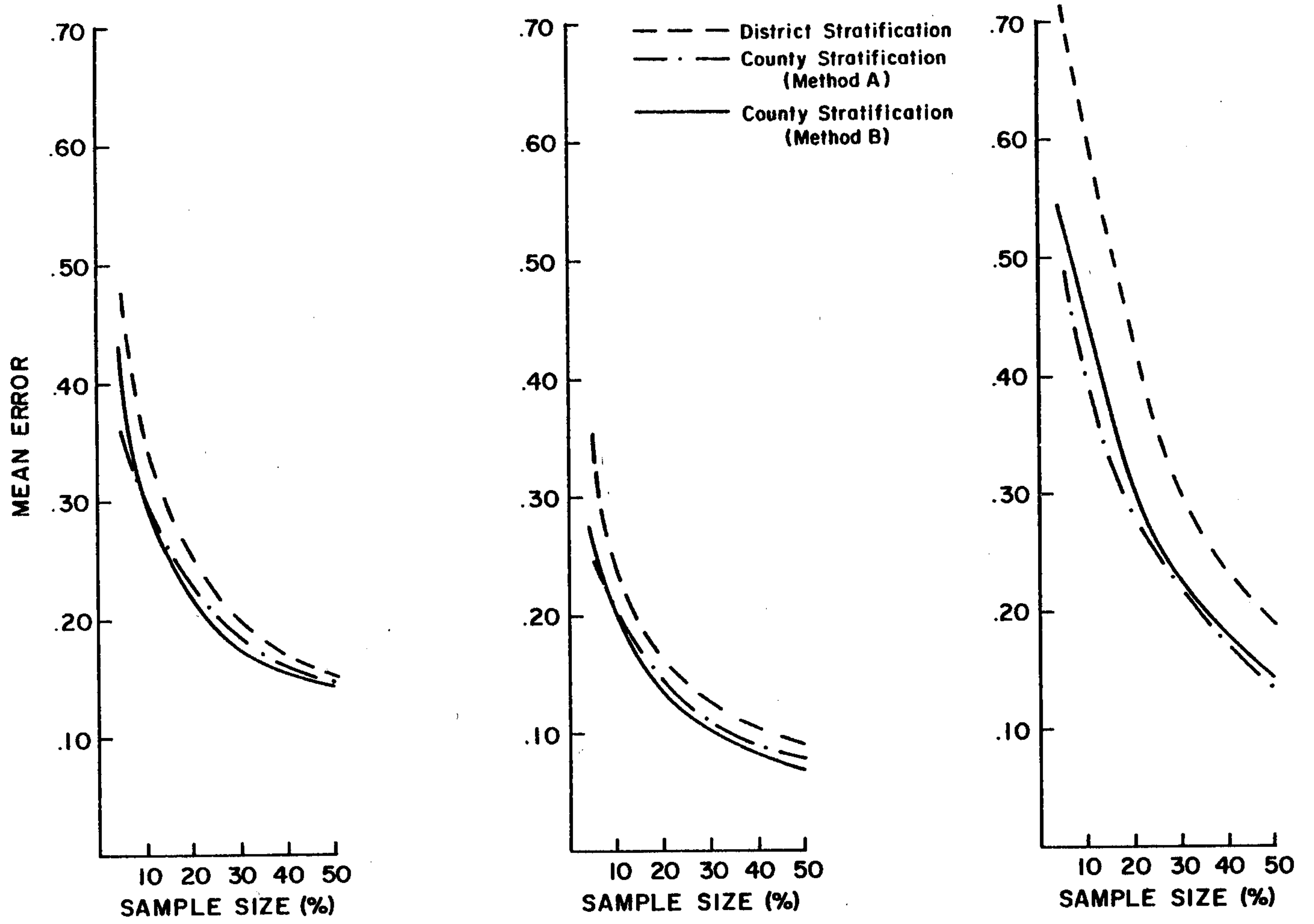


District 8, System SH

District 11, System SH

District 15, System SH

FIGURE 9. Comparison of sample designs used to estimate costs, System SH



District 8, System U.S.

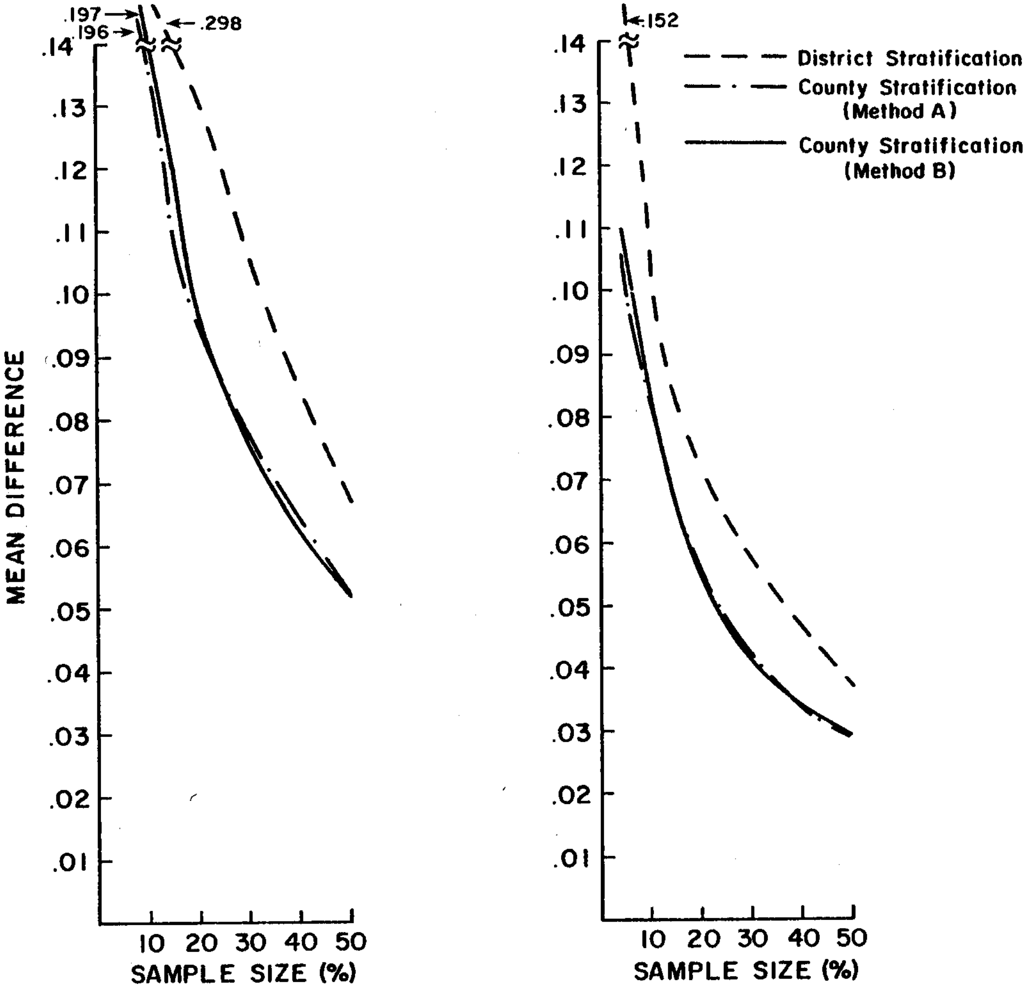
District 11, System U.S.

District 15, System U.S.

FIGURE 10. Comparison of sample designs used to estimate costs, System US

The statistic chosen is the mean absolute maximum difference between the cumulative empirical distribution and the 100 percent sampled cumulative distribution. Again 300 runs of the simulation model are made at each of the ten sample sizes discussed previously. With each run the value of the cumulative distribution is calculated at each of the twenty divisions (5,10,...,100). Each of these values was then subtracted from the corresponding value of the 100 percent sample cumulative distribution and the maximum absolute difference then taken. Thus after 300 iterations a mean of the 300 maximum differences was calculated along with the variance of the absolute maximum difference. In Appendix E, Tables E-13 through E-21 show these statistics for each type of sample design. Figures 11, 12, 13, and 14 show the relationship between the mean maximum difference and the sample size. Again, in every instance as the sample size increases, the maximum difference decreases. In these figures the three sample designs can be easily compared. As can be seen, stratification over counties produce smaller differences than stratification over districts. Methods A and B of the stratification over counties design produce very similar results.

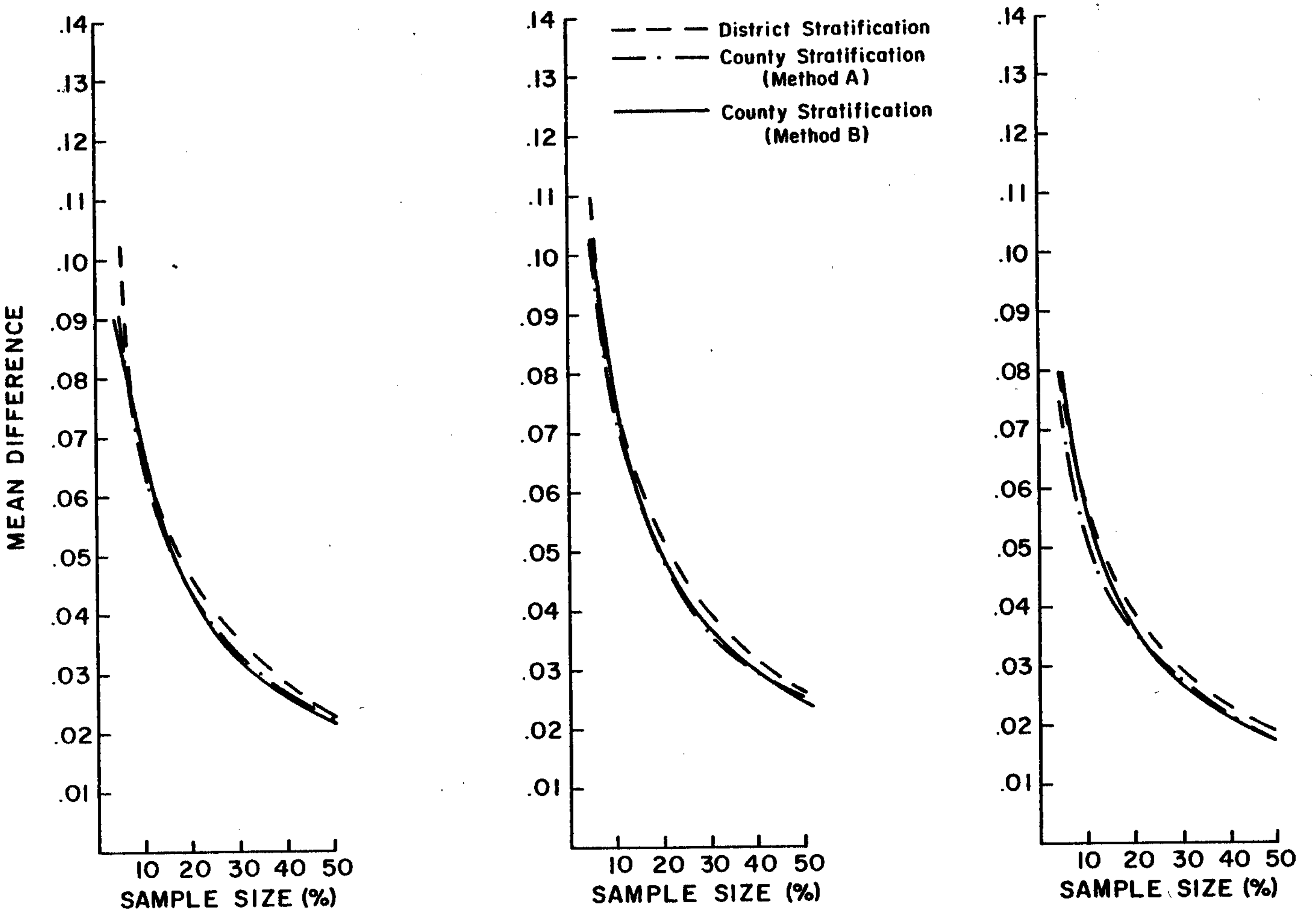
Figures 15, 16, 17, and 18 show the relationship of the mean maximum difference and the sample size using county stratification method A. In these figures a comparison of the district results can be easily compared. In every case except System SH, District 15 has the lowest maximum difference followed by District 8, then District 11. This relationship corresponds identically to the variance of the Pavement Score for each district and roadway system. The results



District 8, System IH

District 15, System IH

FIGURE 11. Comparison of sample designs used to predict pavement condition, System IH

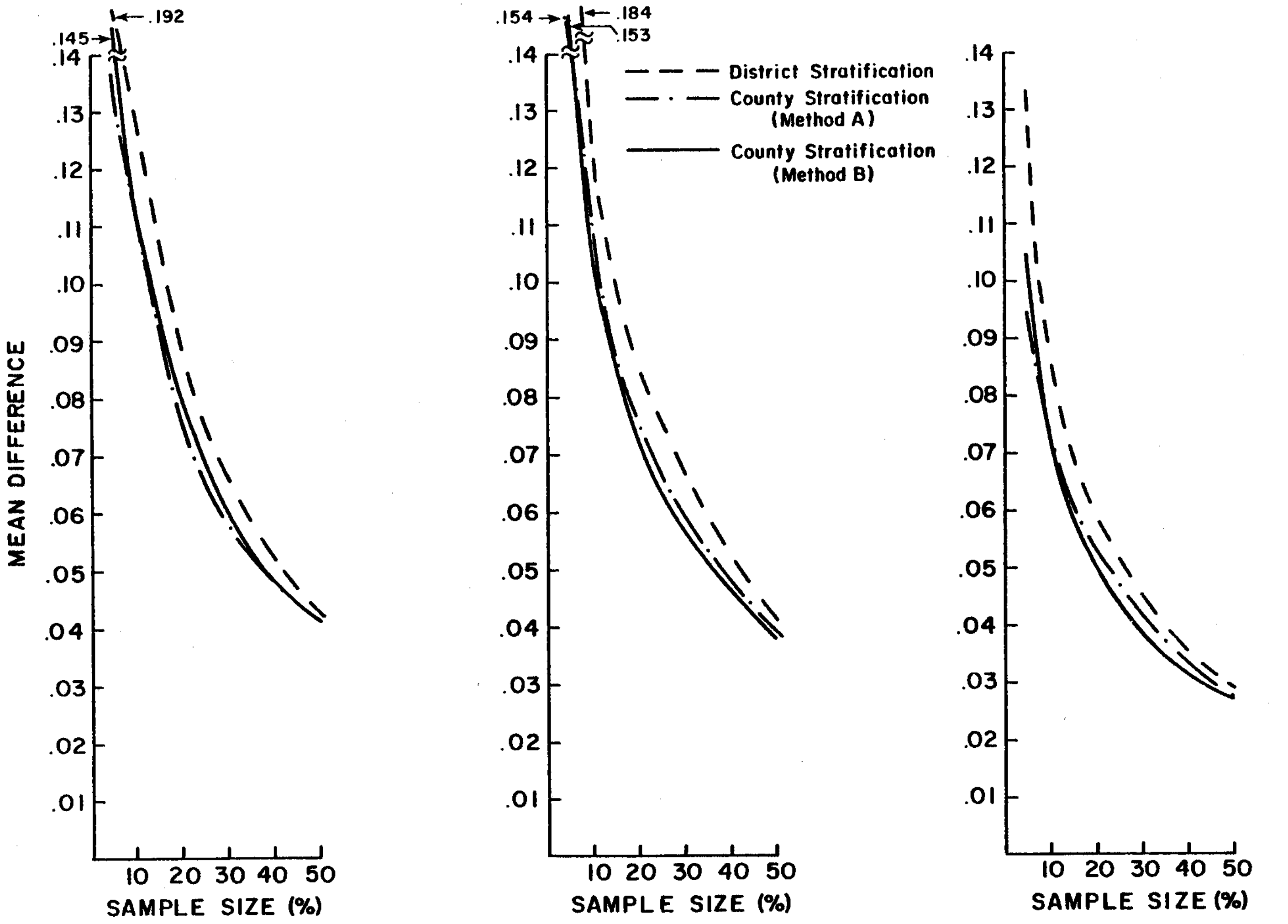


District 8, System FM

District 11, System FM

District 15, System FM

FIGURE 12. Comparison of sample designs used to predict pavement condition, System FM

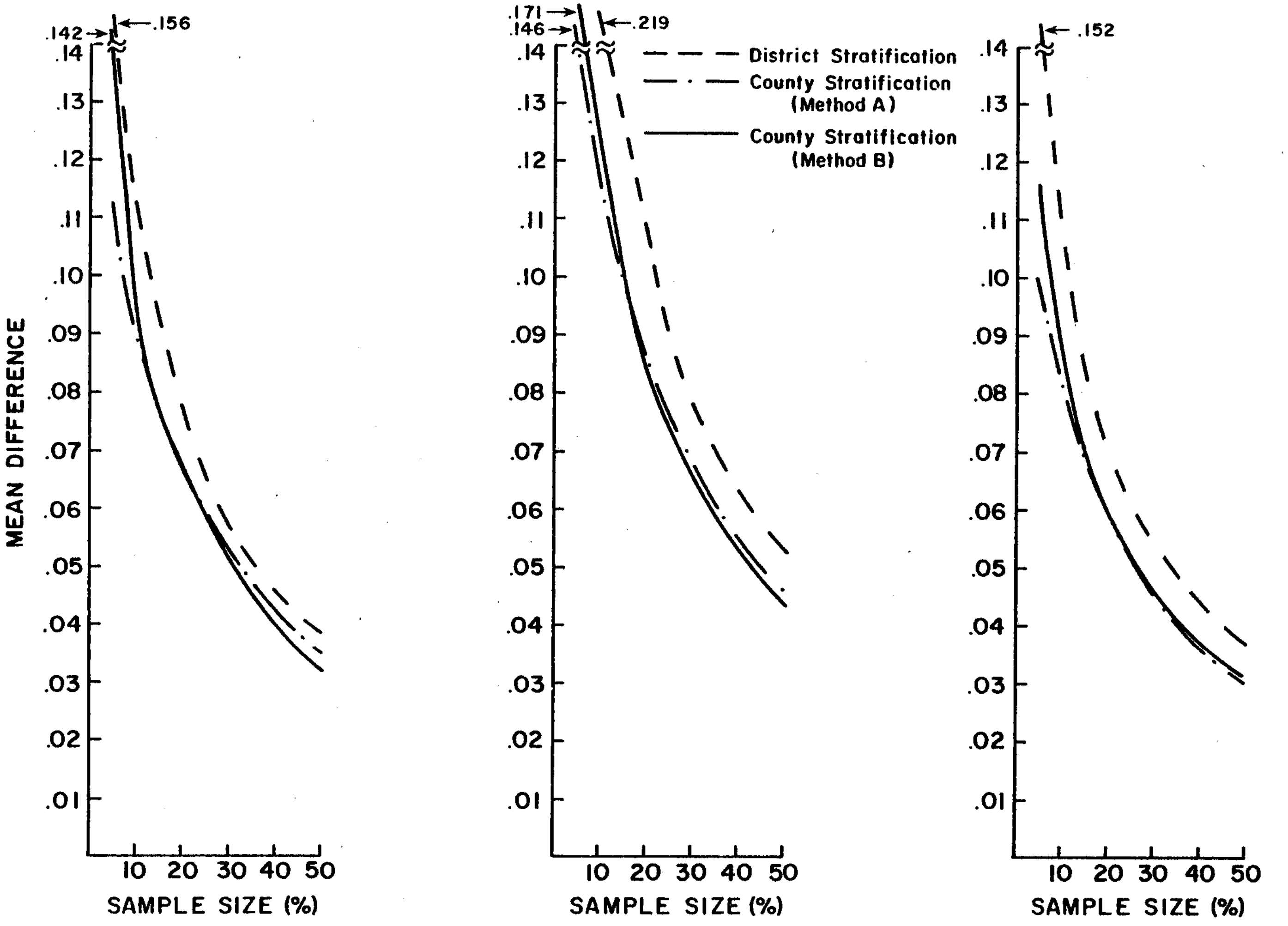


District 8, System SH

District 11, System SH

District 15, System SH

FIGURE 13. Comparison of sample designs used to predict pavement condition, System SH



District 8, System US

District 11, System US

District 15, System US

FIGURE 14. Comparison of sample designs used to predict pavement condition, System US

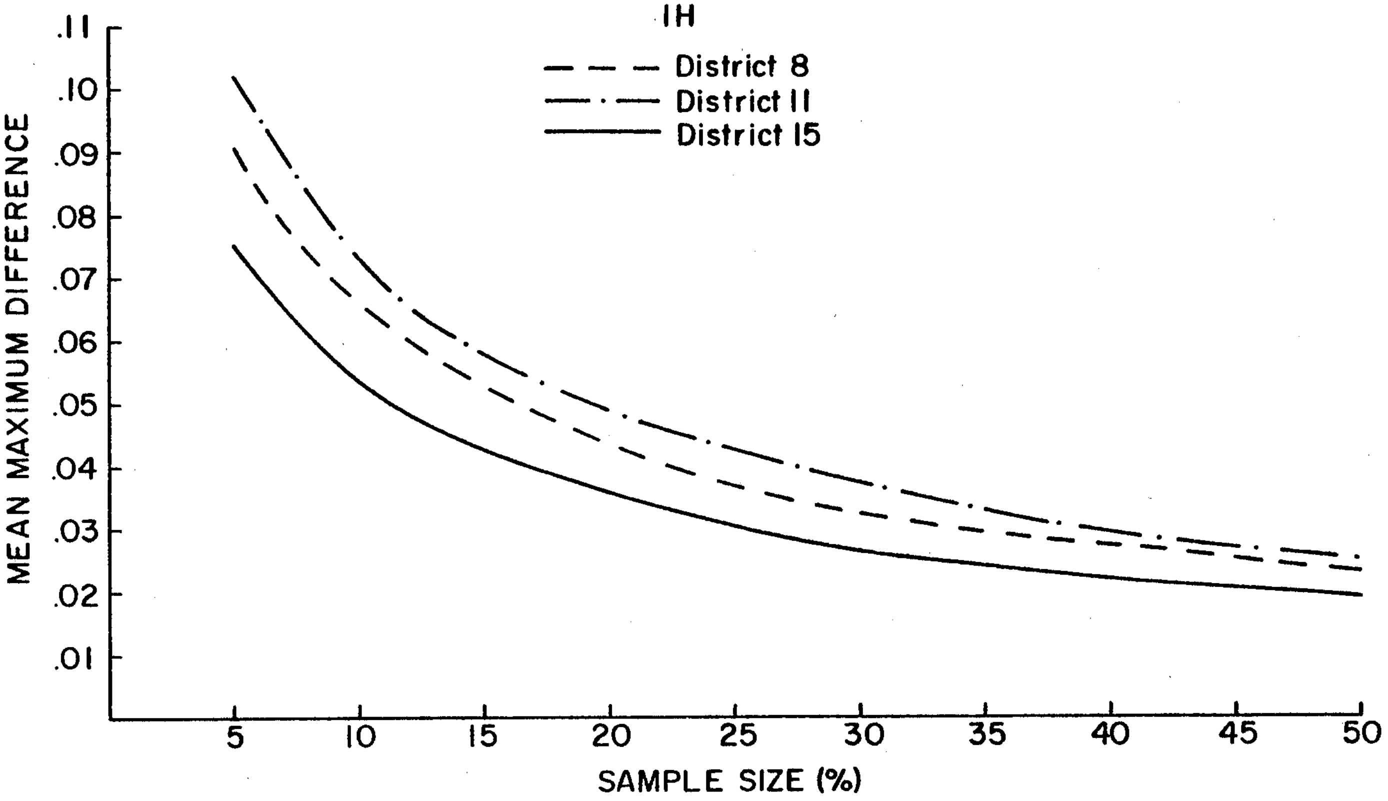


FIGURE 15. Comparison of district results of the mean maximum difference using county stratification method A, System IH

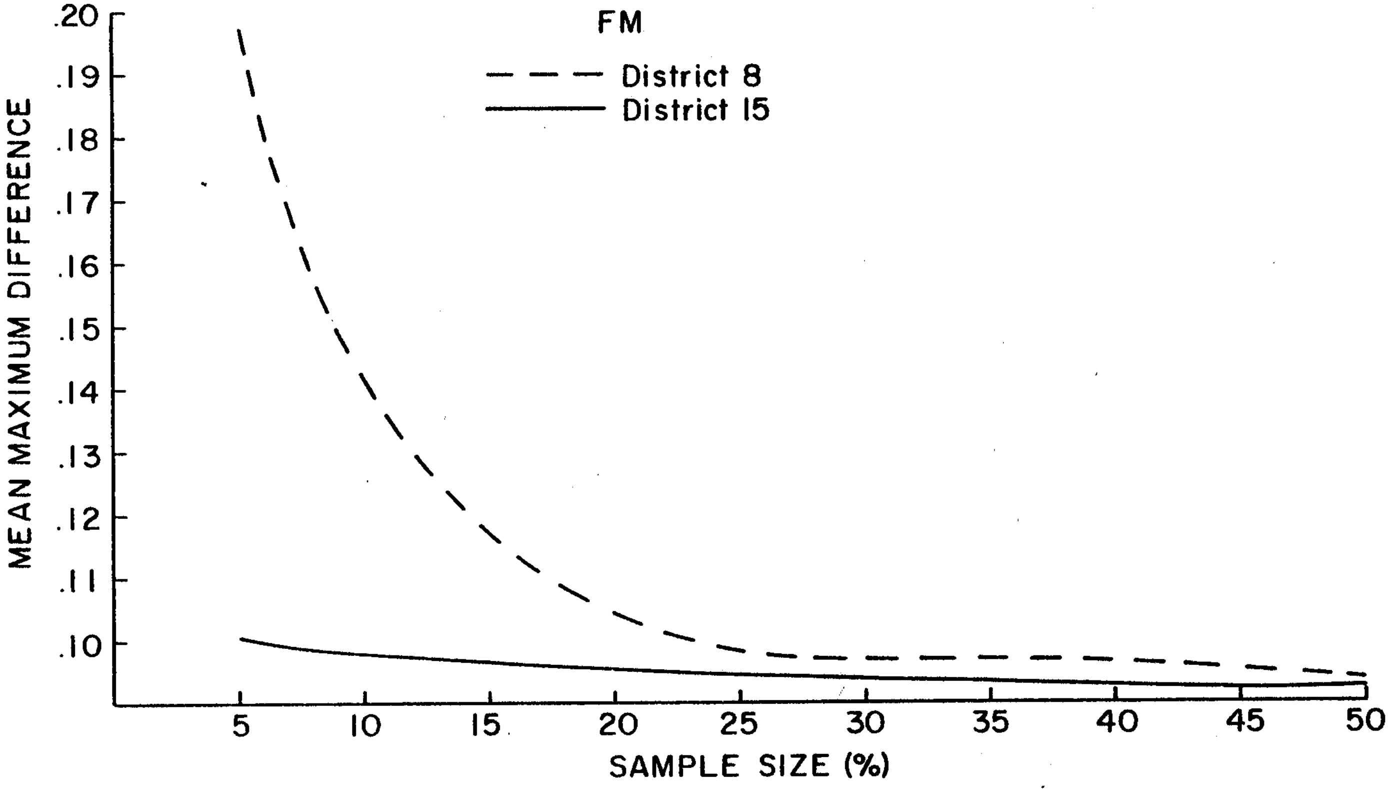


FIGURE 16. Comparison of district results of the mean maximum difference using county stratification method A, System FM

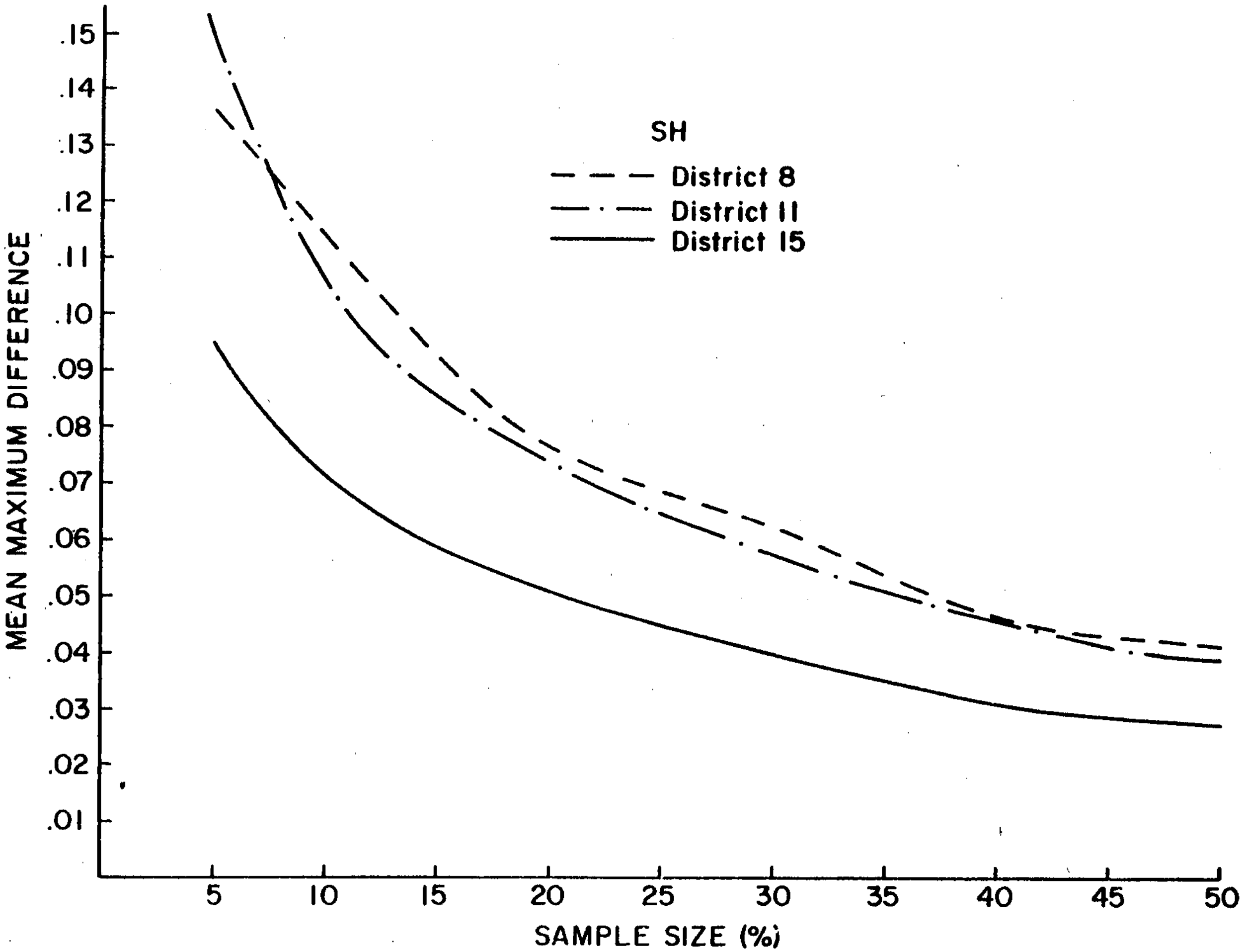


FIGURE 17. Comparison of district results of the mean maximum difference using county stratification method A, System SH

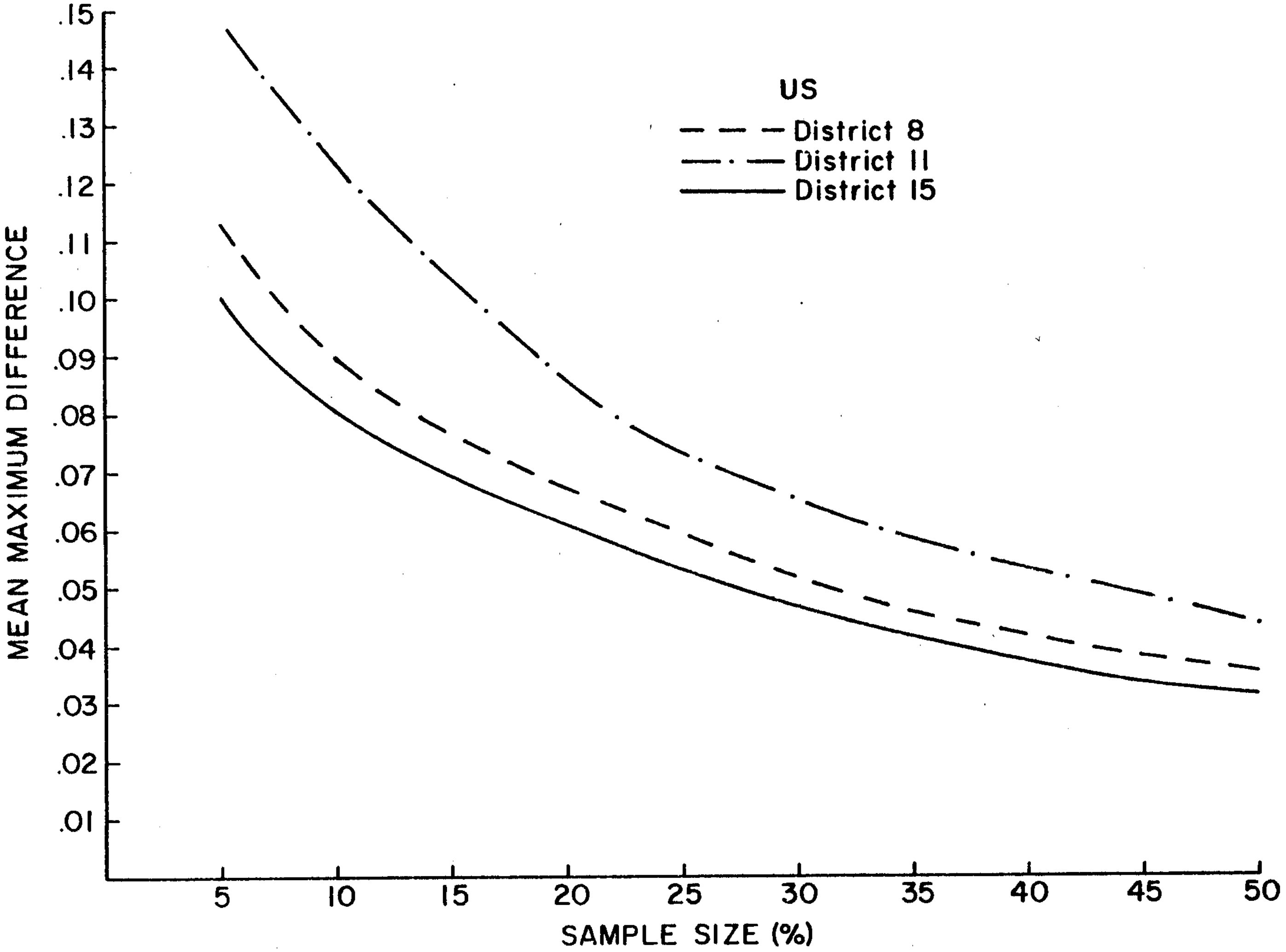


FIGURE 18. Comparison of district results of the mean maximum difference using county stratification method A, System US

shown in Table 12 suggest that the more variable the scores are in a data group, the greater the mean maximum difference.

TABLE 12. Variances of Pavement Scores

System District	IH	FM	SH	US
8	1392.2	639.10	975.4	795.1
11	--	816.34	888.8	1342.9
15	429.1	386.10	634.4	580.5

Analytical Approach to Determining Most Accurate Sampling Plan for Estimating Costs

Cochran (3) states that given a desired degree of precision, an equation exists that connects the sample size with desired precision. This degree of precision is commonly measured by the relative error, r, in the estimated population total or mean.

A first approximation of the sample size is given by:

$$n_0 = \left(\frac{tS}{r\bar{y}}\right)^2 = \frac{1}{C}\left(\frac{S}{\bar{y}}\right)^2 \tag{25}$$

where

n_0 = initial sample size approximation,

t = value of the t distribution associated with some desired level of confidence,

r = relative error desired,

S = standard deviation of the cost per square yard,

$\bar{y}$ = mean cost per square yard, and

c = square of coefficient of variation, fixed as r^2/t^2.

The sample size is then computed as

$$n = \frac{n_0}{1+(n_0/N)} \qquad (26)$$

Table 13 sets out the computed sample sizes with coefficients of variation (CV) at .05 and .10 for District 8.

TABLE 13. Computed Sample Sizes for District 8

	CV = .05	CV = .10
IH (N=154)	94	43
FM (N=938)	194	57
SH (N=270)	72	19
US (N=324)	122	43

In this section three different types of sampling designs are studied in order to find the one which yields the smallest variance of the estimator. Only District 8 data will be analyzed. The designs under investigation are:

(1) Simple random sampling with district stratification

(2) Simple random sampling with county stratification

(3) Two stage random sampling with district stratification.

Design types 1 and 2 were described at the beginning of this chapter. In two stage random sampling the counties would serve as the primary units while the "two-mile" length observations would serve as secondary units.

The equations used to compute the variance of the estimator using the various sample designs follows:

(1) Simple random sampling with district stratification

$$V_{ran}(\bar{y}) = \frac{S^2}{n}(1-f) \tag{27}$$

where

S^2 = variance of cost per square yard in District 8 assuming simple random sampling,

n = sample size (obtained from Table 13),

N = total number of observations in district, and

f = sampling fraction (n/N).

(2) Simple random sampling with county stratification

$$V_{strat}(\bar{y}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{S_h^2}{n_h}\left(1-\frac{n_h}{n}\right) \tag{28}$$

where

S_h^2 = variance of cost within county h,

N = total number of observations in district,

N_h = total number of observations in county h,

n = total number of observations sampled in district,

n_h = total number of observations sampled in county h, and

L = number of counties in district.

(3) Two stage random sampling with district stratification

$$V_{\text{two-stage}}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_b^2 + \frac{1}{nN} \mathop{}_{i=1}^{n} \left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_i^2 \qquad (29)$$

where

N = total number of primary units in district,

M_i = total number of secondary units in i-th primary unit,

n = number of primary units in the sample,

m_i = number of secondary units in the sample from i-th primary unit,

S_b^2 = variance between primary units, and

S_i^2 = variance within secondary units in i-th primary unit.

Tables 14 and 15 show the calculated variances of the estimators under the various sample designs. As can be seen, simple random sampling with county stratification yields the smallest variance in every case. This result reinforces the results obtained with simulation techniques, i.e. simple random sampling with county stratification is the best sample survey design investigated.

TABLE 14. Variances for Three Different Sample Designs for District 8, Assuming a Sample Size Corresponding to a Coefficient of Variation of 0.05

	Simple Random Sampling With District Stratification	Simple Random Sampling With County Stratification	Two Stage Random Sampling with District Stratification Number of Primary Units 8	 6
FM	0.00109	0.00095	0.00133	.00185
IH	0.02016	0.01607	0.08132	0.14285
SH	0.00863	0.00851	0.00903	0.01157
US	0.00485	0.00383	0.00612	0.00754

TABLE 15. Variances for Three Different Sample Designs for District 8, Assuming a Sample Size Corresponding to a Coefficient of Variation of 0.10

	Simple Random Sampling With District Stratification	Simple Random Sampling With County Stratification	Two Stage Random Sampling with District Stratification Number of Primary Units 8	 6
FM	0.00439	0.00326	0.00456	0.00457
IH	0.08095	0.04571	0.08887	0.13725
SH	0.03536	0.03392	0.03981	0.03766
US	0.01990	0.01950	0.02056	0.02131

CHAPTER VII

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The increasing shift from new construction of highways to maintenance or rehabilitation of existing systems requires the SDHPT to know the condition of the existing pavement and the costs of maintaining or rehabilitating that pavement. Because of the limitations on funds, time, equipment and manpower, this information must be obtained through appropriate sampling procedures.

Summary

This report develops a methodology for predicting the percentage of roads with a Pavement Score of 40 or below with a five percent sample and also establishes a sampling scheme (sample survey design and sample size) to assess pavement condition and estimate maintenance and rehabilitation costs.

In order to achieve these objectives, data from three districts in Texas whose roads had been sampled 100 percent were analyzed. An approximate two-mile single lane highway strip served as a sample observation. For each of these observations a Pavement Score and cost of maintenance and rehabilitation were calculated from the raw data collected at the observation sight. These data were divided into four populations by roadway system (IH, FM, SH and US). For purposes of analysis, data from each district and each system were combined to form eleven data groups. (District 11 has no Interstate Highways.)

The methodology used to accomplish the objectives of this research can be divided into three steps:

(a) Fit a density function to the Pavement Score data and determine the more accurate procedure for estimating the percentage of roads below a score of 40 with a five percent sample,

(b) Determine the more accurate estimator of mean cost per square yard for maintenance and rehabilitation, and

(c) Determine sample size and sampling plan needed to accurately estimate pavement condition and maintenance and rehabilitation costs.

Conclusion

Through the use of simulation procedures described in this report the following basic results were concluded:

(a) It was found that the beta distribution fits the Pavement Score data in each of the eleven data groups.

(b) The percentage of roads with a Pavement Score of 40 or less can best be estimated through the use of the cumulative beta distribution with estimation of its parameters by the method of moments. This procedure leads to a better estimate of the percentage below 40 than by a direct estimate from the sample.

(c) The mean of ratios was found to be the more accurate estimator of the cost per square yard of maintenance and rehabilitation costs when the minimum mean squared error served as the selection criterion.

(d) Pavement condition can be estimated accurately with a smaller sample size than is required for costs of maintenance and rehabilitation.

(e) Simple random sampling with county stratification usually leads to smaller relative errors in predicting costs than simple random sampling with district stratification.

(f) Methods A and B of the stratified-over-counties plan lead to similar results both in predicting pavement condition and estimating costs.

(g) The structure of this population does not lend itself to systematic sampling.

(h) The better the condition of roads within a district, the more difficult it is to accurately estimate a mean cost per square yard.

(i) The more variable the Pavement Scores are within a district, the greater the mean maximum difference between the empirical cumulative distribution and the 100 percent sample cumulative distribution.

Recommendations

A. Recommendations for SDHPT

There exist two methods to combine results of Chapter 6 in order to determine the relative errors associated with various sample sizes for each sample design on a state-wide basis. One method is to take a weighted average of the mean errors at each sample size over all three districts. This weighted average can be defined as follows:

$$\hat{\mu}_{\text{mean error}} = \frac{\sum_{h=1}^{3} N_h \, \hat{\mu}_{\text{error}(h)}}{N} \tag{30}$$

where

$\hat{\mu}_{\text{error}(h)}$ = estimated mean error in district h,

N_h = number of observations in district h, and

N = number of observations in all three districts combined.

A second method would be to use the maximum mean error between the three districts at each sample size. The results of the two methods are summarized in Tables 16, 17, 18, 19, 20, and 21 for each sample design.

TABLE 16. Weighted averages over the three 100 percent Sampled Districts of the Mean Error with District Stratification

System Sample Size (%)	IH	FM	SH	US
5	0.8719	0.4198	0.5028	0.5376
10	0.6257	0.3114	0.3643	0.4134
15	0.4922	0.2502	0.3188	0.3431
20	0.3895	0.2075	0.2577	0.2904
25	0.3439	0.1896	0.2446	0.2561
30	0.3284	0.1767	0.2304	0.2259
35	0.2885	0.1563	0.2180	0.2029
40	0.2639	0.1513	0.2074	0.1815
45	0.2303	0.1442	0.1934	0.1588
50	0.2197	0.1401	0.1872	0.1535

TABLE 17. Maximum mean error between the three 100 percent Sampled Districts with District Stratification

System Sample Size (%)	IH	FM	SH	US
5	1.0185	0.5214	0.5360	0.7432
10	0.7337	0.4129	0.4057	0.5925
15	0.5854	0.3233	0.3640	0.5095
20	0.4495	0.2795	0.3294	0.4313
25	0.4048	0.2510	0.3047	0.3860
30	0.3932	0.2510	0.2998	0.3047
35	0.3338	0.2248	0.2922	0.2852
40	0.3146	0.2237	0.2849	0.2467
45	0.2703	0.2196	0.2718	0.2016
50	0.2594	0.2214	0.2715	0.1896

TABLE 18. Weighted averages over the three 100 percent Sampled Districts of the Mean Error with County Stratification (method A)

System Sample Size (%)	IH	FM	SH	US
5	0.5957	0.3879	0.4586	0.3776
10	0.4705	0.2857	0.3530	0.1997
15	0.3897	0.2346	0.3009	0.2753
20	0.3286	0.2071	0.2688	0.2247
25	0.2931	0.1912	0.2481	0.2057
30	0.2821	0.1782	0.2401	0.1832
35	0.2457	0.1651	0.2229	0.1582
40	0.2396	0.1527	0.2175	0.1462
45	0.2179	0.1411	0.2100	0.1368
50	0.1946	0.1361	0.2073	0.1243

TABLE 19. Maximum mean error between the three 100 percent Sampled Districts with County Stratification (method A)

System Sample Size (%)	IH	FM	SH	US
5	0.7010	0.4745	0.5446	0.4893
10	0.5662	0.3759	0.4375	0.3681
15	0.4743	0.3018	0.3911	0.3564
20	0.4007	0.2744	0.3574	0.2853
25	0.3549	0.2722	0.3306	0.2507
30	0.3499	0.2584	0.3277	0.2276
35	0.3016	0.2451	0.3247	0.1964
40	0.2974	0.2284	0.3183	0.1705
45	0.2739	0.2214	0.3181	0.1601
50	0.2402	0.2186	0.3180	0.1438

TABLE 20. Weighted averages over the three 100 percent Sampled Districts of the Mean Error with County Stratification (method B)

System Sample Size (%)	IH	FM	SH	US
5	0.6976	0.3913	0.4407	0.4269
10	0.4835	0.2861	0.3399	0.3098
15	0.3964	0.2341	0.2887	0.2774
20	0.3291	0.2084	0.2561	0.2247
25	0.2886	0.1921	0.2470	0.2047
30	0.2790	0.1743	0.2339	0.1825
35	0.2417	0.1612	0.2181	0.1599
40	0.2364	0.1533	0.2114	0.1430
45	0.2131	0.1415	0.2058	0.1369
50	0.1933	0.1371	0.2000	0.1234

TABLE 21. Maximum mean error between the three 100 percent Sampled Districts with County Stratification (method B)

System Sample Size (%)	IH	FM	SH	US
5	0.8521	0.4732	0.5202	0.5476
10	0.5809	0.3725	0.4026	0.3958
15	0.4845	0.3002	0.3675	0.3756
20	0.3999	0.2764	0.3295	0.2925
25	0.3510	0.2740	0.3254	0.2573
30	0.3444	0.2485	0.3129	0.2343
35	0.2957	0.2397	0.3111	0.2028
40	0.2919	0.2294	0.3080	0.1736
45	0.2667	0.2220	0.3060	0.1566
50	0.2384	0.2209	0.3002	0.1413

It is recommended that simple random sampling with county stratification method A be used in the future. This plan was chosen over method B since it is easier to compute and county estimates are not needed. It is further recommended that a sample size of at least 35 percent be used to estimate costs per square yard. This sample size produces the following weighted mean errors:

System	Weighted Mean Error
IH	25%
FM	17%
SH	22%
US	16%

Table 22 shows the mean maximum difference between the empirical cumulative distribution and the 100 percent sample cumulative distribution of all data groups using simple random sampling with county stratification method A and a 35% sample.

TABLE 22. Mean Maximum Difference (in percent) using County Stratification method A with a 35% sample

System District	IH	FM	SH	US
8	7.0	3.0	5.4	4.6
11	--	3.4	5.1	6.1
15	3.8	2.5	3.6	4.2

B. Recommendations for Further Study

The following are recommendations for further study in this area:

(a) Determine accuracy of the cumulative beta distribution with sample sizes other than five percent,

(b) Determine accuracy of simple random sampling with county stratification using optimal allocation among the counties, and

(c) Develop a relationship between costs and Pavement Scores whereby costs are not estimated directly by sampling but through the sampled Pavement Scores.

REFERENCES

1. Leon M. Noel. Pavement Management as Related to Current Shortages of Materials, Finances and Manpower. Proceedings - Paving Conference (17th), University of New Mexico, 1980.

2. U.S. Department of Transportation. The Status of the Nation's Highways: Conditions and Performance. Report of the Secretary of Transportation to the United States Congress, January, 1981.

3. William G. Cochran. Sampling Techniques. John Wiley & Sons, New York, New York, 1953.

4. R. L. Scheaffer, W. Mendenhall, and L. Ott. Elementary Survey Sampling. Duxbury Press, Belmont, California, 1979.

5. Vic Burnett. Elements of Sampling Theory. The English University Press Ltd., London, England, 1974.

6. H. S. Konijn. Statistical Theory of Sample Survey Design and Analysis. North-Holland Publishing Company, Amsterdam, Holland, 1973.

7. J. P. Mahoney and R. L. Lytton. Measurements of Pavement Performance Using Statistical Sampling Techniques. Research Report 207-2, Texas Transportation Institute, Texas A&M University, 1978.

8. J. A. Epps, C. W. Shaw, G. G. Harvey, J. P. Mahoney, and W. W. Scott. Operational Characteristics of Mays Ride Meter. Texas Transportation Institute Research Report 151-3, September, 1976.

9. J. A. Epps, A. H. Meyer, I. E. Larrimore, and H. L. Jones. Roadway Maintenance Evaluation User's Manual. Texas Transportation Institute Research Report 151-2, September, 1974.

10. U. S. Department of Transportation, Federal Highway Administration, Program Management Division. Highway Performance Monitoring System, Phase 1 Report. September, 1978.

11. U. S. Department of Transportation, Federal Highway Administration, Office of Highway Planning. Highway Performance Monitoring System, Field Manual. September, 1980.

12. U. S. Department of Transportation, Federal Highway Administration, Office of Highway Planning. Highway Performance Monitoring System, Revised Field Manual. October, 1981.

13. Robert L. Lytton, Michael I. Darter, Mohamed Y. Shahin, and J. Brent Rauhut. State-Level Pavement Monitoring Program. Transportation Research Record 846, Pavement Management, National Academy of Sciences, Washington, D.C.

14. Frederick W. Morgan and James L. Burati, Jr. Statistical Evaluation of Random vs. Stratified Random Sampling for Pavement Test Section. A paper presented at the 62nd Annual Meeting of the Transportation Research Board, Washington, D.C., 1983.

15. W. J. Conover. Practical Nonparametric Statistics. John Wiley & Sons Inc., New York, 1971.

16. Bernard Ostle and Richard W. Mensing. Statistics in Research. The Iowa State University Press, Ames, Iowa, 1979.

17. Robert V. Hogg and Allen T. Craig. Introduction to Mathematical Statistics. MacMillan Publishing Co., Inc., New York, 1978.

18. Wayne W. Daniel. Applied Nonparametric Statistics. Houghton Mifflin Company, Boston, Massachusetts, 1978.

19. Frank O. Massey, Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the America Statistical Association, Vol. 46, 1951, pp. 68-78.

APPENDICES

APPENDIX A

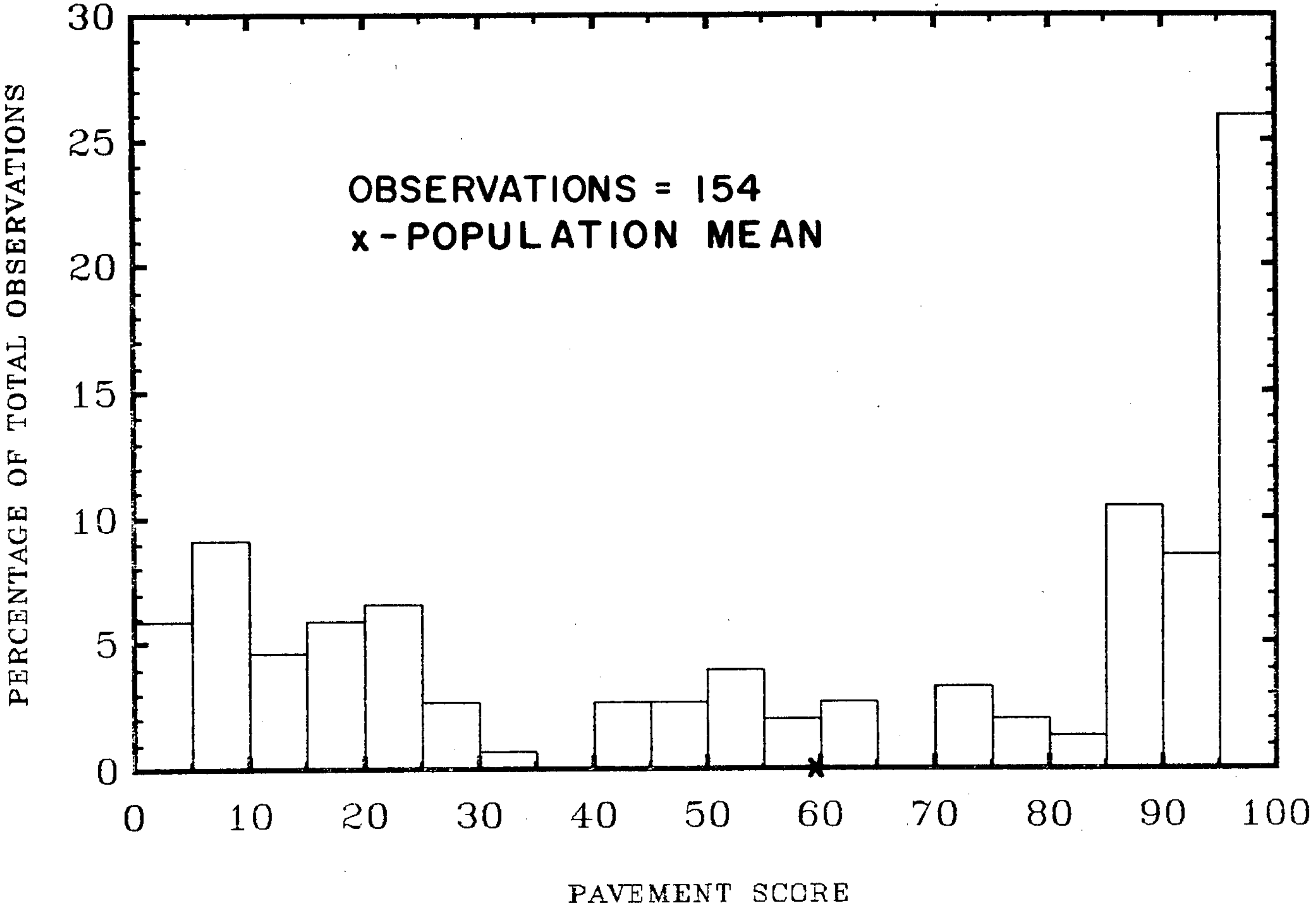


FIGURE A-1. Pavement Score histogram, 100% sample for District 8, System IH

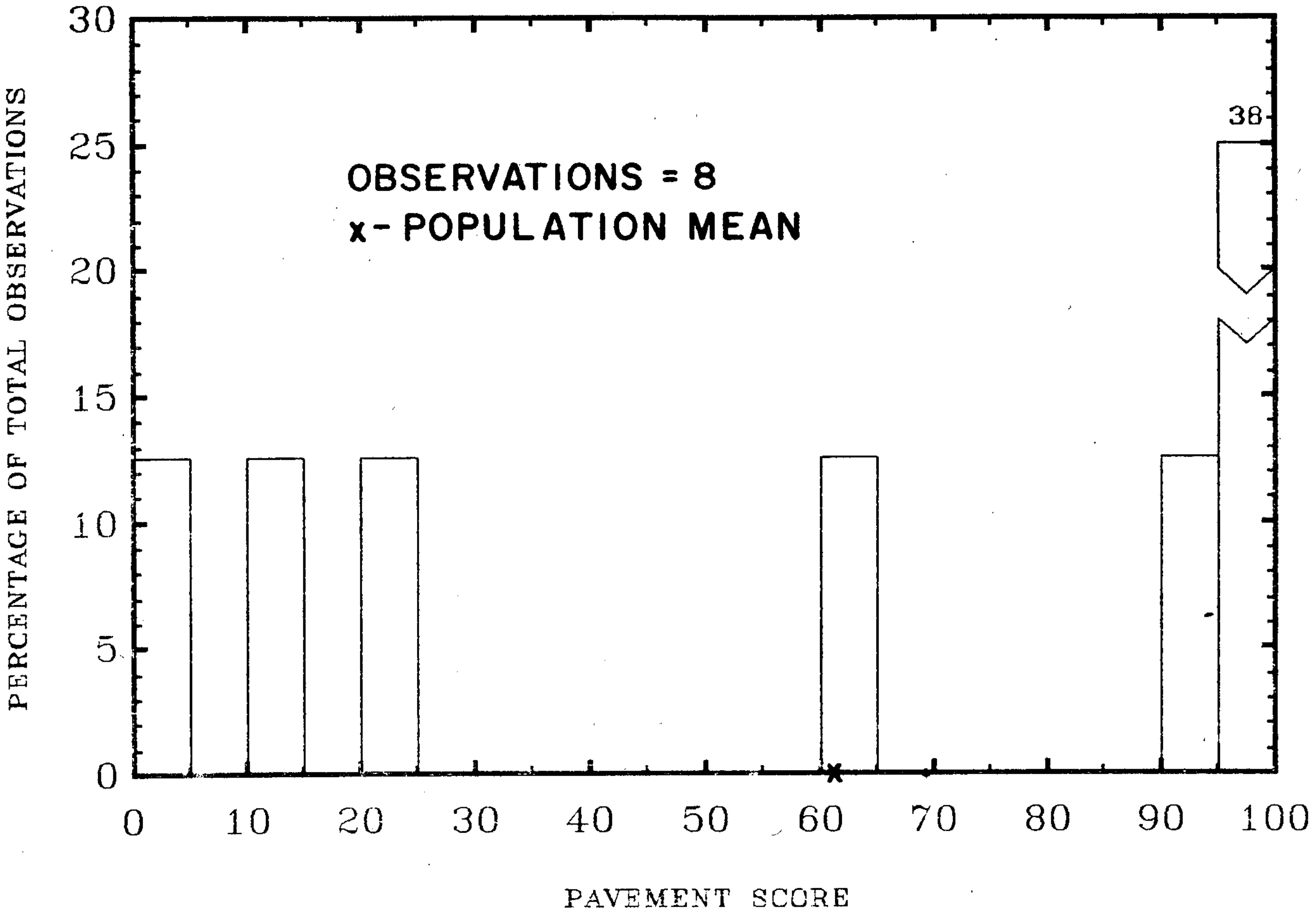


FIGURE A-2. Pavement Score histogram, 5% sample for District 8, System IH

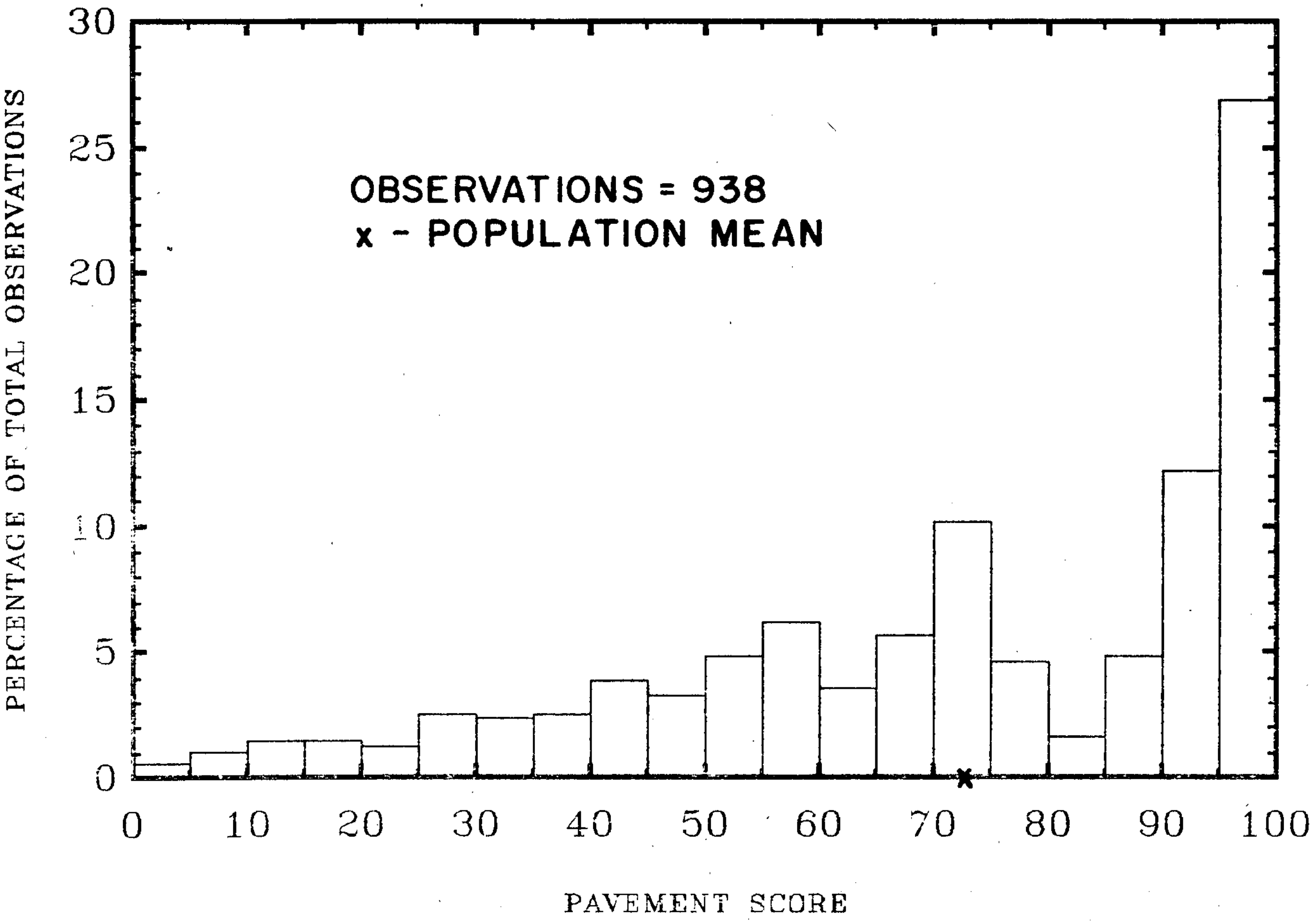


FIGURE A-3. Pavement Score histogram, 100% sample for District 8, System FM

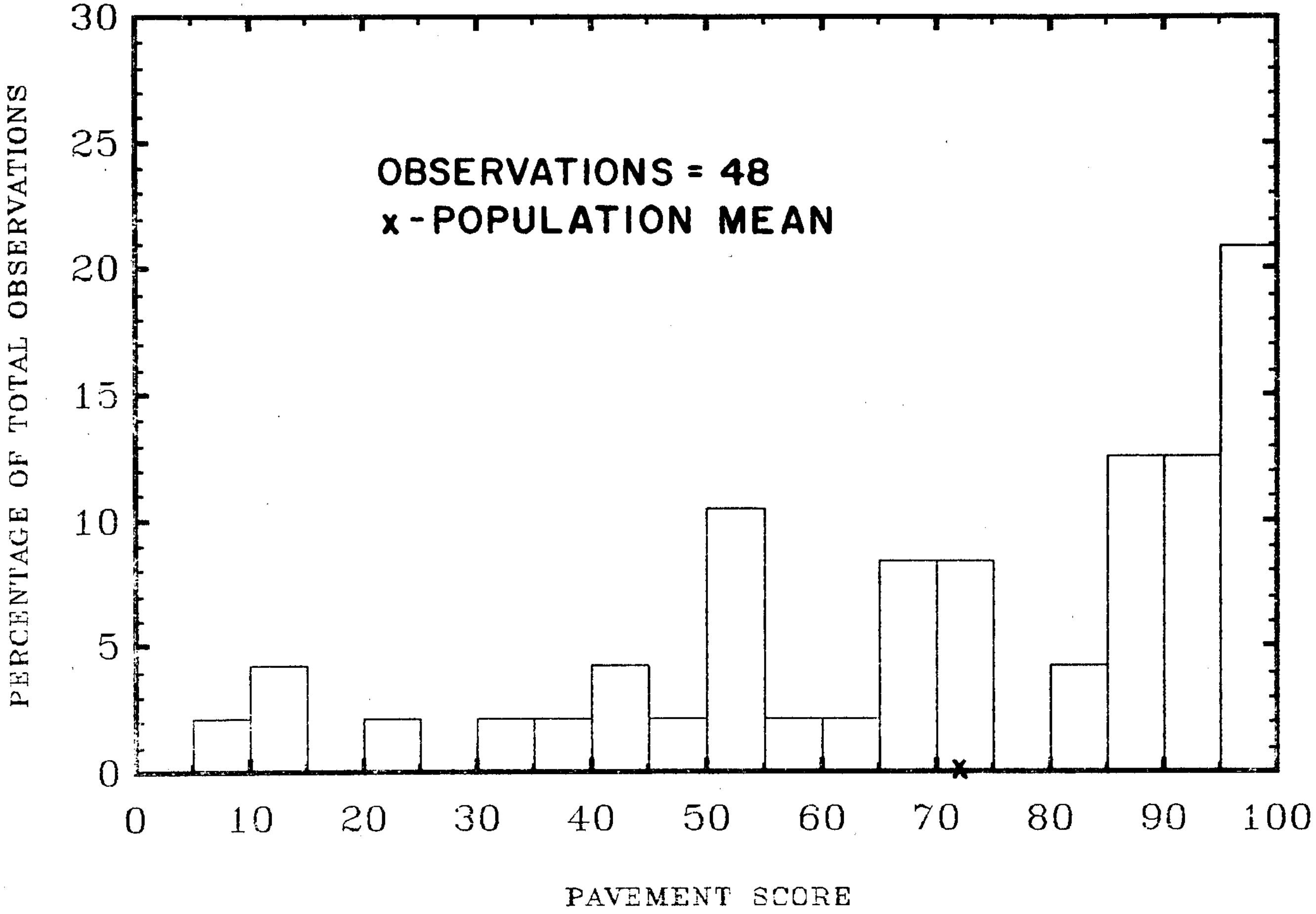


FIGURE A-4. Pavement Score histogram, 5% sample for District 8, System FM

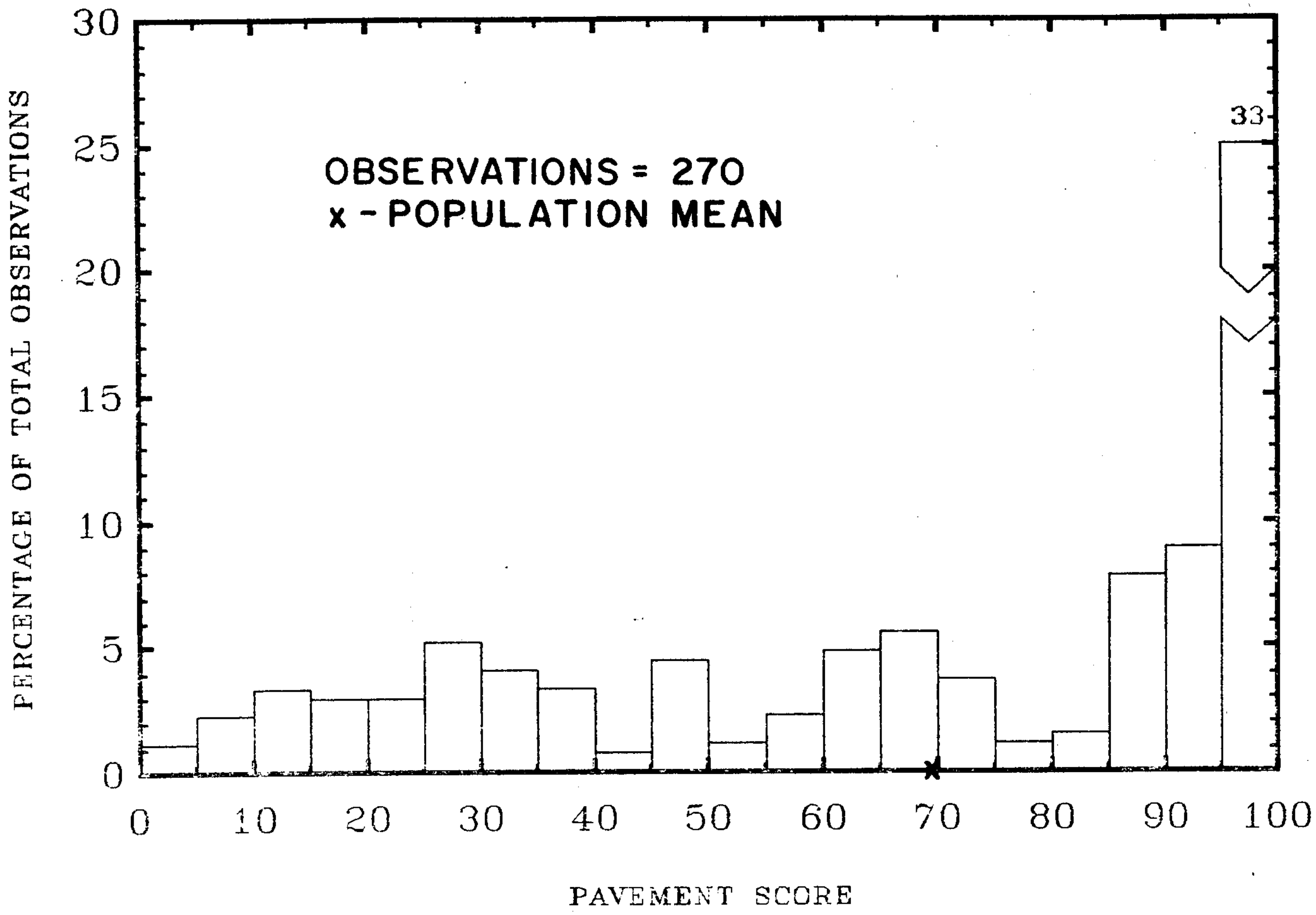


FIGURE A-5. Pavement Score histogram, 100% sample for District 8, System SH

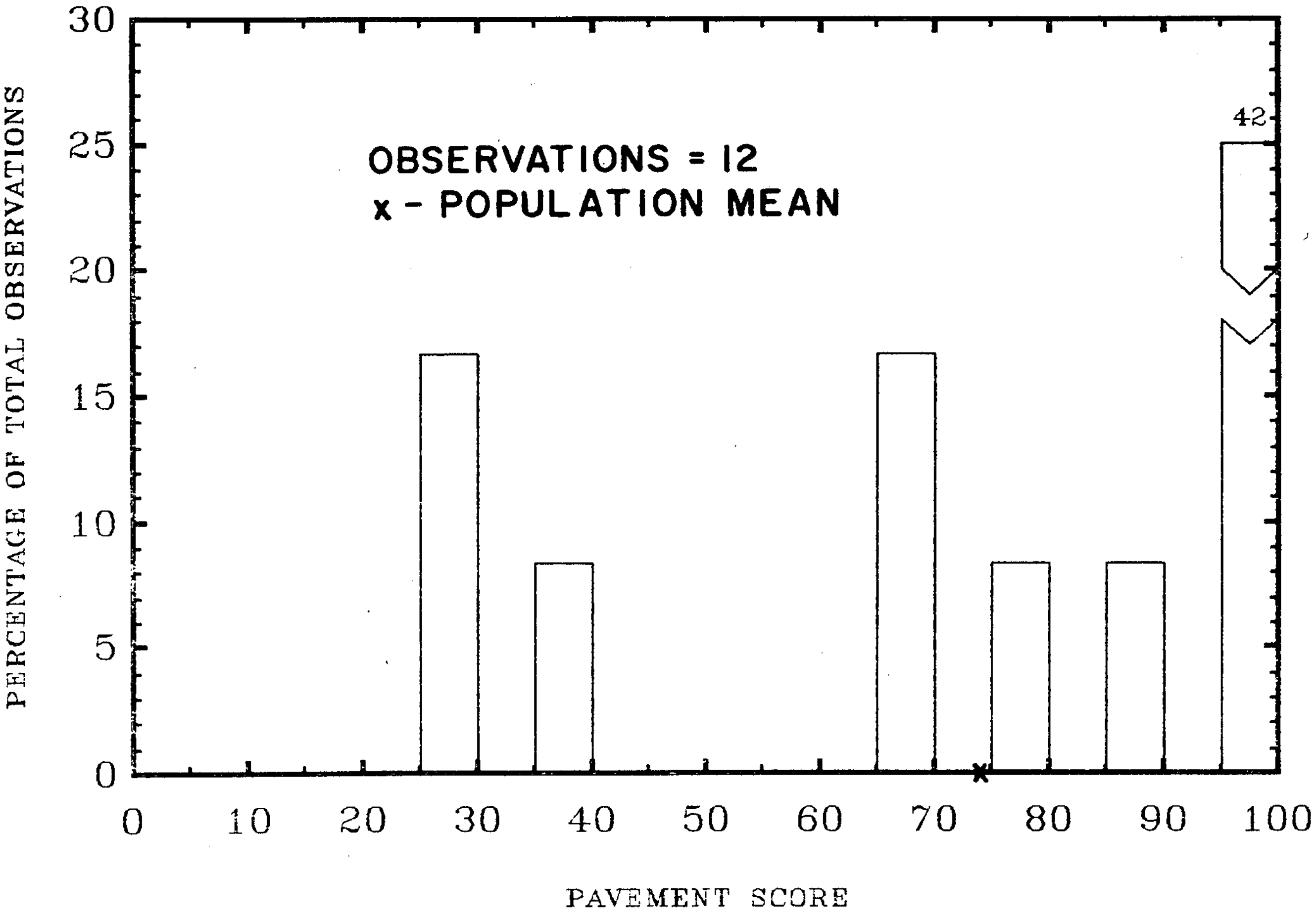


FIGURE A-6. Pavement Score histogram, 5% sample for District 8, System SH

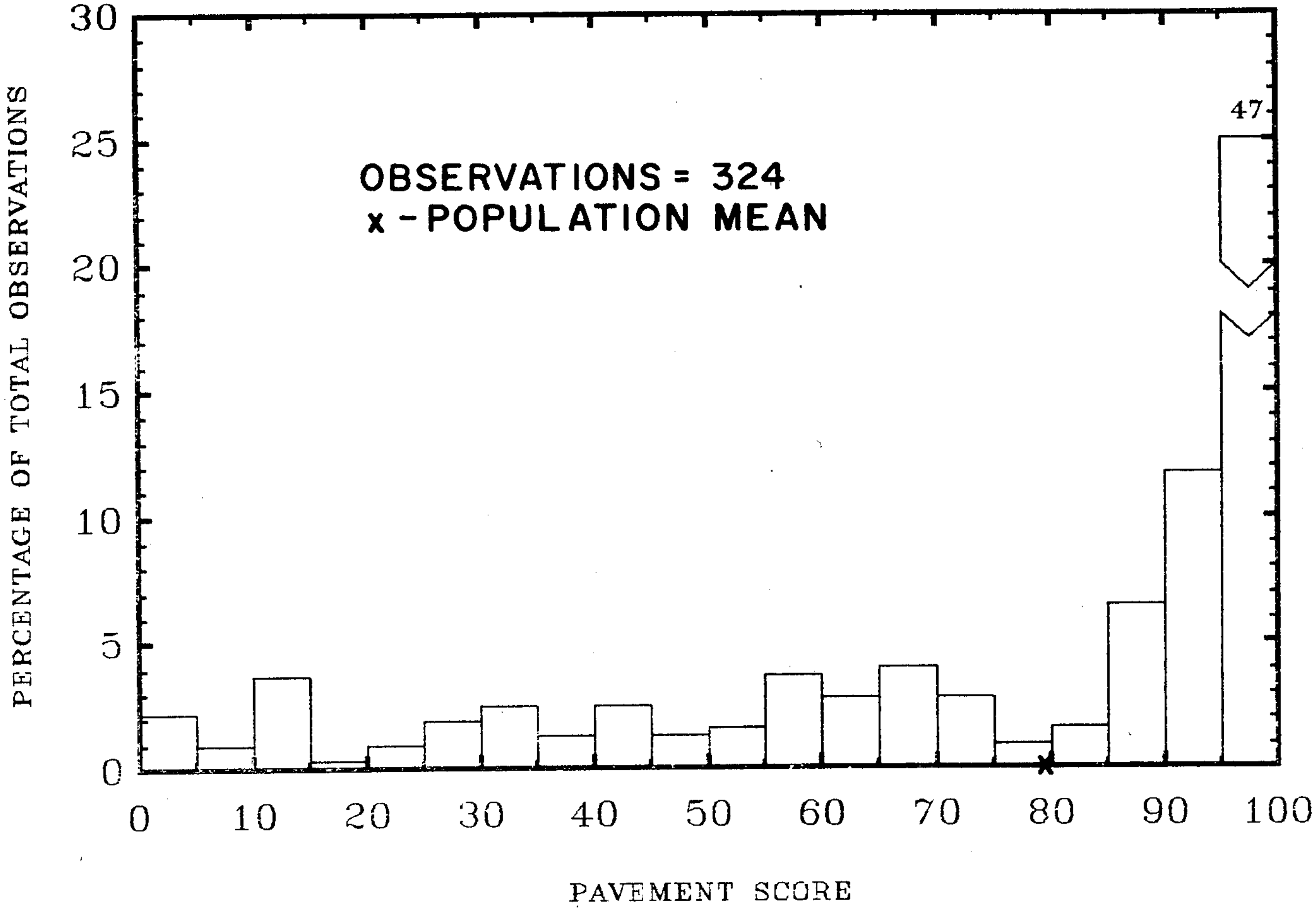


FIGURE A-7. Pavement Score histogram, 100% sample for District 8, System US.

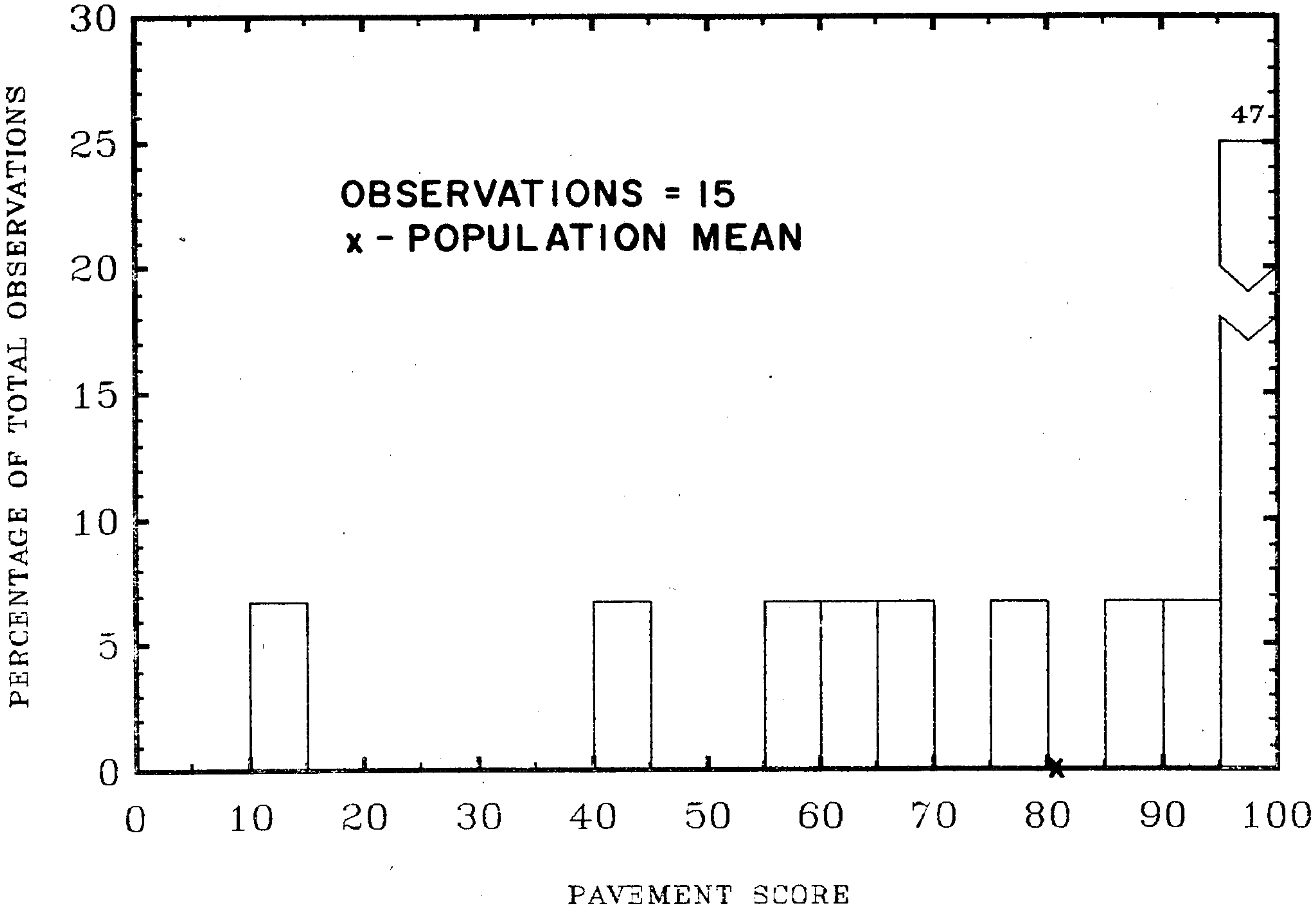


FIGURE A-8. Pavement Score histogram, 5% sample for District 8, System US

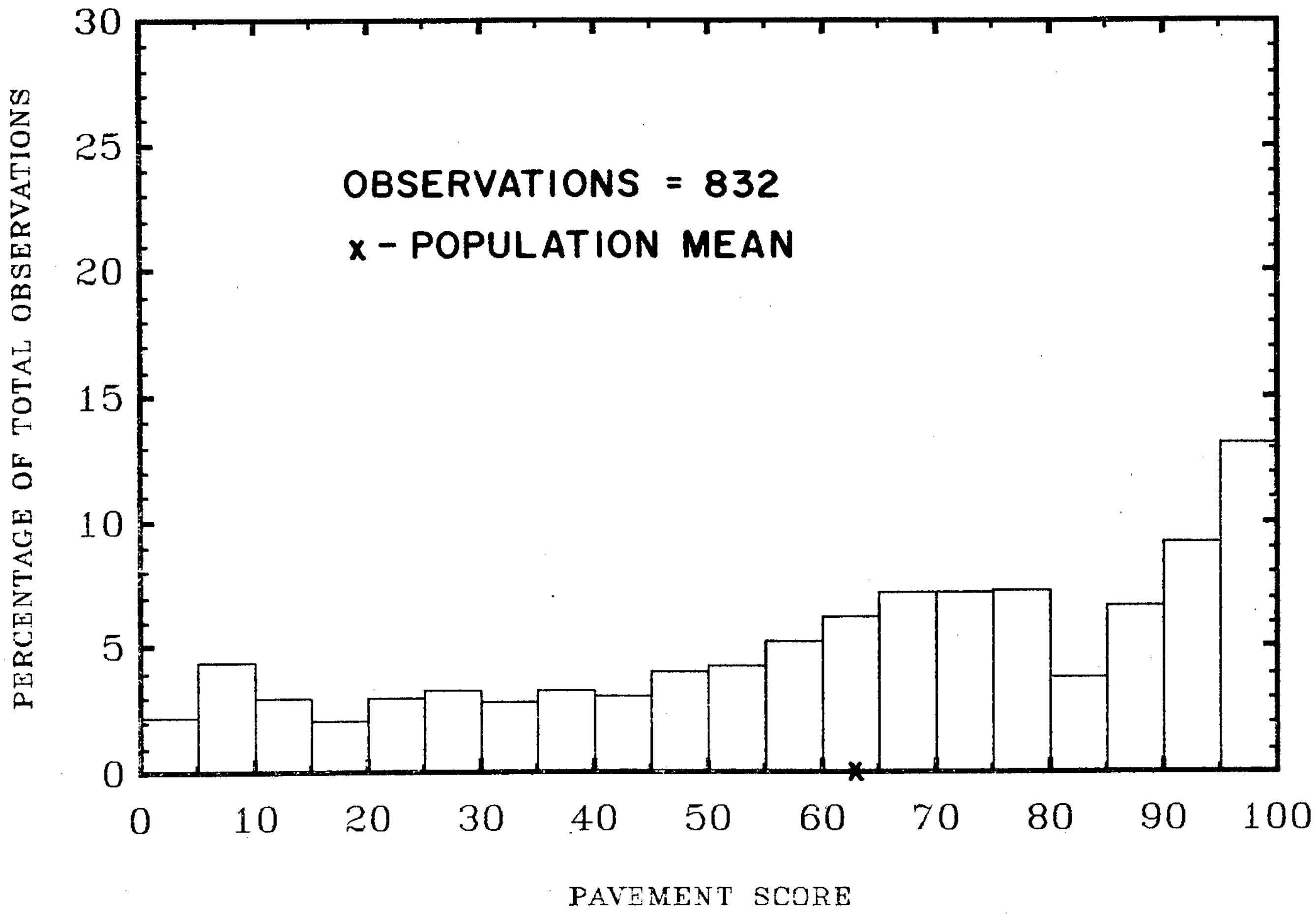


FIGURE A-9. Pavement Score histogram, 100% sample for District 11, System FM

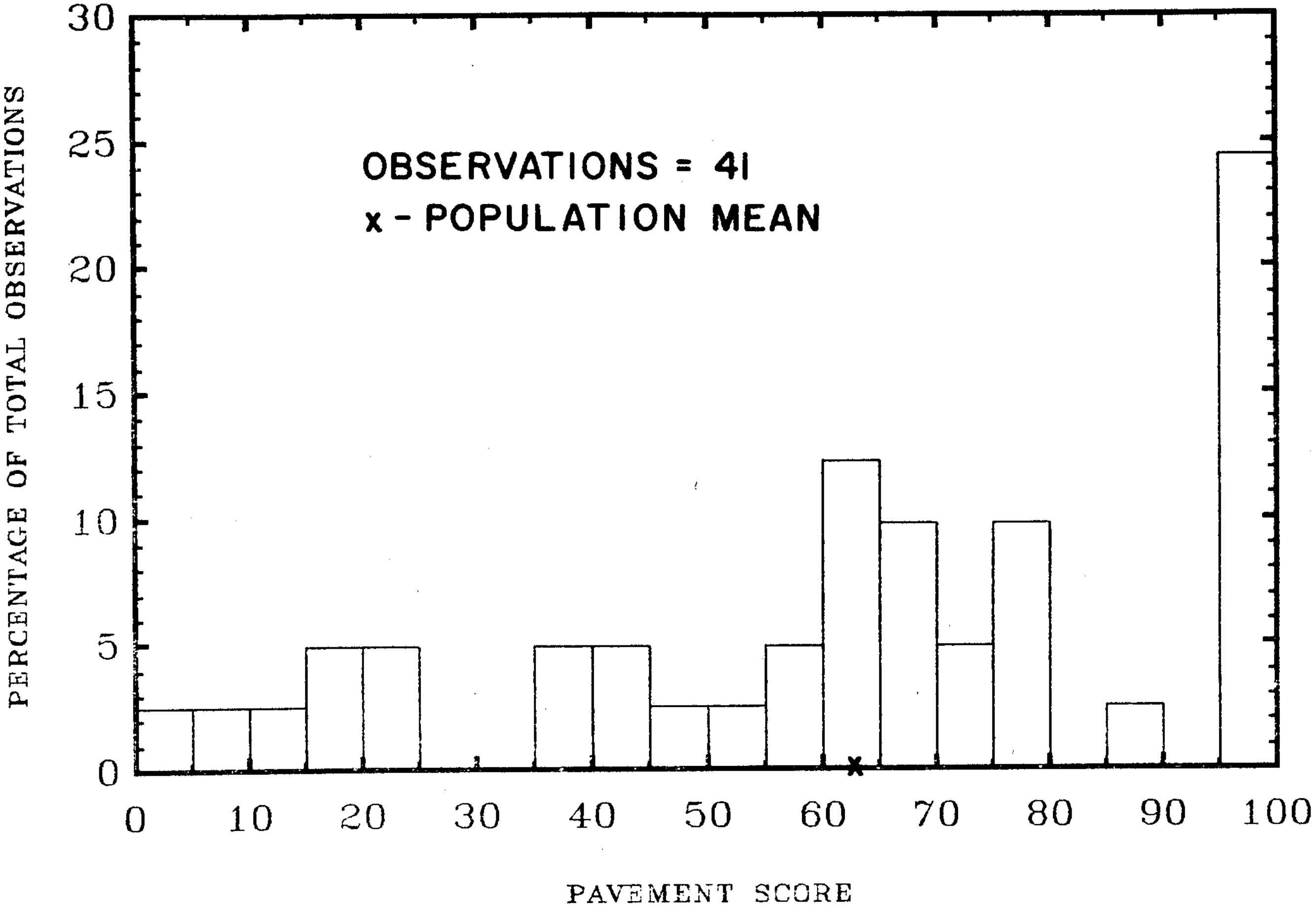


FIGURE A-10. Pavement Score histogram, 5% sample for District 11, System FM

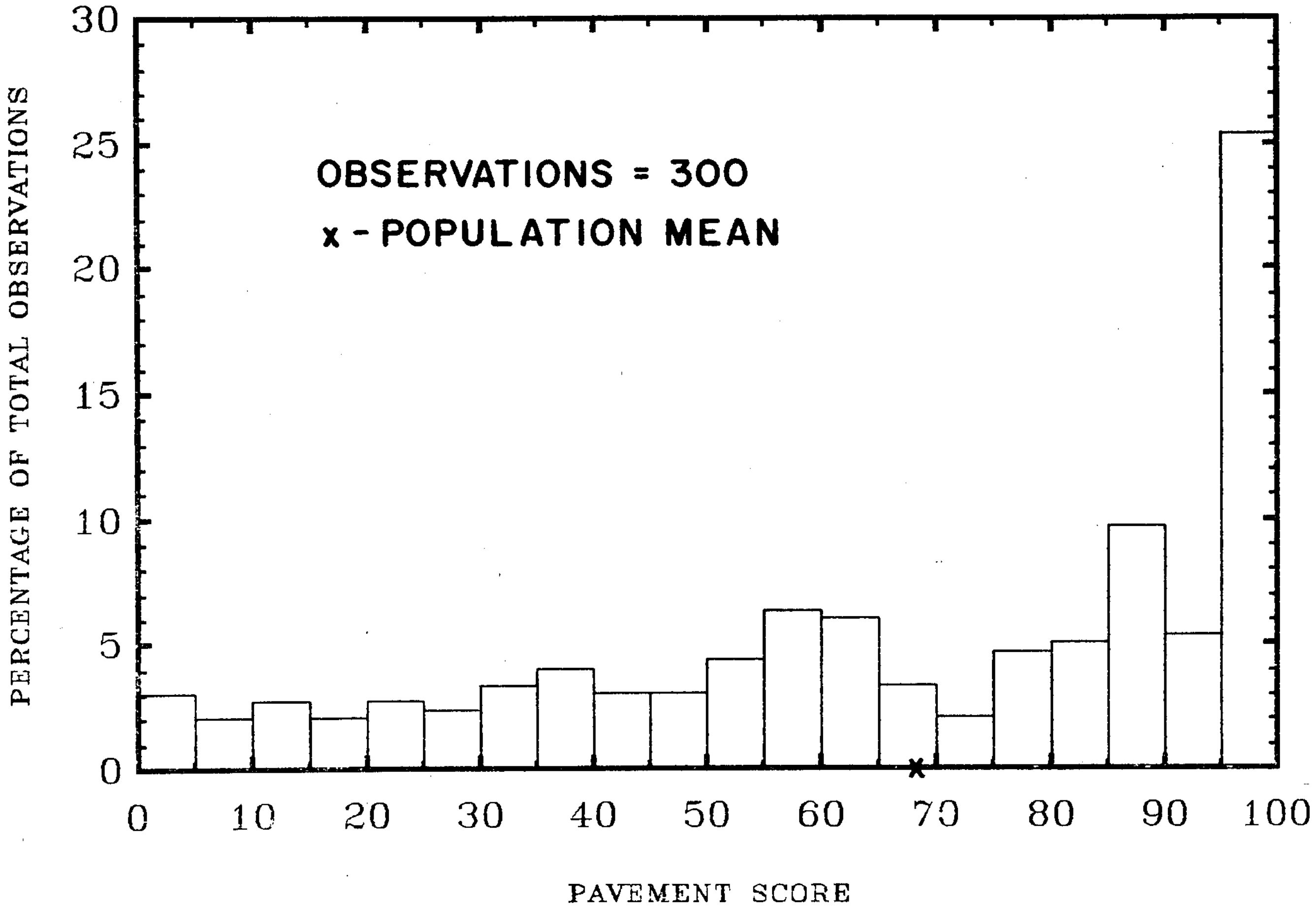


FIGURE A-11. Pavement Score histogram, 100% sample for District 11, System SH

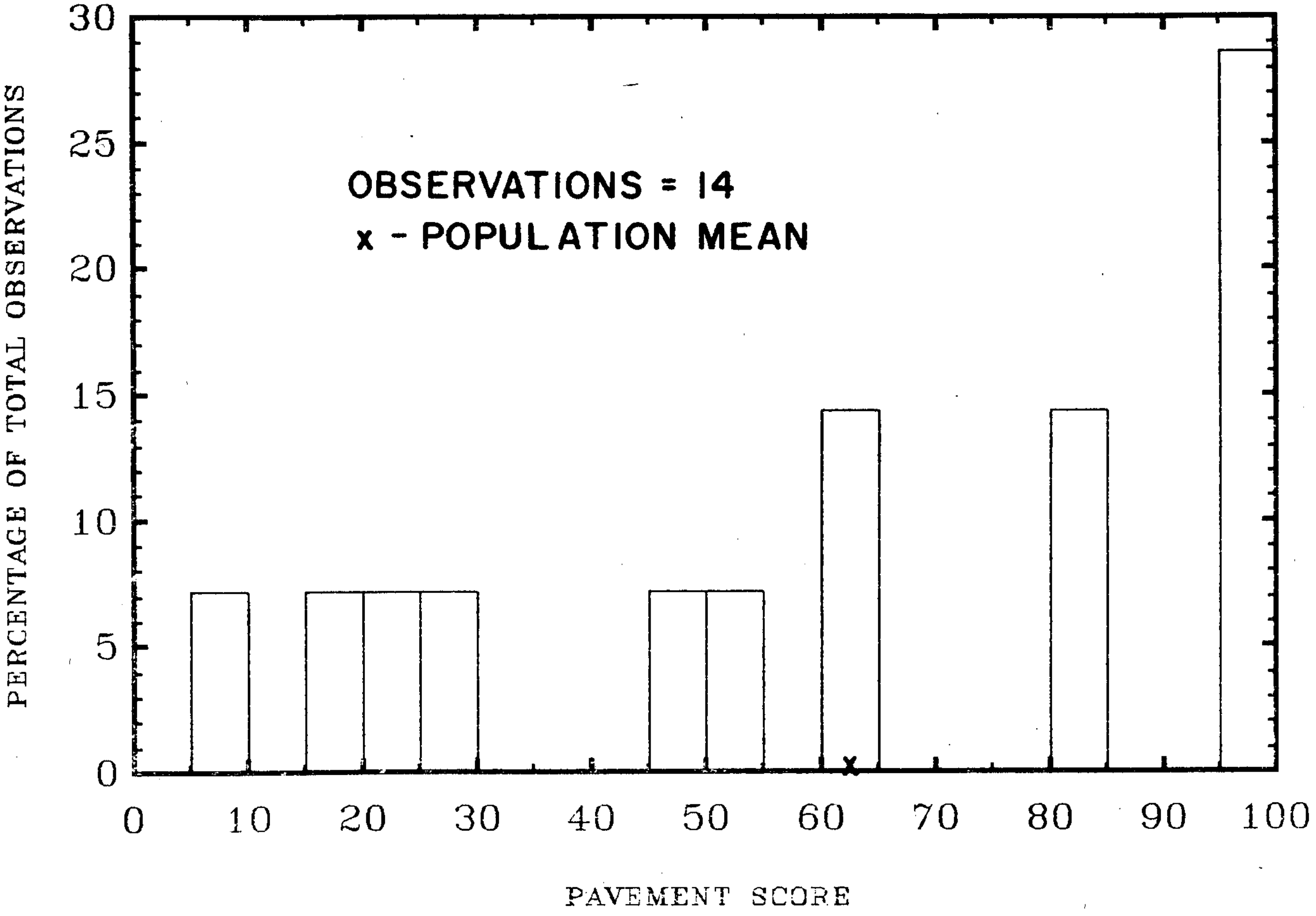


FIGURE A-12. Pavement Score histogram, 5% sample for District 11, System SH

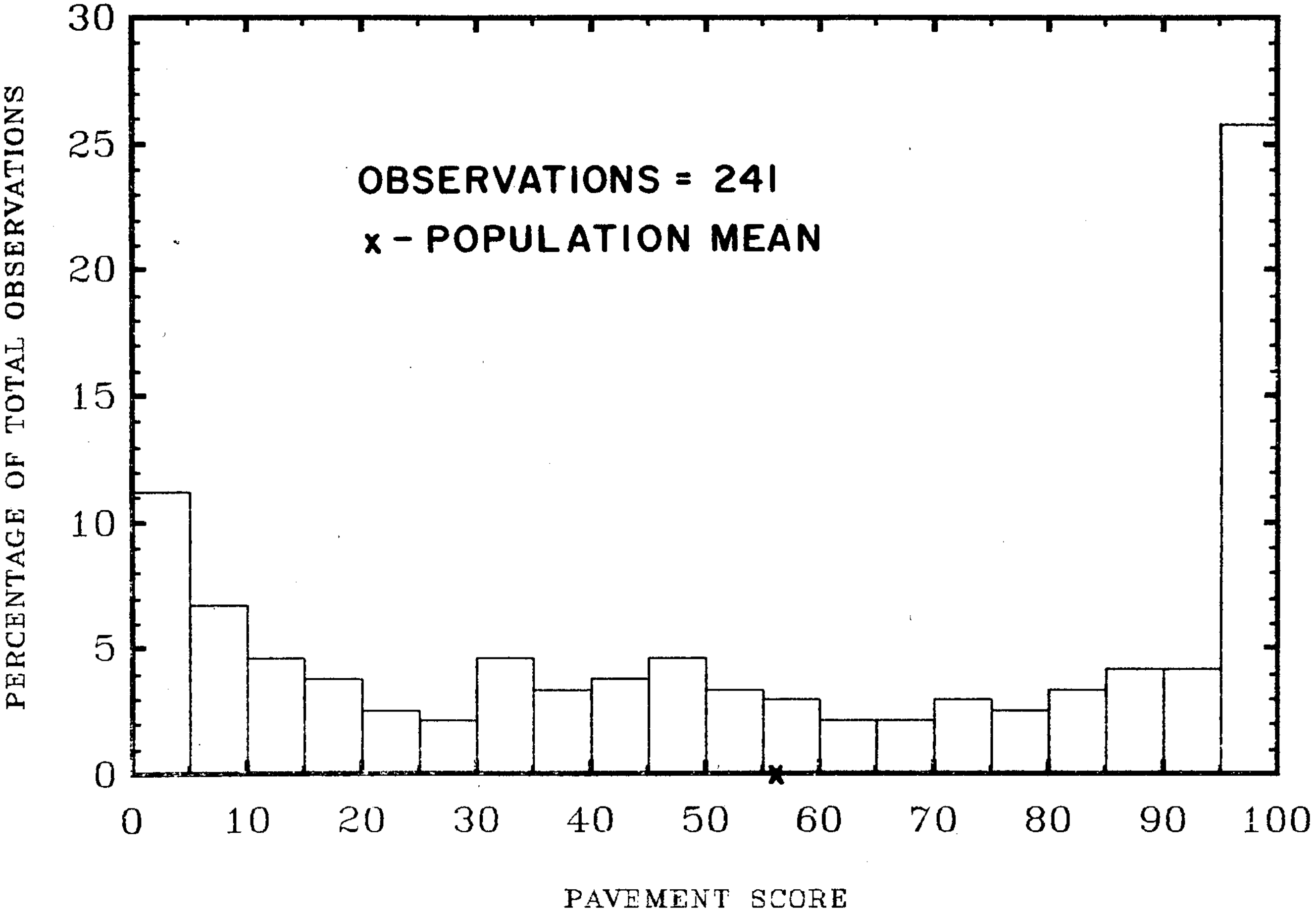


FIGURE A-13. Pavement Score histogram, 100% sample for District 11, System US

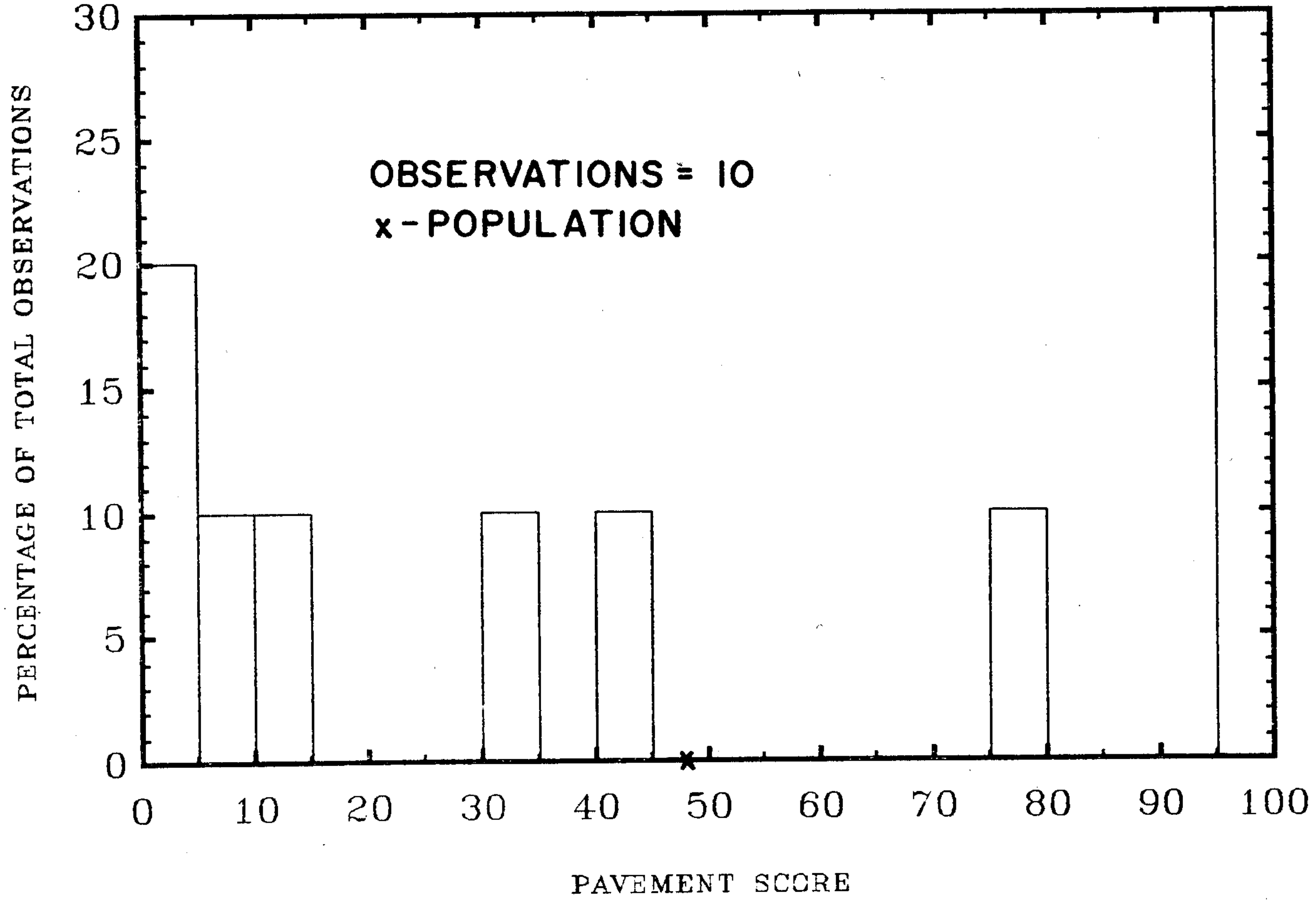


FIGURE A-14. Pavement Score histogram, 5% sample for District 11, System US

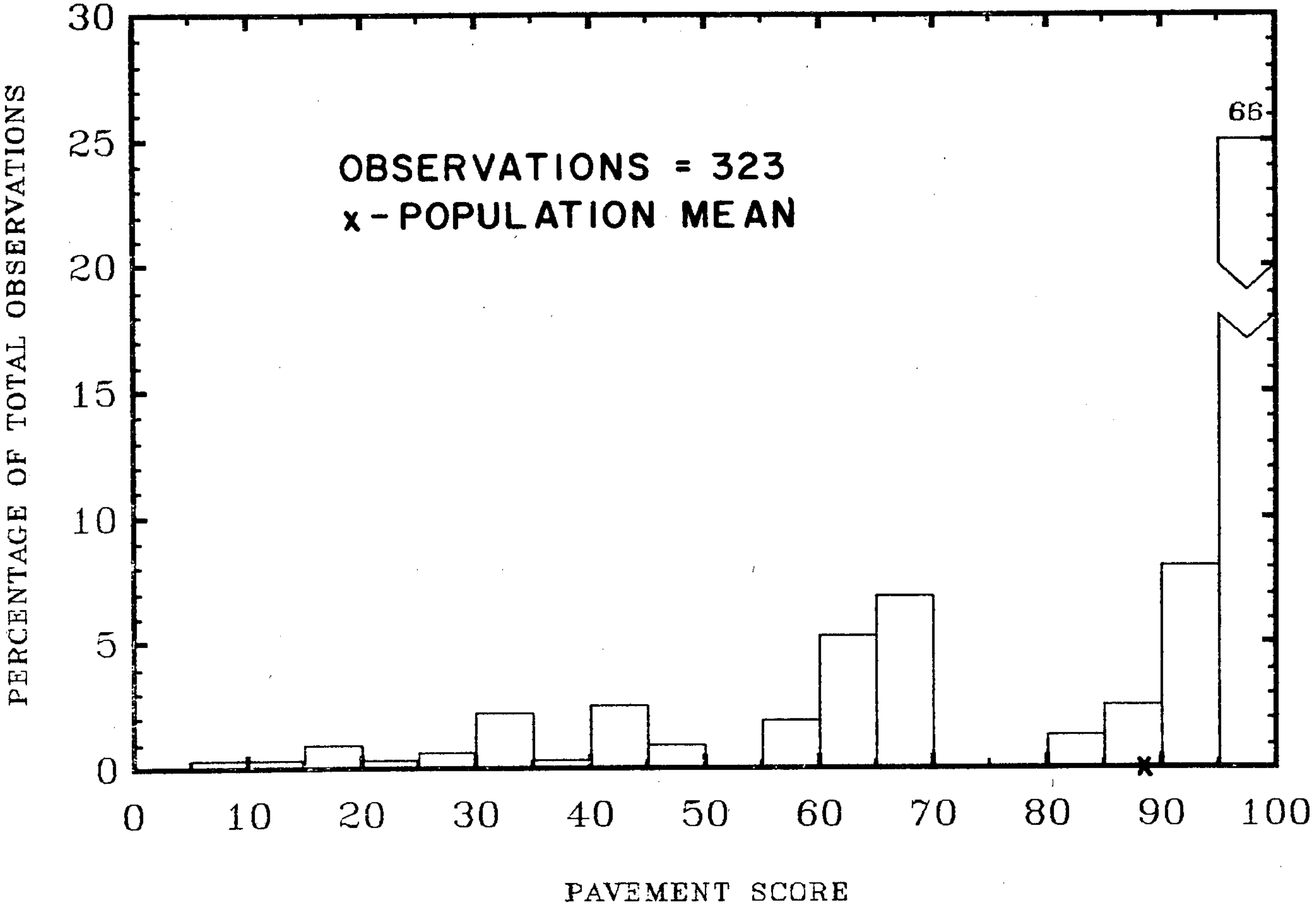


FIGURE A-15. Pavement Score histogram, 100% sample for District 15, System IH

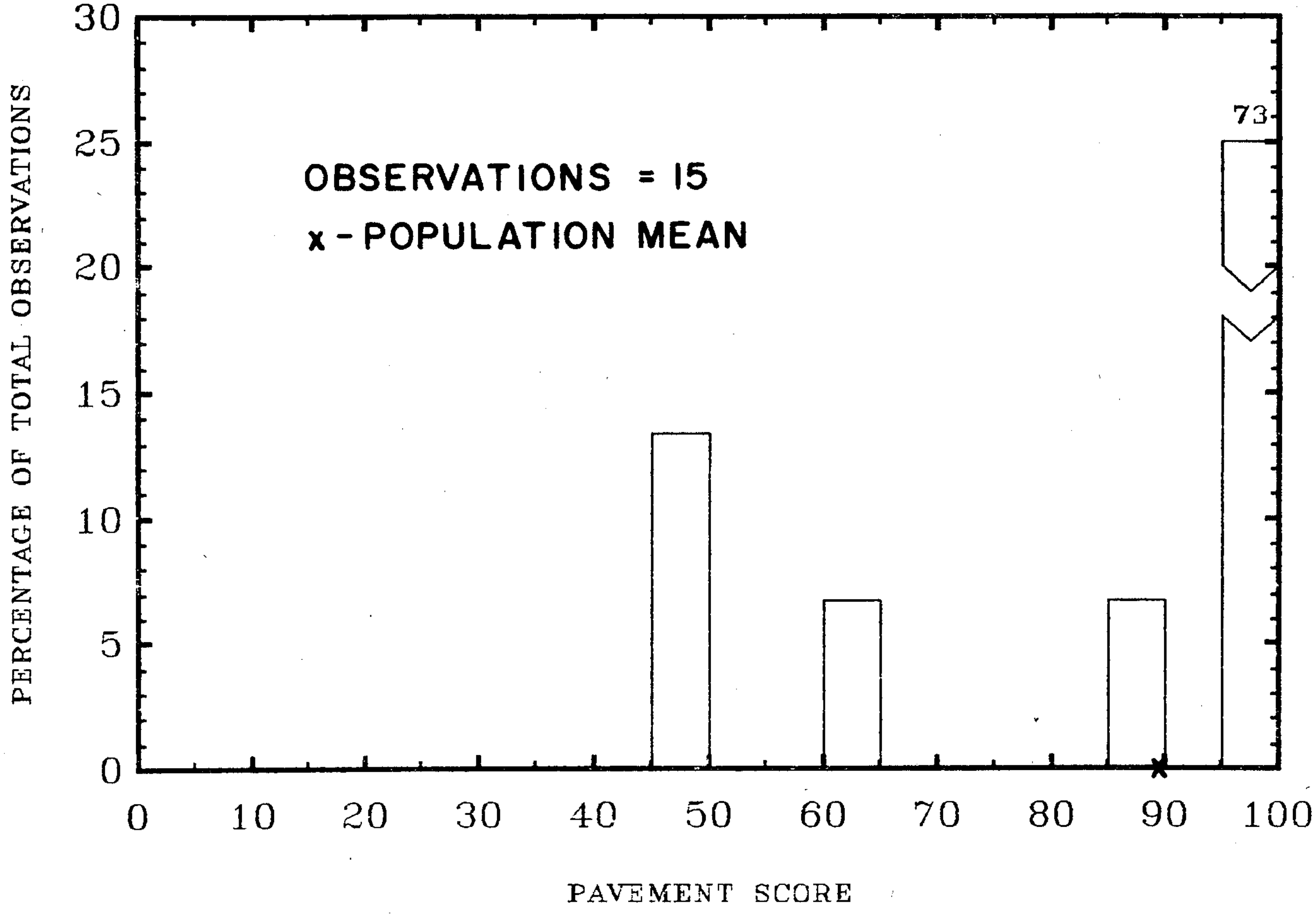


FIGURE A-16. Pavement Score histogram, 5% sample for District 15, System IH

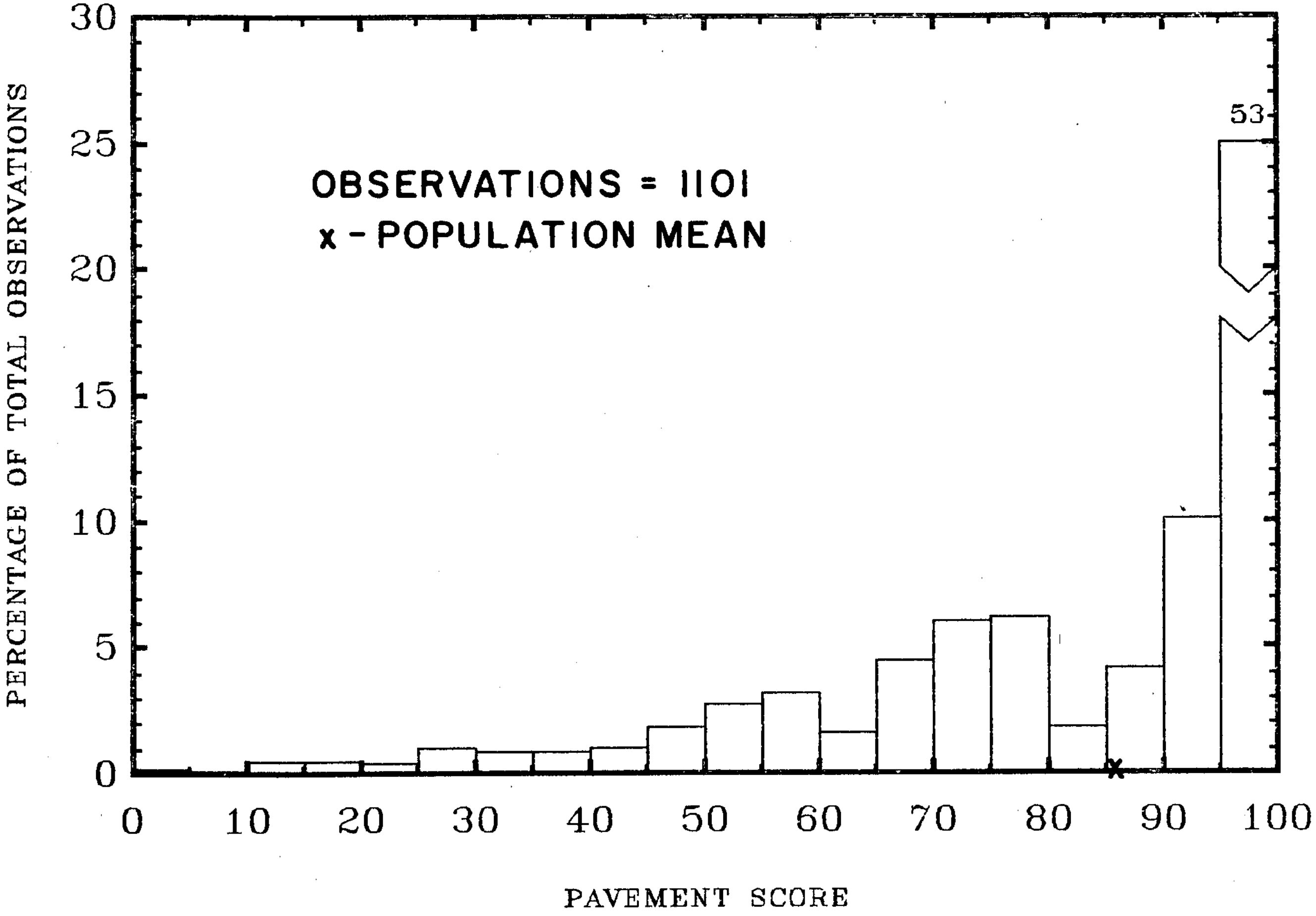


FIGURE A-17. Pavement Score histogram, 100% sample for District 15, System FM

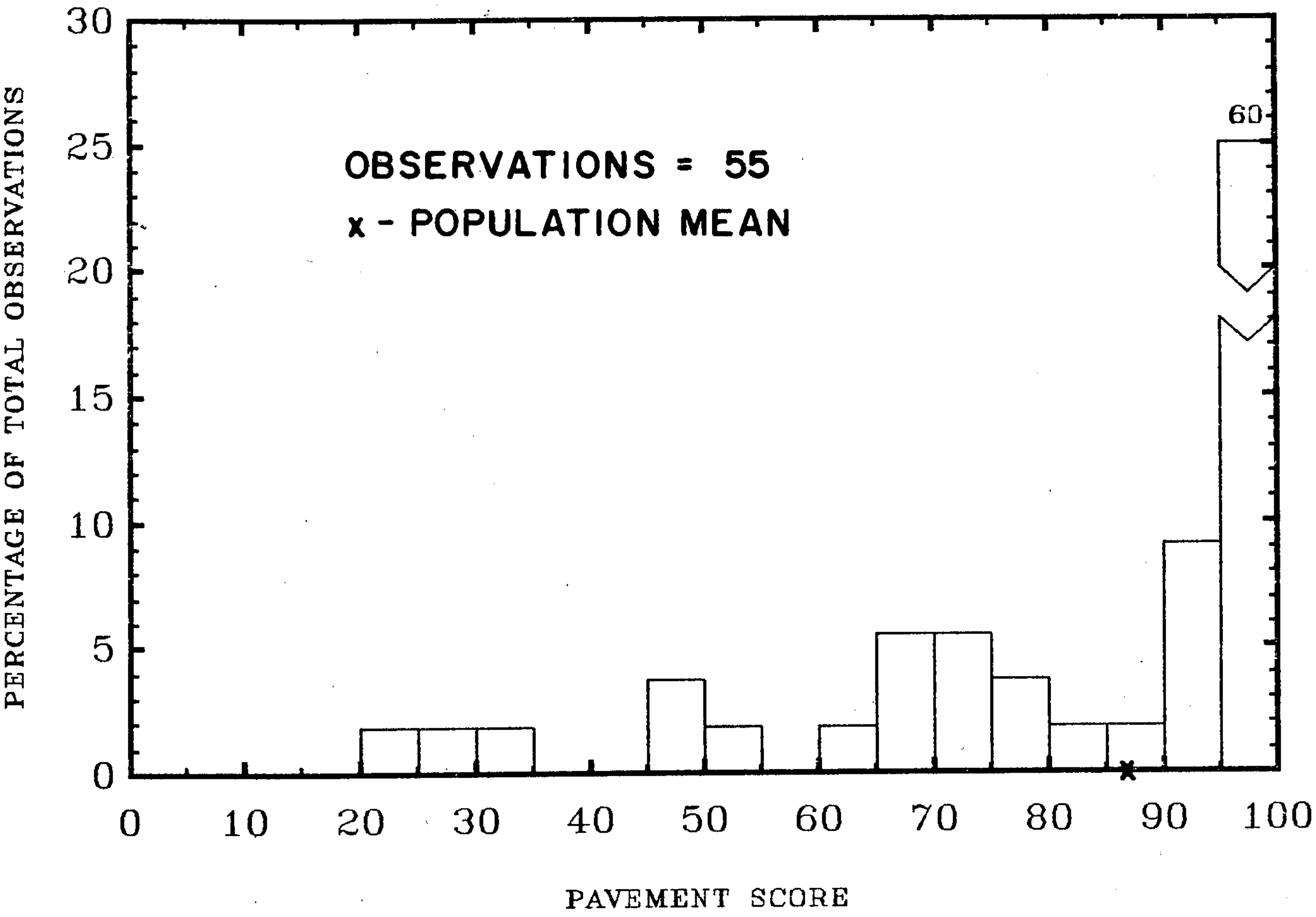


FIGURE A-18. Pavement Score histogram, 5% sample for District 15, System FM

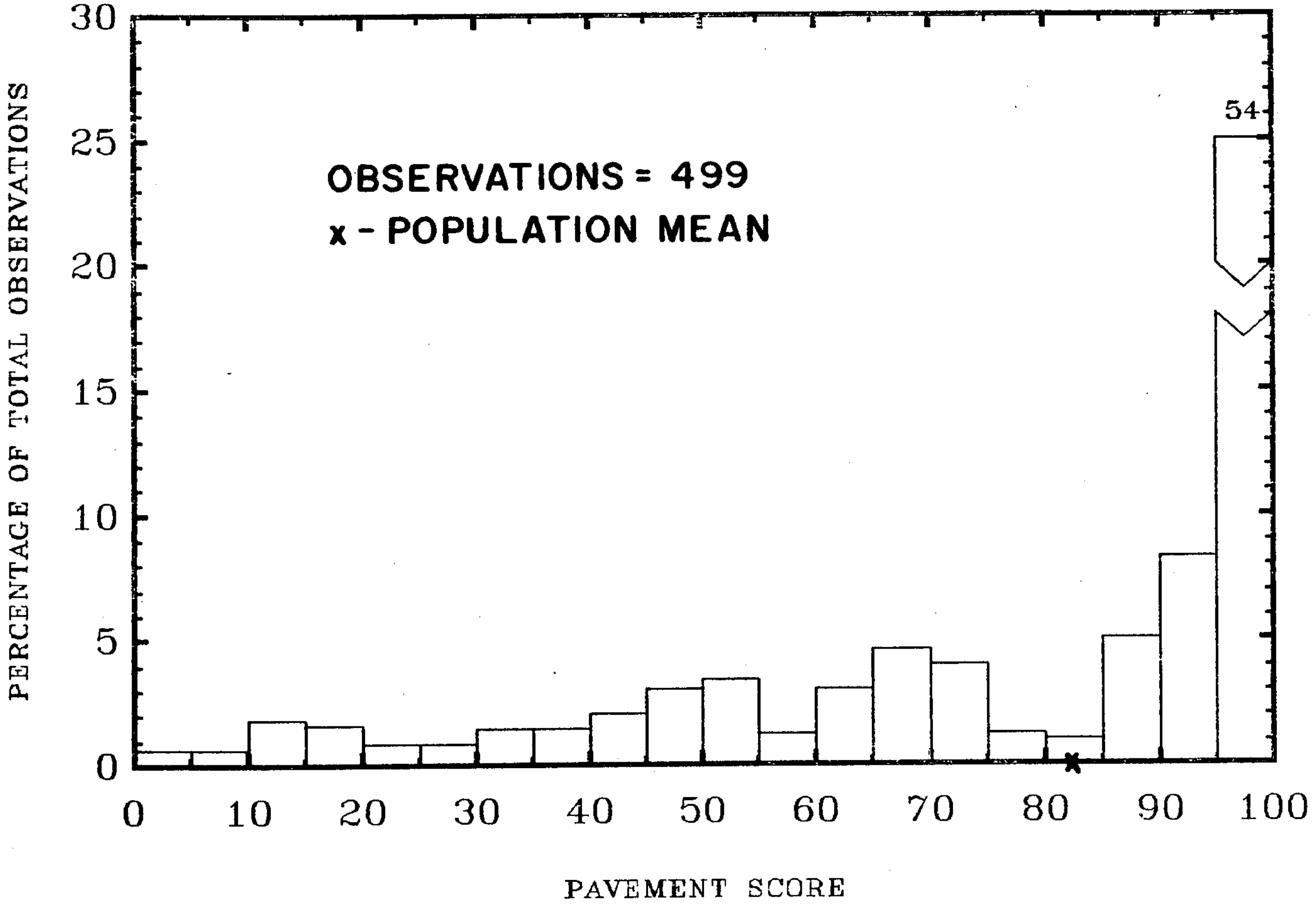


FIGURE A-19. Pavement Score histogram, 100% sample for District 15, System SH

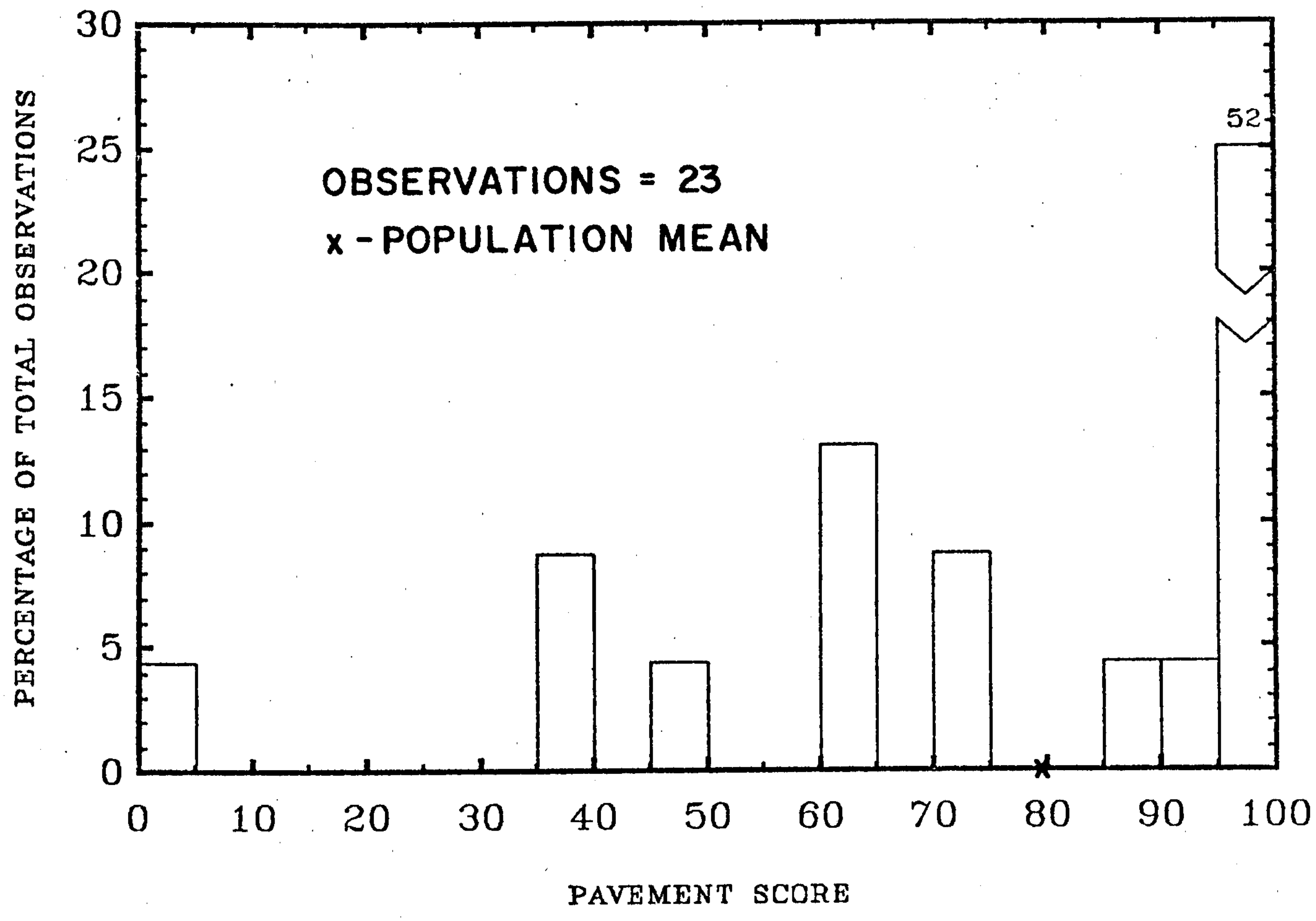


FIGURE A-20. Pavement Score histogram, 5% sample for District 15, System SH

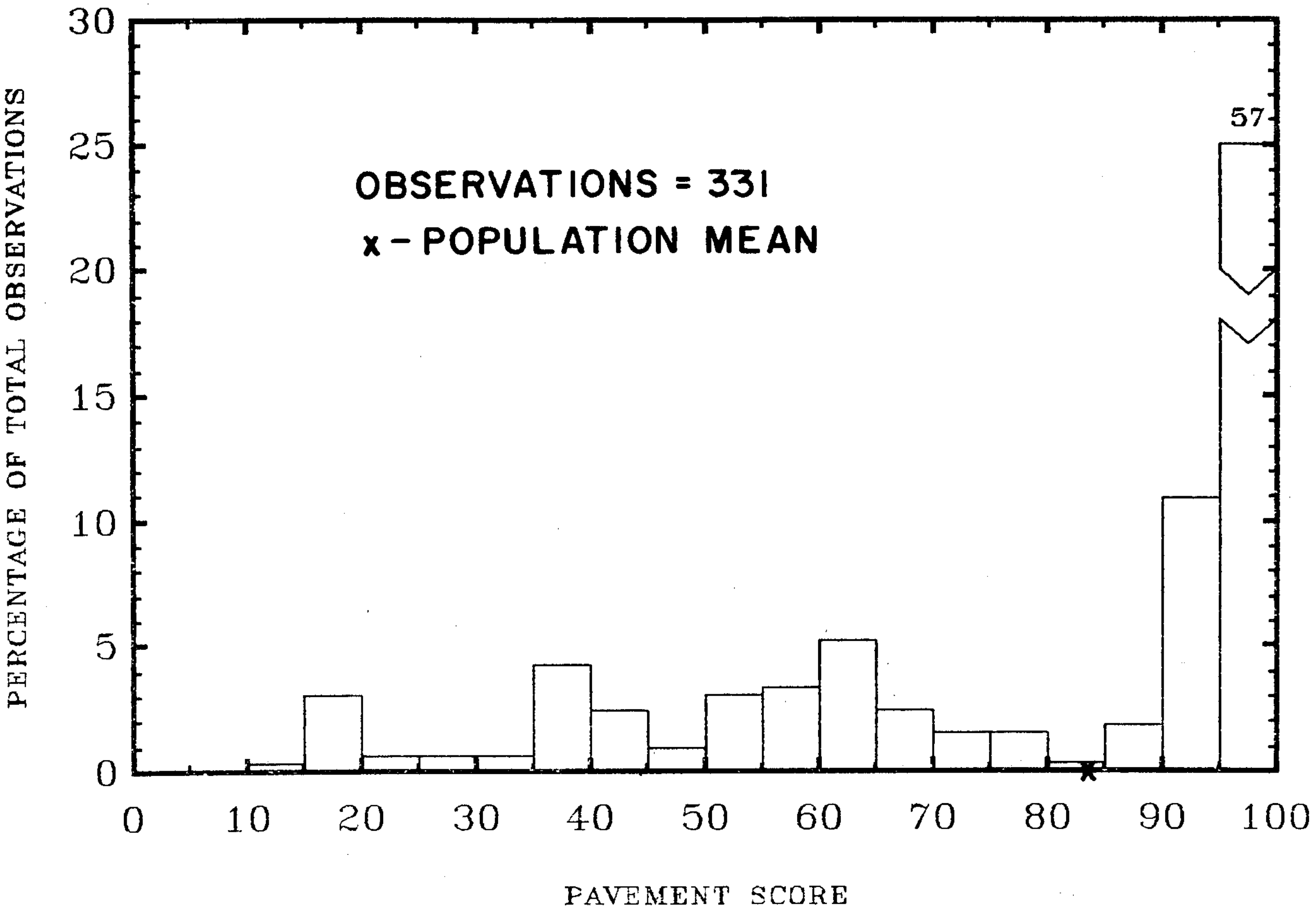


FIGURE A-21. Pavement Score histogram, 100% sample for District 15, System US

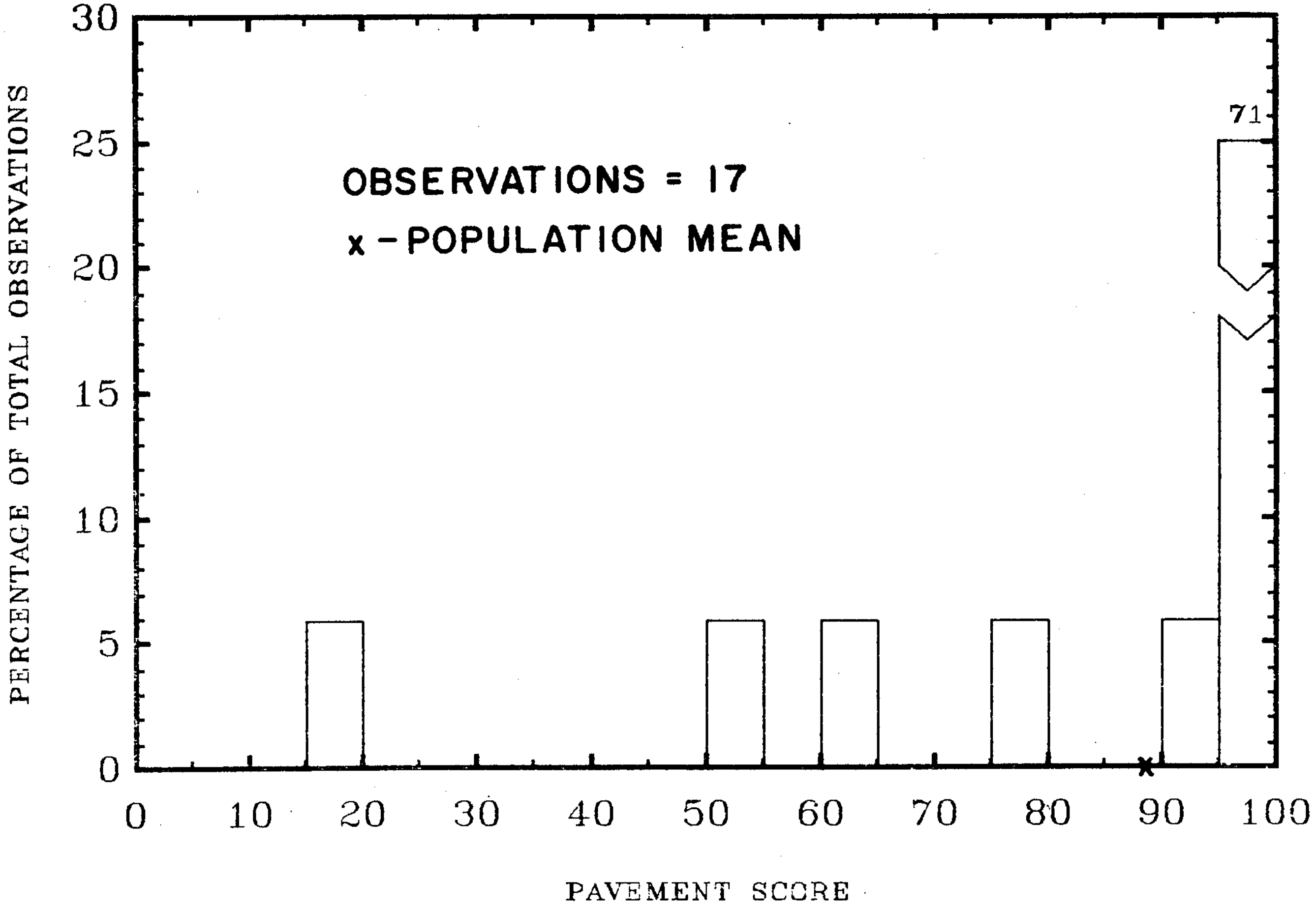


FIGURE A-22. Pavement Score histogram, 5% sample for District 15, System US

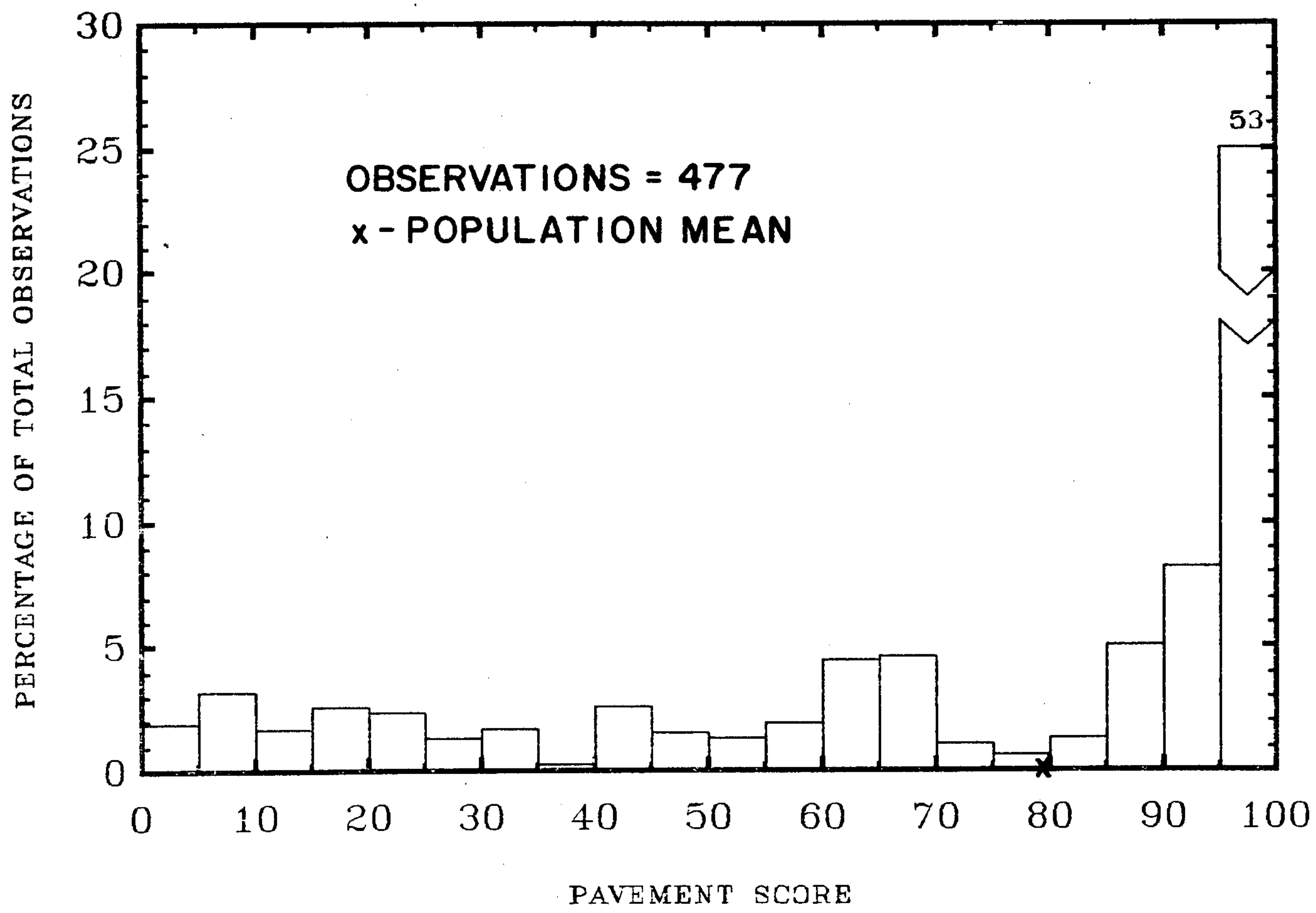


FIGURE A-23. Pavement Score histogram, 100% sample for District 8,15, System IH

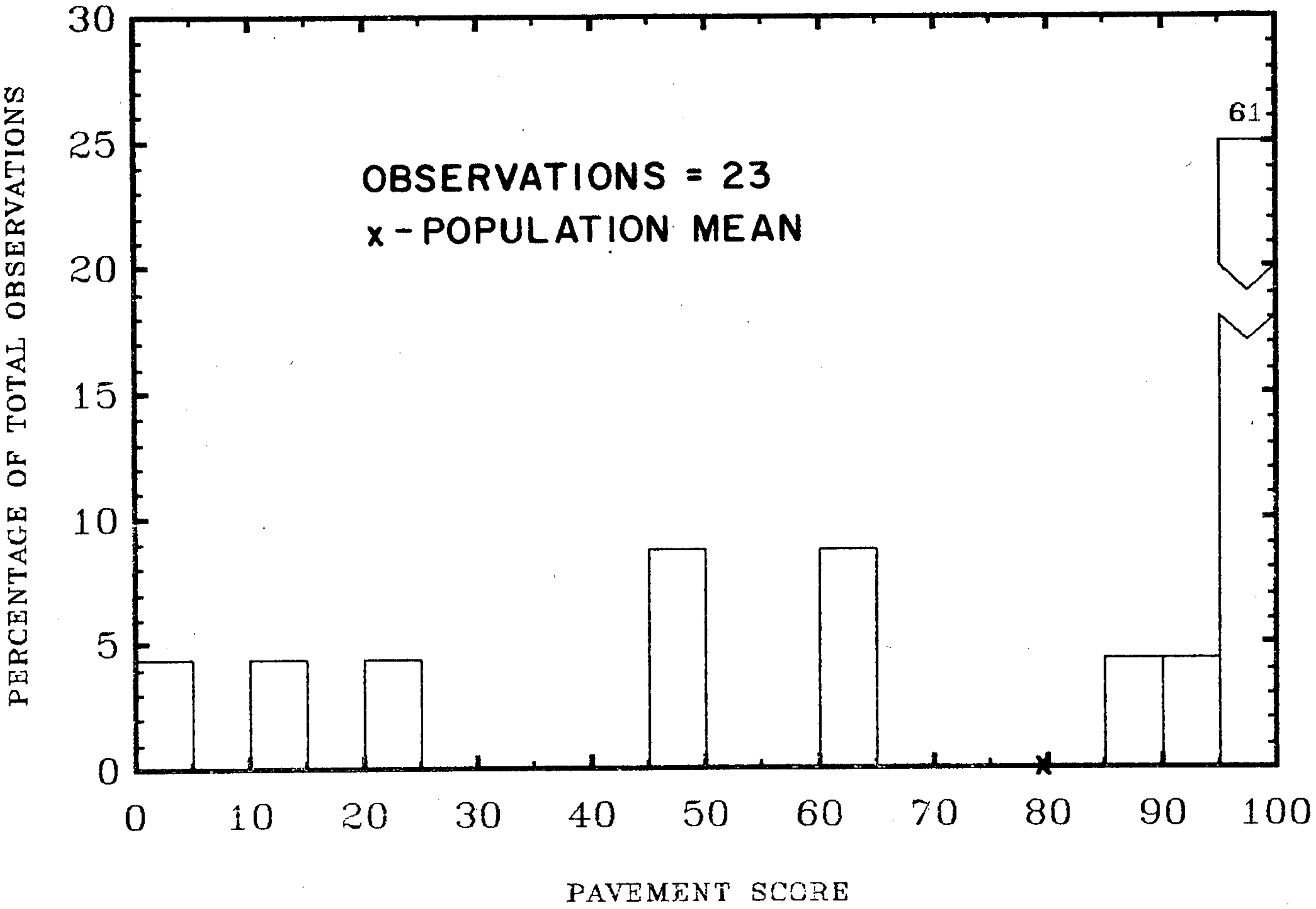


FIGURE A-24. Pavement Score histogram, 5% sample for District 8,15, System IH

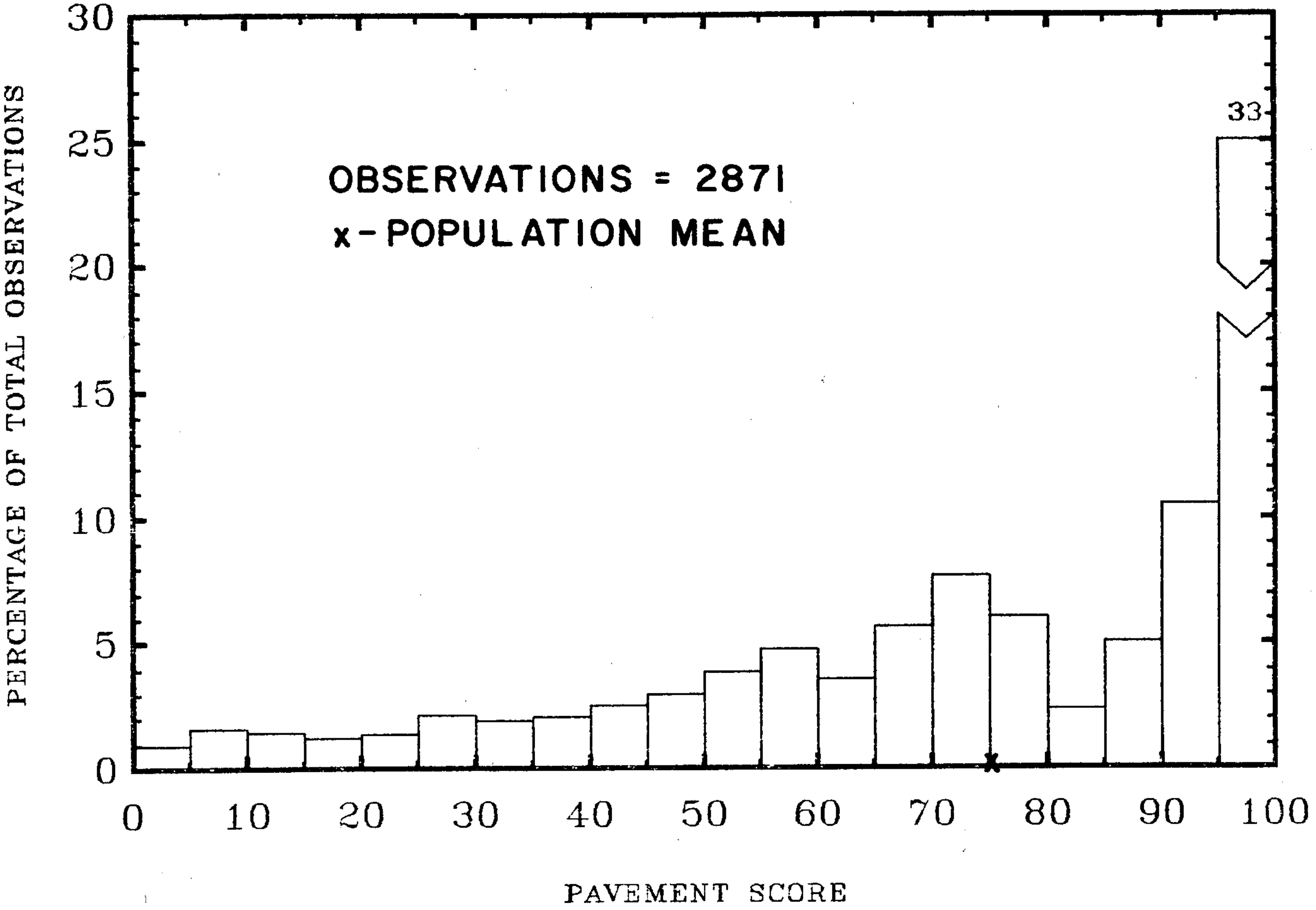


FIGURE A-25. Pavement Score histogram, 100% sample for Districts 8,11,15, System FM

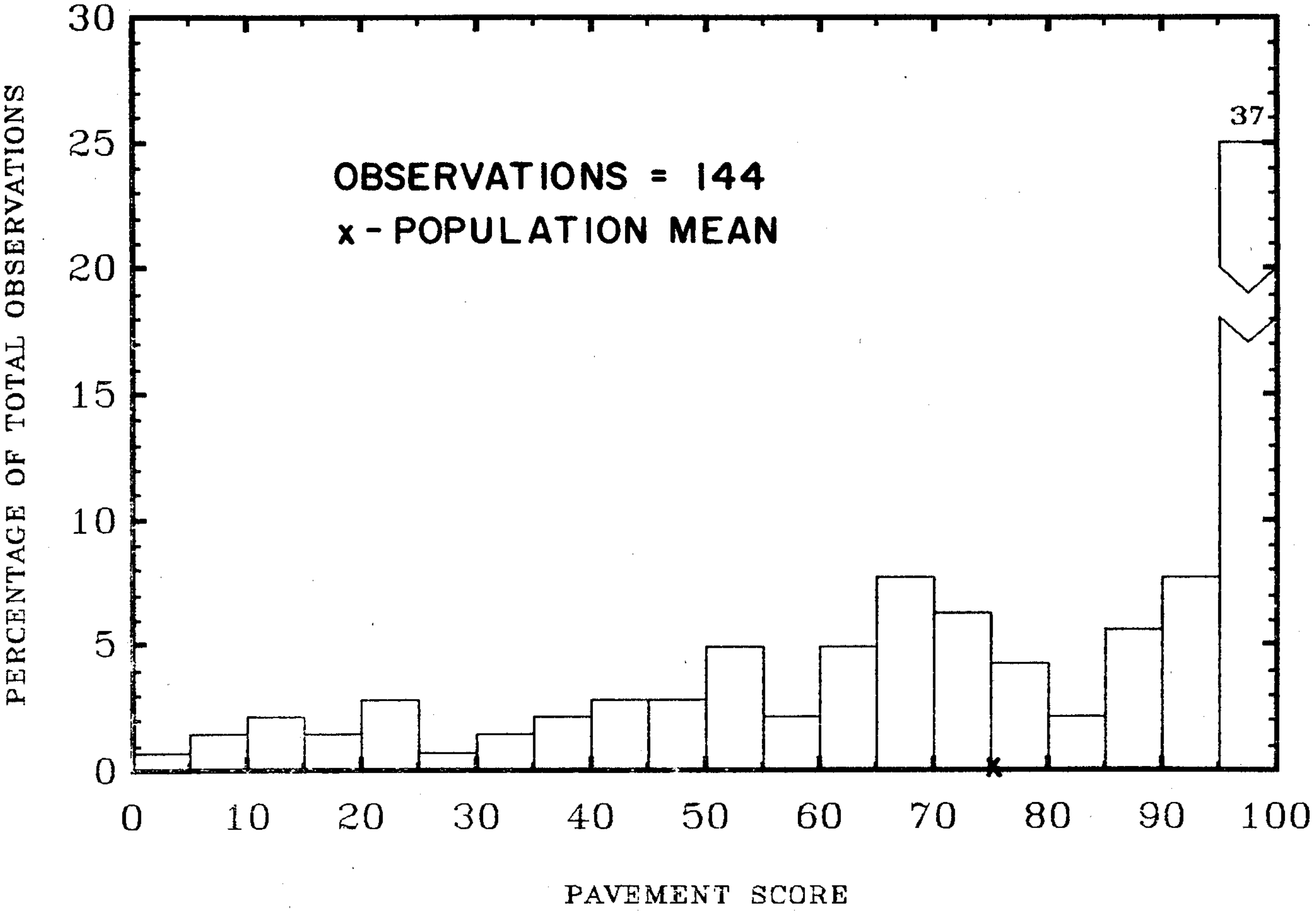


FIGURE A-26. Pavement Score histogram, 5% sample for Districts 8,11,15, System FM

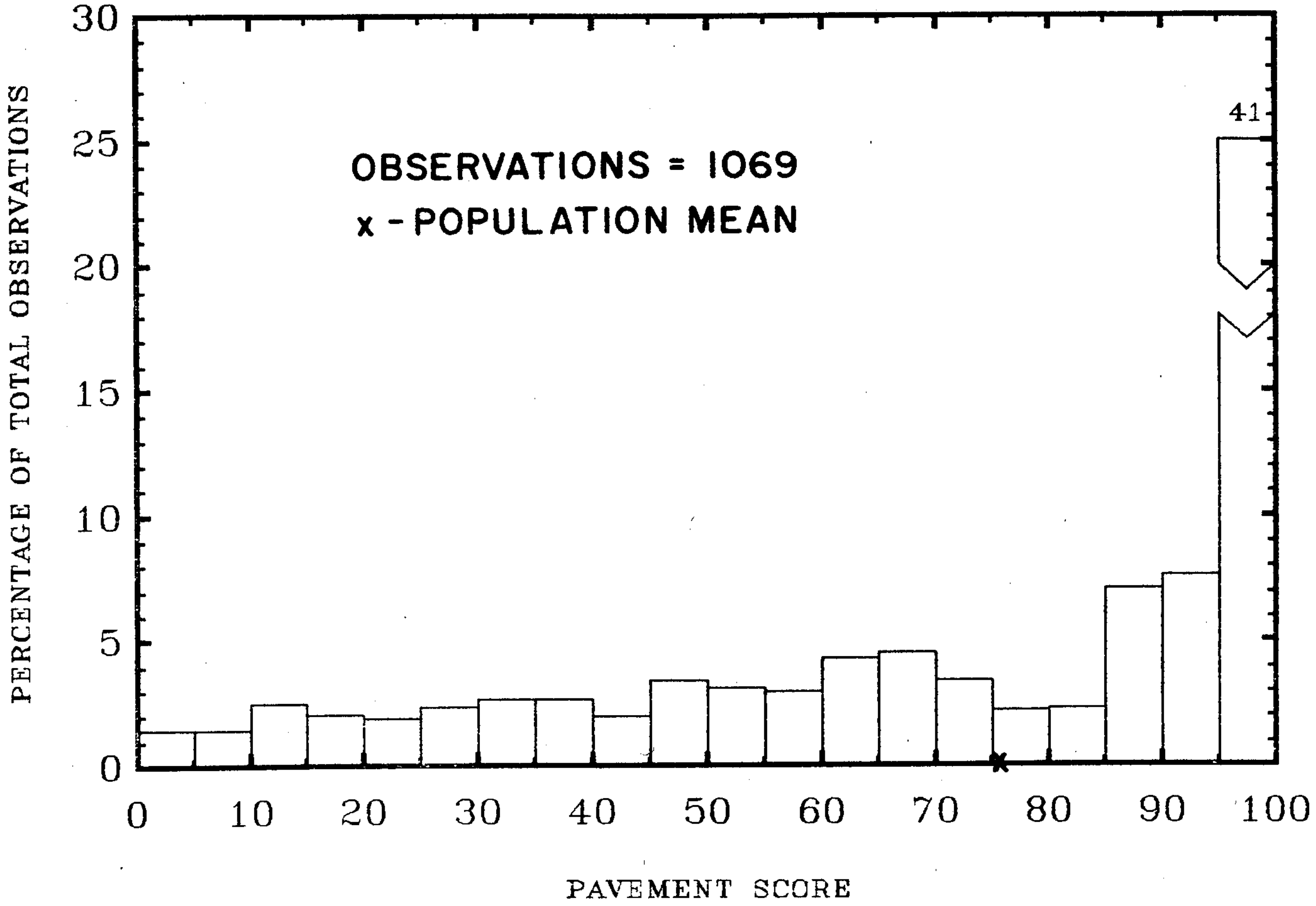


FIGURE A-27. Pavement Score histogram, 100% sample for District 8,11,15, System SH

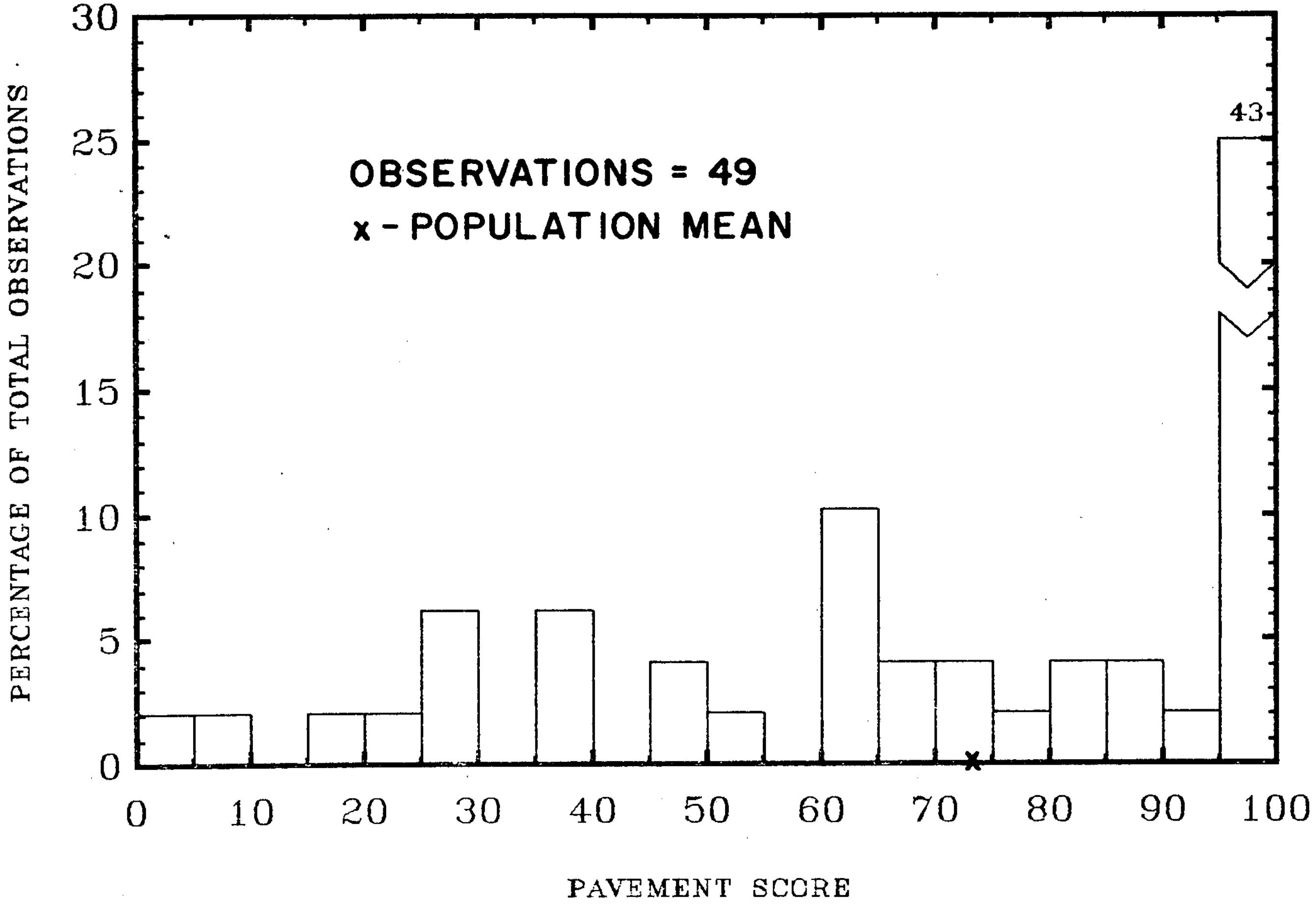


FIGURE A-28. Pavement Score histogram, 5% sample for Districts 8,11,15, System SH

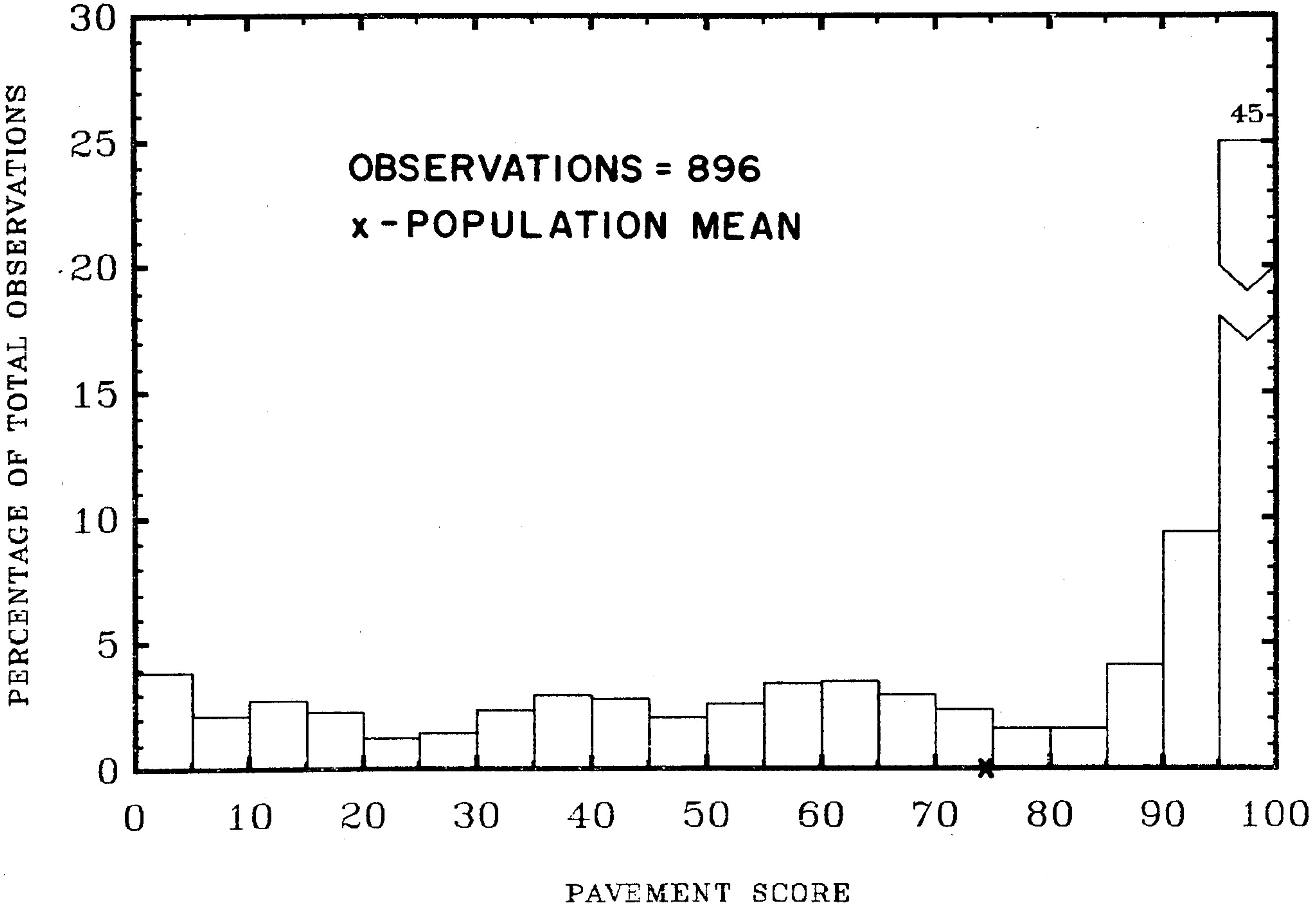


FIGURE A-29. Pavement Score histogram, 100% sample for Districts 8,11,15, System US

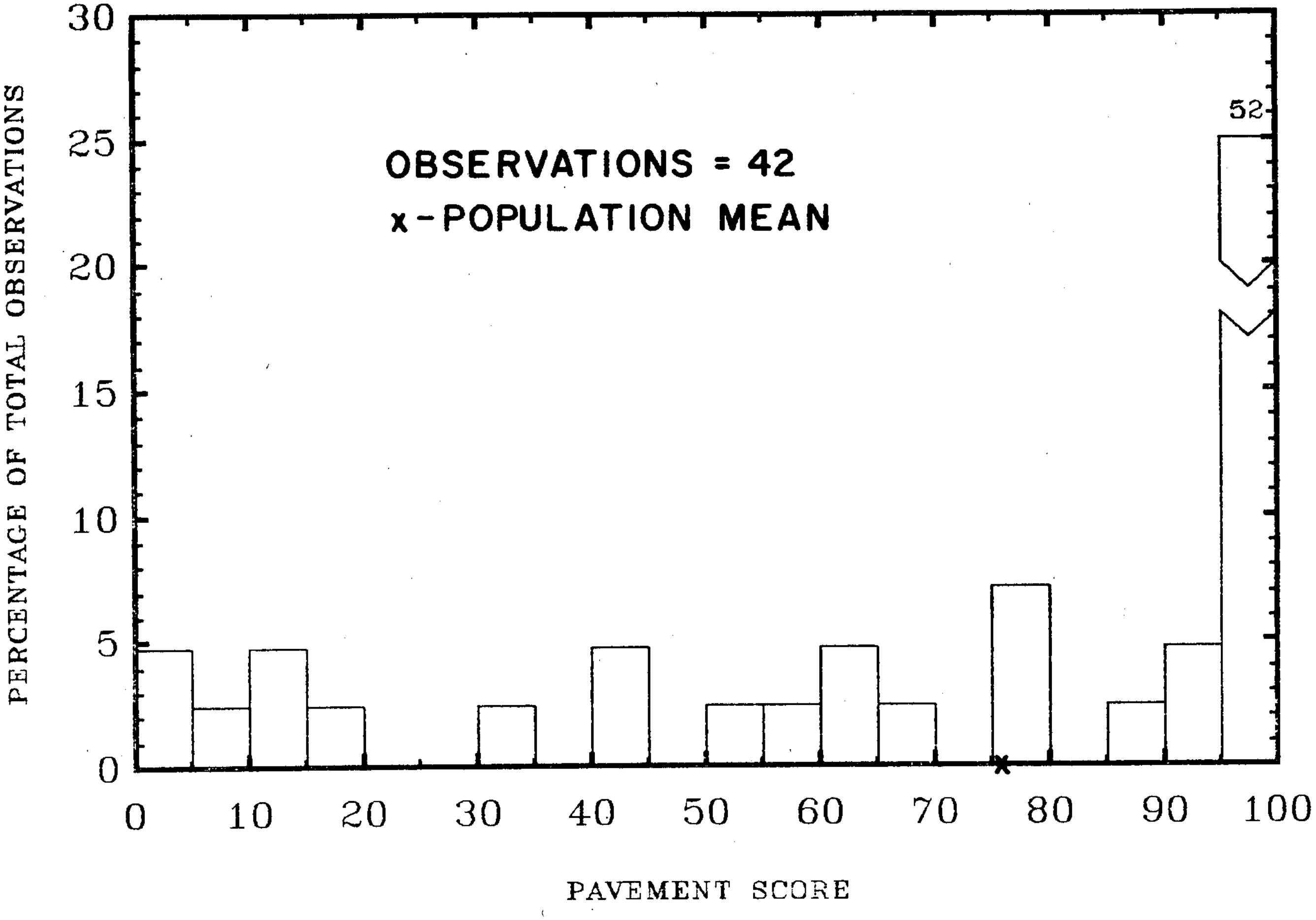


FIGURE A-30. Pavement Score histogram, 5% sample for Districts 8,11,15, System US

TABLE A-1. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 8, System IH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.1250	.0584	.0666
6-10	.1250	.1493	.0243
11-15	.2500	.1948	.0552
16-20	.2500	.2532	.0032
21-25	.3750	.3181	.0569
26-30	.3750	.3441	.0309
31-35	.3750	.3506	.0244
36-40	.3750	.3506	.0244
41-45	.3750	.3766	.0016
46-50	.3750	.4026	.0276
51-55	.3750	.4416	.0666
56-60	.3750	.4611	.0861
61-65	.5000	.4871	.0129
66-70	.5000	.4871	.0129
71-75	.5000	.5196	.0196
76-80	.5000	.5391	.0391
81-85	.5000	.5521	.0521
86-90	.5000	.6560	.1560*
91-95	.6250	.7404	.1154
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-2. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 8, System FM

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0053	.0053
6-10	.0208	.0149	.0059
11-15	.0625	.0288	.0337
16-20	.0625	.0427	.0198
21-25	.0833	.0544	.0289
26-30	.0833	.0789	.0044
31-35	.1041	.1024	.0017
36-40	.1249	.1269	.0020
41-45	.1666	.1653	.0013
46-50	.1874	.1973	.0099
51-55	.2916	.2453	.0463
56-60	.3124	.3071	.0053
61-65	.3332	.3423	.0091
66-70	.4165	.3988	.0177
71-75	.4998	.5001	.0003
76-80	.4998	.5459	.0461
81-85	.5415	.5619	.0204
86-90	.6665	.6099	.0566
91-95	.7915	.7314	.0601*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-3. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 8, System SH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0111	.0111
6-10	.0000	.0333	.0333
11-15	.0000	.0666	.0666
16-20	.0000	.0962	.0962
21-25	.0000	.1258	.1258
26-30	.1667	.1777	.0110
31-35	.1667	.2184	.0517
36-40	.2500	.2517	.0017
41-45	.2500	.2591	.0091
46-50	.2500	.3035	.0535
51-55	.2500	.3146	.0646
56-60	.2500	.3368	.0868
61-65	.2500	.3850	.1350*
66-70	.4167	.4406	.0239
71-75	.4167	.4776	.0609
76-80	.5000	.4887	.0113
81-85	.5000	.5035	.0035
86-90	.5833	.5813	.0020
91-95	.5833	.6702	.0869
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-4. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 8, System US

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0216	.0216
6-10	.0000	.0309	.0309
11-15	.0667	.0679	.0012
16-20	.0667	.0710	.0043
21-25	.0667	.0803	.0136
26-30	.0667	.0988	.0321
31-35	.0667	.1235	.0568
36-40	.0667	.1359	.0692
41-45	.1334	.1606	.0272
46-50	.1334	.1730	.0396
51-55	.1334	.1884	.0550
56-60	.2001	.2254	.0253
61-65	.2668	.2532	.0136
66-70	.3335	.2933	.0402
71-75	.3335	.3211	.0124
76-80	.4002	.3304	.0698*
81-85	.4002	.3458	.0544
86-90	.4669	.4106	.0563
91-95	.5336	.5279	.0057
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-5. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 11, System FM

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0244	.0216	.0028
6-10	.0488	.0649	.0161
11-15	.0732	.0938	.0206
16-20	.1220	.1142	.0078
21-25	.1708	.1431	.0277
26-30	.1708	.1756	.0048
31-35	.1708	.2032	.0324
36-40	.2196	.2357	.0161
41-45	.2684	.2658	.0026
46-50	.2928	.3055	.0127
51-55	.3172	.3476	.0304
56-60	.3660	.3993	.0333
61-65	.4880	.4606	.0274
66-70	.5856	.5315	.0541
71-75	.6344	.6024	.0320
76-80	.7320	.6745	.0575
81-85	.7320	.7118	.0202
86-90	.7564	.7779	.0215
91-95	.7564	.8693	.1129*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-6. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 11, System SH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0300	.0300
6-10	.0714	.0500	.0214
11-15	.0714	.0767	.0053
16-20	.1428	.0967	.0461
21-25	.2142	.1234	.0908
26-30	.2856	.1467	.1389*
31-35	.2856	.1800	.1056
36-40	.2856	.2200	.0656
41-45	.2856	.2500	.0356
46-50	.3570	.2800	.0770
51-55	.4284	.3233	.1051
56-60	.4284	.3866	.0418
61-65	.5713	.4466	.1247
66-70	.5713	.4799	.0914
71-75	.5713	.4999	.0714
76-80	.5713	.5466	.0247
81-85	.7142	.5966	.1176
86-90	.7142	.6933	.0209
91-95	.7142	.7466	.0324
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-7. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 11, System US

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.2000	.1120	.0880
6-10	.3000	.1784	.1216
11-15	.4000	.2240	.1760
16-20	.4000	.2613	.1387
21-25	.4000	.2862	.1138
26-30	.4000	.3070	.0930
31-35	.5000	.3526	.1474
36-40	.5000	.3858	.1142
41-45	.6000	.4231	.1769*
46-50	.6000	.4687	.1313
51-55	.6000	.5019	.0981
56-60	.6000	.5310	.0690
61-65	.6000	.5518	.0482
66-70	.6000	.5726	.0274
71-75	.6000	.6017	.0017
76-80	.7000	.6266	.0734
81-85	.7000	.6598	.0402
86-90	.7000	.7013	.0013
91-95	.7000	.7428	.0428
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-8. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 15, System IH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0000	.0000
6-10	.0000	.0031	.0031
11-15	.0000	.0062	.0062
16-20	.0000	.0155	.0155
21-25	.0000	.0186	.0186
26-30	.0000	.0248	.0248
31-35	.0000	.0465	.0465
36-40	.0000	.0496	.0496
41-45	.0000	.0744	.0744*
46-50	.1333	.0837	.0496
51-55	.1333	.0837	.0496
56-60	.1333	.1023	.0310
61-65	.2000	.1549	.0451
66-70	.2000	.2230	.0230
71-75	.2000	.2230	.0230
76-80	.2000	.2230	.0230
81-85	.2000	.2354	.0354
86-90	.2667	.2602	.0065
91-95	.2667	.3407	.0740
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-9. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 15, System FM

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0018	.0018
6-10	.0000	.0027	.0027
11-15	.0000	.0072	.0072
16-20	.0000	.0117	.0117
21-25	.0182	.0153	.0029
26-30	.0364	.0253	.0111
31-35	.0546	.0335	.0211
36-40	.0546	.0417	.0129
41-45	.0546	.0517	.0029
46-50	.0910	.0699	.0211
51-55	.1092	.0972	.0120
56-60	.1092	.1290	.0198
61-65	.1274	.1444	.0170
66-70	.1820	.1889	.0069
71-75	.2366	.2489	.0123
76-80	.2730	.3107	.0377
81-85	.2912	.3289	.0377
86-90	.3094	.3698	.0604
91-95	.4003	.4706	.0703*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-10. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 15, System SH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0435	.0060	.0375
6-10	.0435	.0120	.0315
11-15	.0435	.0300	.0135
16-20	.0435	.0460	.0025
21-25	.0435	.0540	.0105
26-30	.0435	.0620	.0185
31-35	.0435	.0760	.0325
36-40	.1305	.0900	.0405
41-45	.1305	.1100	.0205
46-50	.1740	.1401	.0339
51-55	.1740	.1742	.0002
56-60	.1740	.1862	.0122
61-65	.3044	.2163	.0881
66-70	.3044	.2624	.0420
71-75	.3914	.3025	.0889*
76-80	.3914	.3145	.0769
81-85	.3914	.3245	.0669
86-90	.4349	.3746	.0603
91-95	.4784	.4568	.0216
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-11. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in District 15, System US

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0000	.0000	.0000
6-10	.0000	.0000	.0000
11-15	.0000	.0030	.0030
16-20	.0588	.0332	.0256
21-25	.0588	.0392	.0196
26-30	.0588	.0452	.0136
31-35	.0588	.0512	.0076
36-40	.0588	.0935	.0347
41-45	.0588	.1177	.0589
46-50	.0588	.1268	.0680
51-55	.1176	.1570	.0394
56-60	.1176	.1902	.0726
61-65	.1764	.2416	.0652
66-70	.1764	.2658	.0894
71-75	.1764	.2809	.1045
76-80	.2352	.2960	.0608
81-85	.2352	.2990	.0638
86-90	.2352	.3171	.0819
91-95	.2940	.4259	.1319*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-12. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in Districts 8,15, System IH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0435	.0189	.0246
6-10	.0435	.0504	.0069
11-15	.0870	.0672	.0198
16-20	.0870	.0924	.0054
21-25	.1305	.1155	.0150
26-30	.1305	.1281	.0024
31-35	.1305	.1449	.0144
36-40	.1305	.1470	.0165
41-45	.1305	.1722	.0417
46-50	.2175	.1869	.0306
51-55	.2175	.1995	.0180
56-60	.2175	.2184	.0009
61-65	.3045	.2624	.0421
66-70	.3045	.3085	.0040
71-75	.3045	.3190	.0145
76-80	.3045	.3253	.0208
81-85	.3045	.3379	.0334
86-90	.3480	.3882	.0402
91-95	.3915	.4700	.0785*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-13. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in Districts 8,11,15, System FM

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0069	.0087	.0018
6-10	.0208	.0247	.0039
11-15	.0416	.0393	.0023
16-20	.0555	.0515	.0040
21-25	.0833	.0651	.0182
26-30	.0902	.0864	.0038
31-35	.1041	.1052	.0011
36-40	.1249	.1258	.0009
41-45	.1527	.1509	.0018
46-50	.1805	.1798	.0007
51-55	.2291	.2181	.0110
56-60	.2499	.2655	.0156
61-65	.2985	.3007	.0022
66-70	.3749	.3568	.0181
71-75	.4374	.4334	.0040
76-80	.4791	.4931	.0140
81-85	.4999	.5161	.0162
86-90	.5555	.5666	.0111
91-95	.6319	.6714	.0395*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-14. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in Districts 8,11,15, System SH

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0204	.0140	.0064
6-10	.0408	.0280	.0128
11-15	.0408	.0523	.0115
16-20	.0612	.0729	.0117
21-25	.0816	.0916	.0100
26-30	.1428	.1150	.0278
31-35	.1428	.1412	.0016
36-40	.2040	.1674	.0366
41-45	.2040	.1870	.0170
46-50	.2448	.2207	.0241
51-55	.2652	.2516	.0136
56-60	.2652	.2806	.0154
61-65	.3672	.3236	.0436
66-70	.4080	.3685	.0395
71-75	.4488	.4022	.0466
76-80	.4692	.4237	.0455
81-85	.5100	.4462	.0638*
86-90	.5508	.5164	.0344
91-95	.5712	.5922	.0210
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

TABLE A-15. Comparison of five percent sample cumulative distribution and 100 percent sample cumulative distribution in Districts 8,11,15, System US

PAVEMENT SCORE DIVISION	RELATIVE CUMULATIVE FREQUENCY (5% SAMPLE) Sn(x)	RELATIVE CUMULATIVE FREQUENCY (100% SAMPLE) F(x)	|F(x)-Sn(x)|
0-5	.0476	.0380	.0096
6-10	.0714	.0592	.0122
11-15	.1190	.0860	.0330
16-20	.1428	.1083	.0345
21-25	.1428	.1206	.0222
26-30	.1428	.1351	.0077
31-35	.1666	.1585	.0081
36-40	.1666	.1875	.0209
41-45	.2142	.2154	.0012
46-50	.2142	.2355	.0213
51-55	.2380	.2612	.0232
56-60	.2618	.2947	.0329
61-65	.3094	.3293	.0199
66-70	.3332	.3583	.0251
71-75	.3332	.3817	.0485
76-80	.4046	.3973	.0073
81-85	.4046	.4129	.0083
86-90	.4284	.4542	.0258
91-95	.4760	.5480	.0720*
96-100	1.0000	1.0000	.0000

* Denotes Least Upper Bound on |F(x)-Sn(x)|.

APPENDIX B

APPENDIX B

Confidence Bands on F(x)

A valuable feature of the Kolmogorov-Smirnov test statistic is that its $(1-\alpha)$ percentile may be used to form a confidence band for the true cumulative distribution F(x). In order to form this band, the $(1-\alpha)$ percentile of the test statistic must be found. This percentile can be defined as:

$$K_{(1-\alpha)} = \frac{K_{T\ (1-\alpha)}}{\sqrt{n}} \tag{B-1}$$

where

$K_{(1-\alpha)}$ = $(1-\alpha)$ percentile of the Kolmogorov-Smirnov test statistic,

$K_{T(1-\alpha)}$ = table value relating to the $(1-\alpha)$ level of confidence, and

n = sample size.

The upper and lower limits of the band can then be calculated respectively as follows:

$$U(X) = S_n(X) + K_{(1-\alpha)} \tag{B-2}$$

and

$$L(X) = S_n(X) - K_{(1-\alpha)} \tag{B-3}$$

where

$U(X)$ = upper limit of the confidence band on $F(X)$,

$L(X)$ = lower limit of the confidence band on $F(X)$,

$S_n(X)$ = empirical cumulative distribution, and

$K_{(1-\alpha)}$ = $(1-\alpha)$ percentile of the Kolmogorov-Smirnov test statistic.

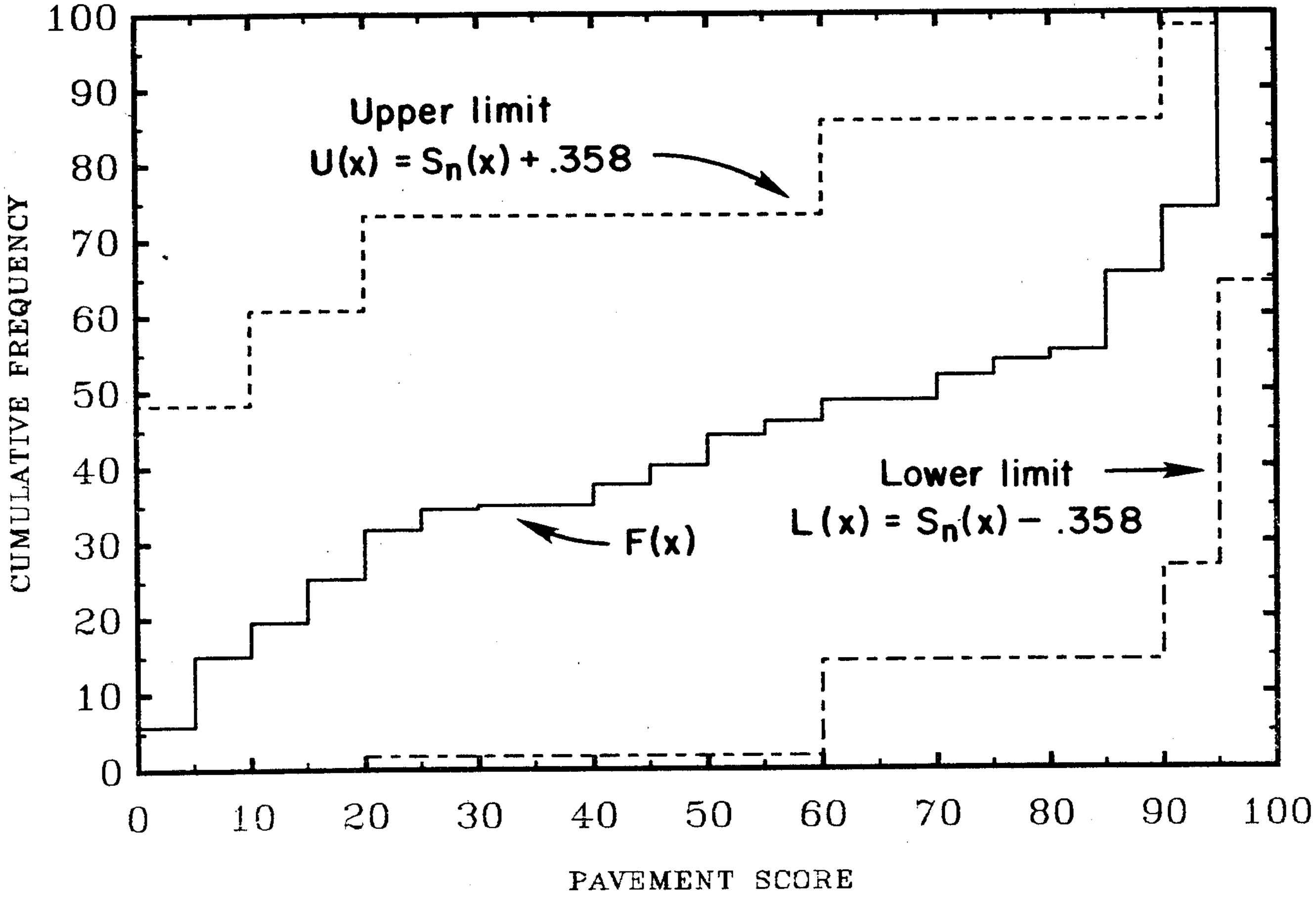


FIGURE B-1. Confidence band for cumulative distribution function of Pavement Scores F(X), District 8, System IH

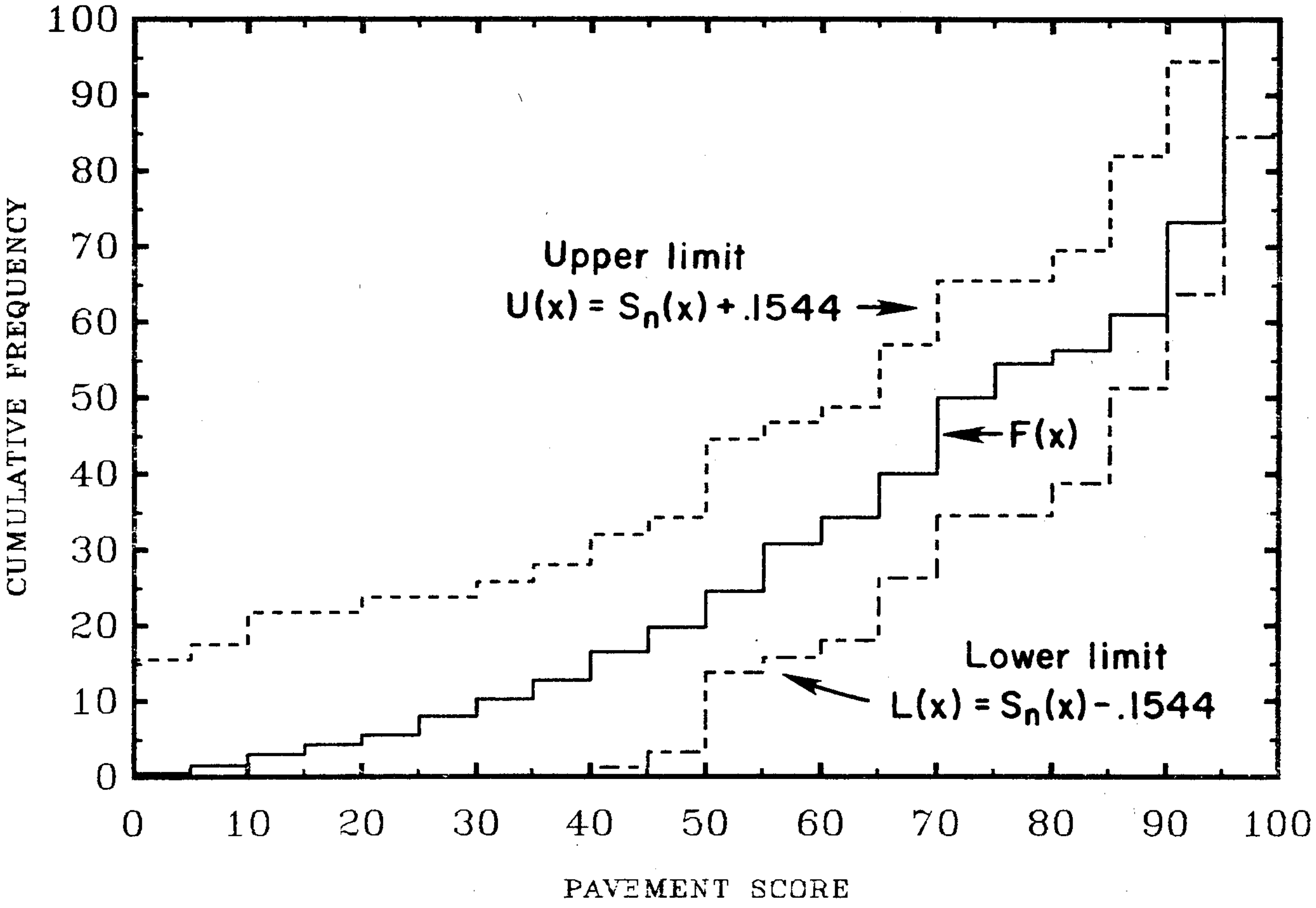


FIGURE B-2. Confidence band for cumulative distribution function of Pavement Scores F(X), District 8, System FM

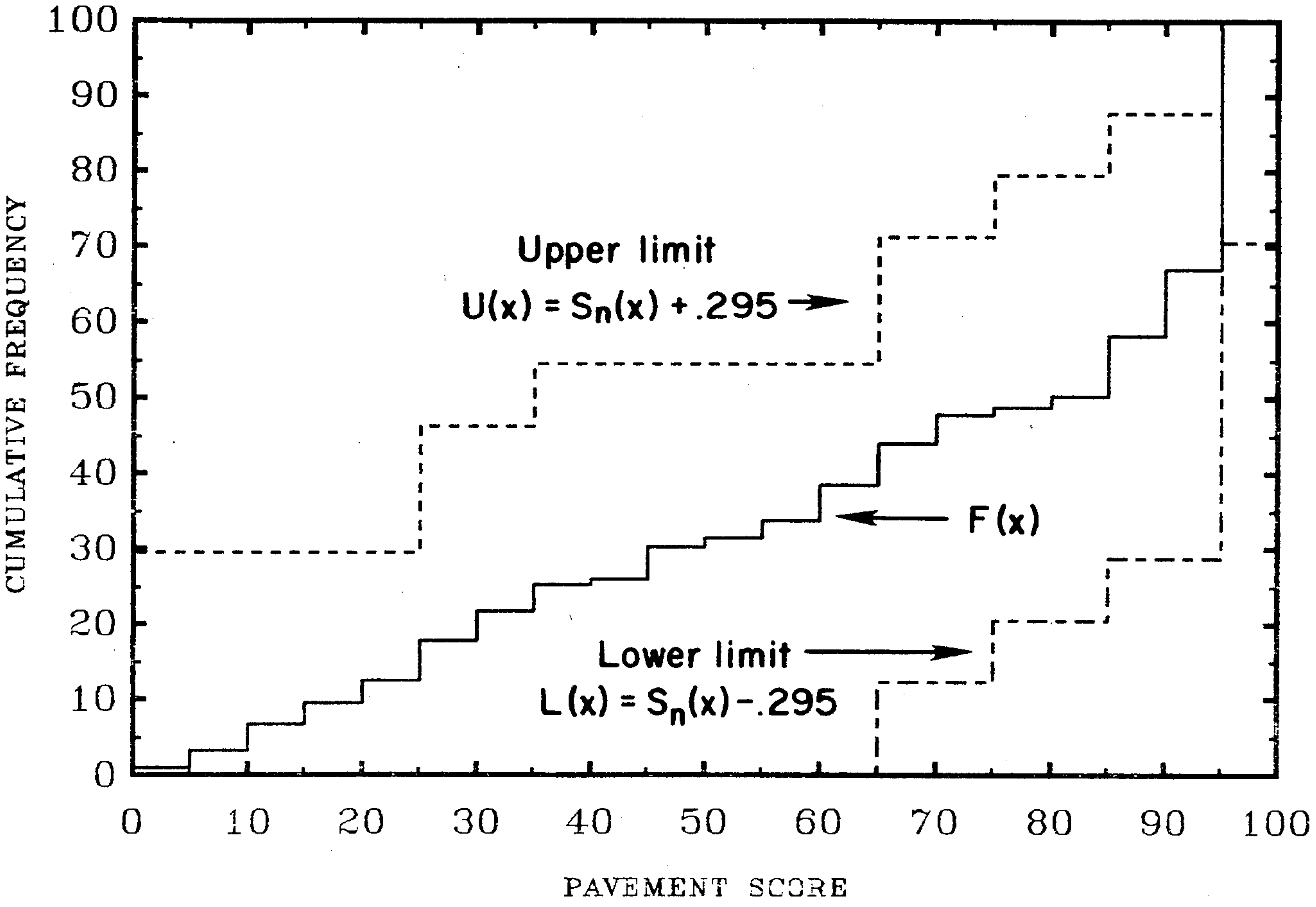


FIGURE B-3. Confidence band for cumulative distribution function of Pavement Scores F(X), District 8, System SH

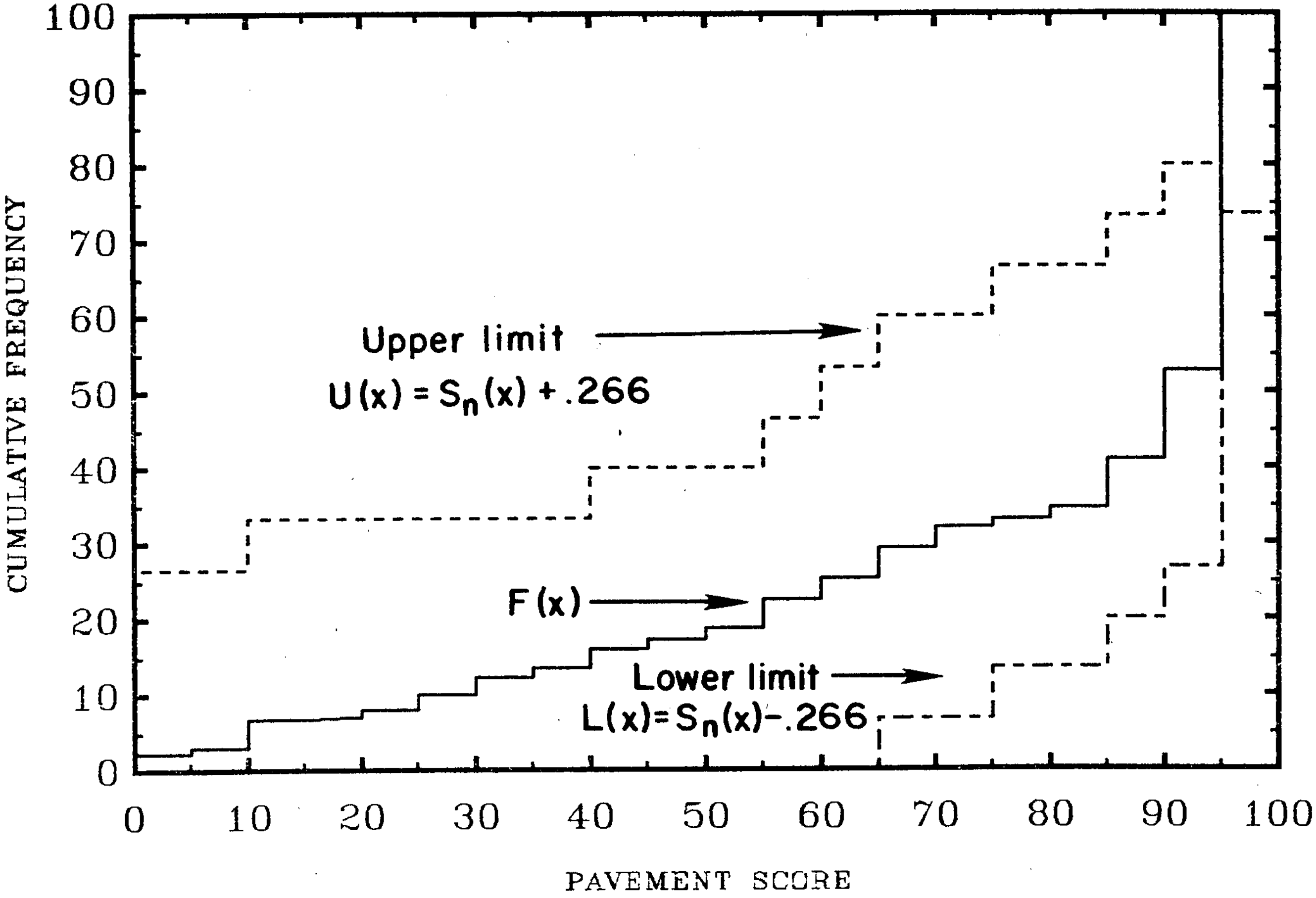


FIGURE B-4. Confidence band for cumulative distribution function of Pavement Scores F(X), District 8, System US

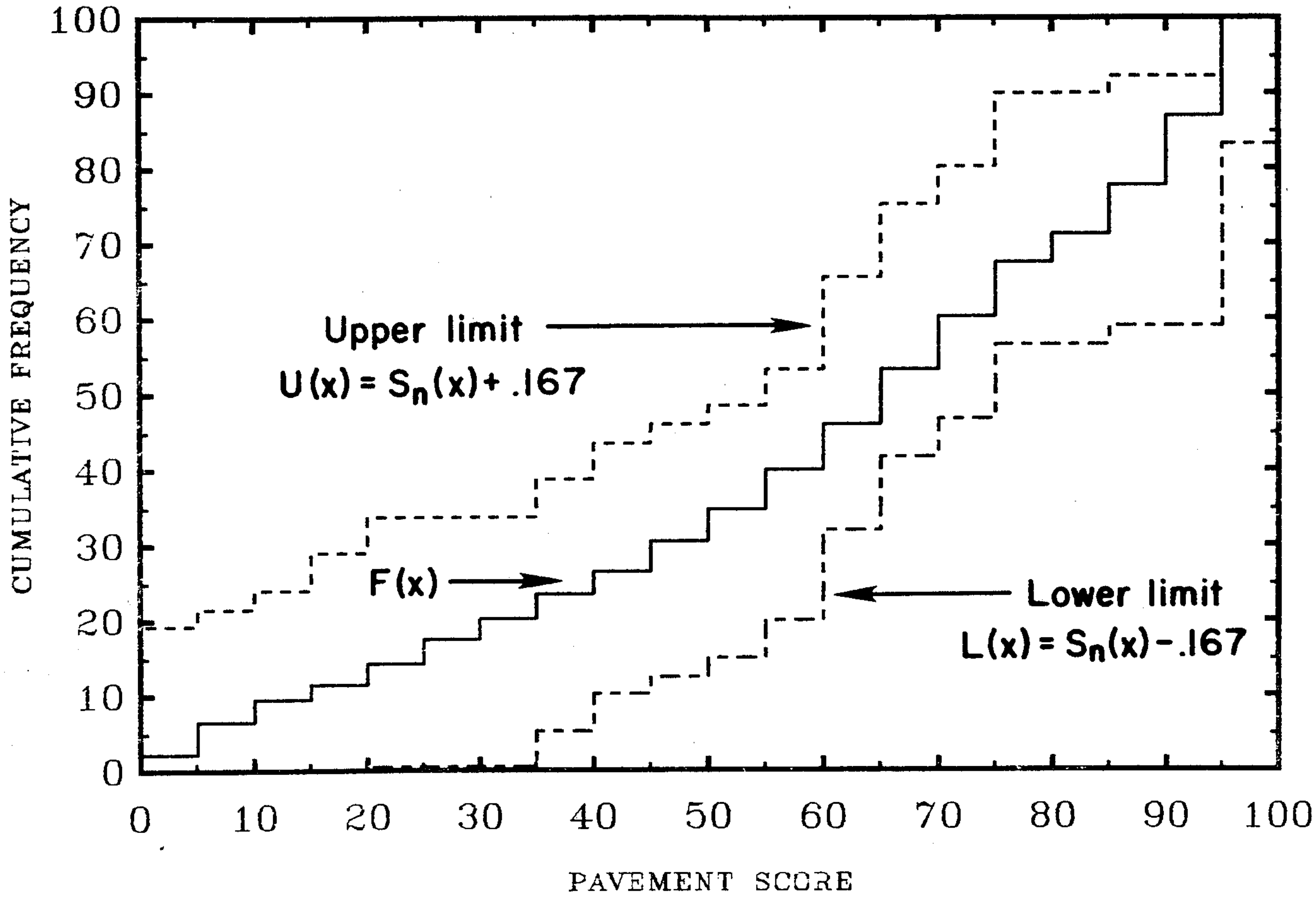


FIGURE B-5. Confidence band for cumulative distribution function of Pavement Scores F(X), District 11, System FM

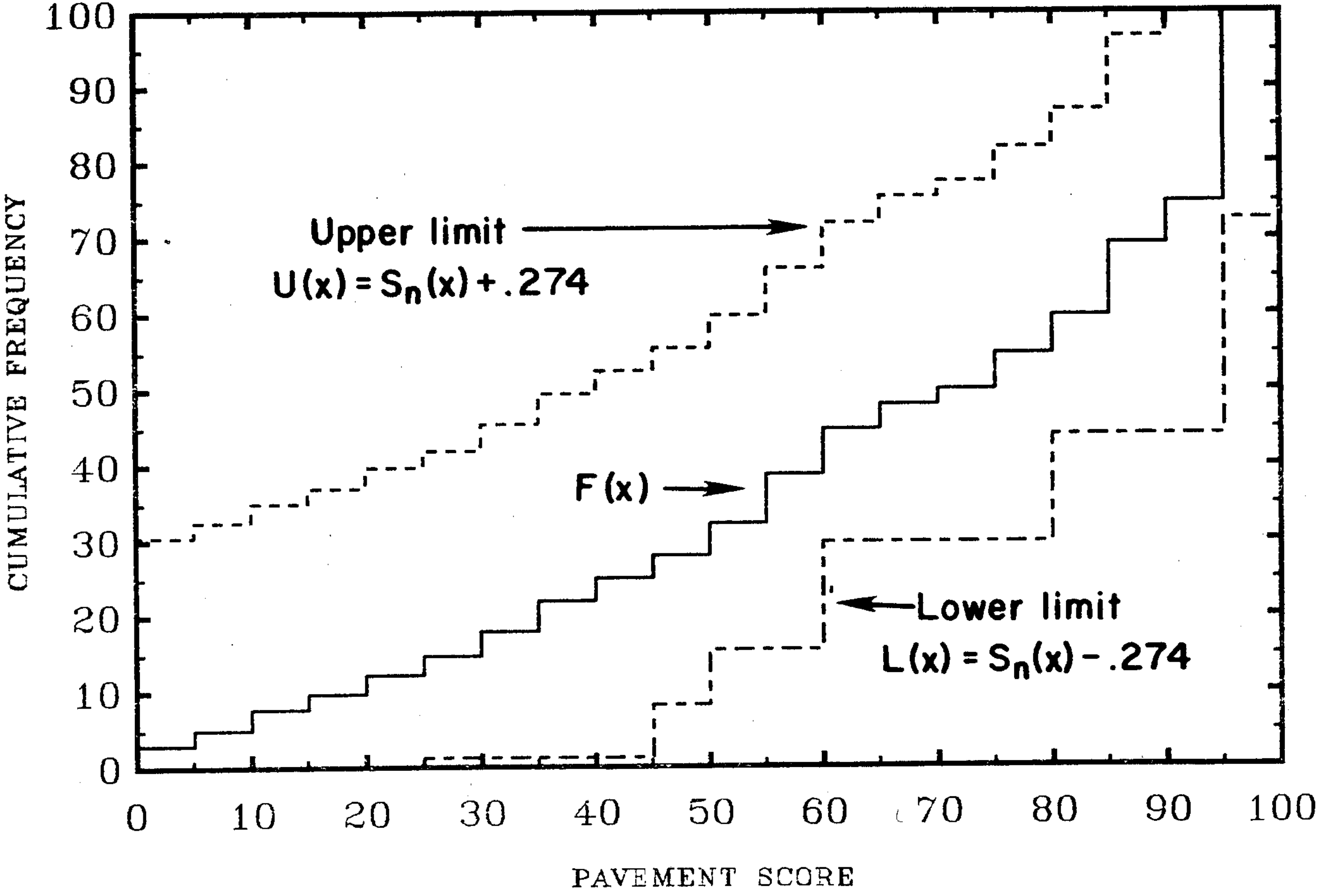


FIGURE B-6. Confidence band for cumulative distribution function of Pavement Scores F(X), District 11, System SH

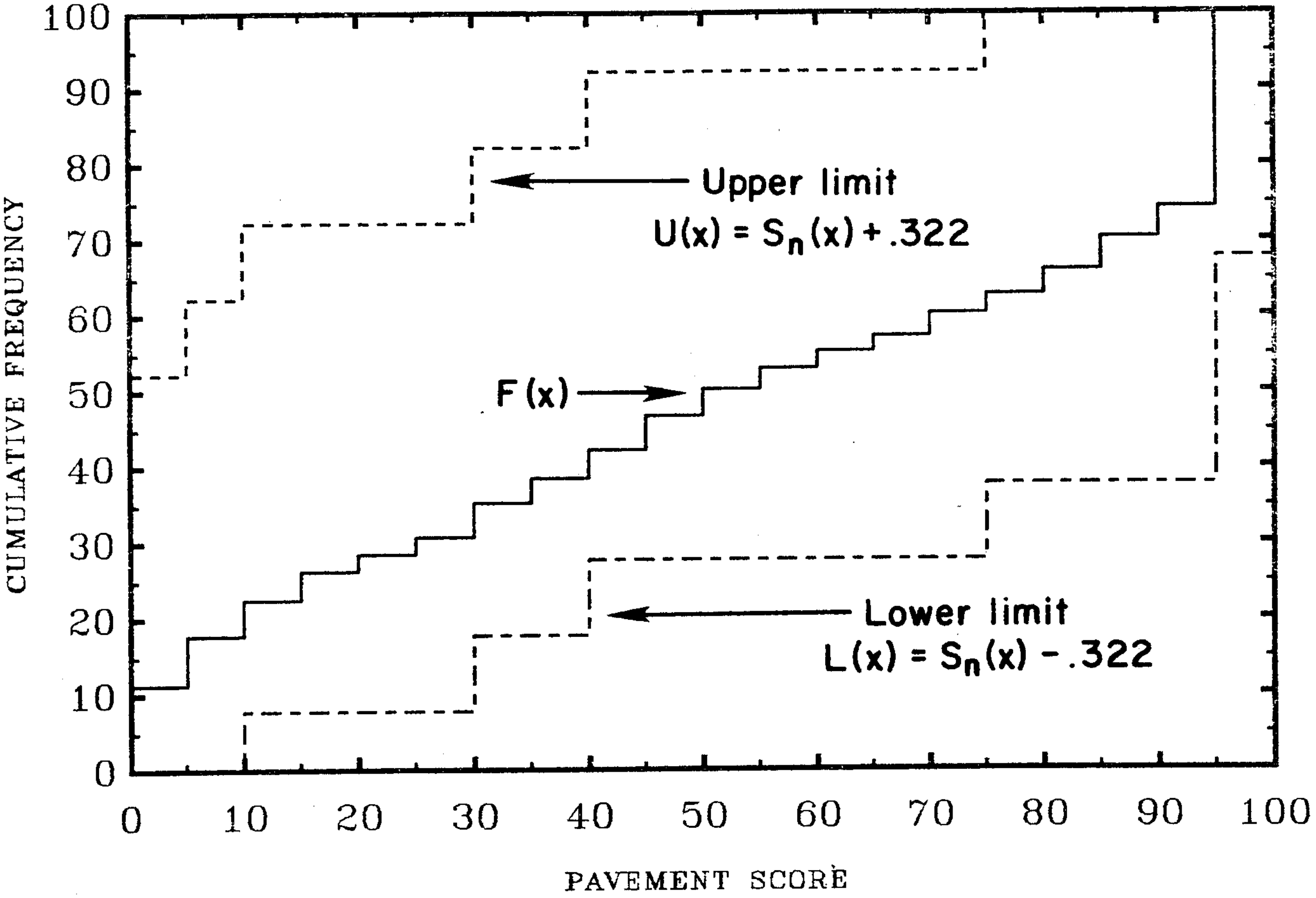


FIGURE B-7. Confidence band for cumulative distribution function of Pavement Scores F(X), District 11, System US

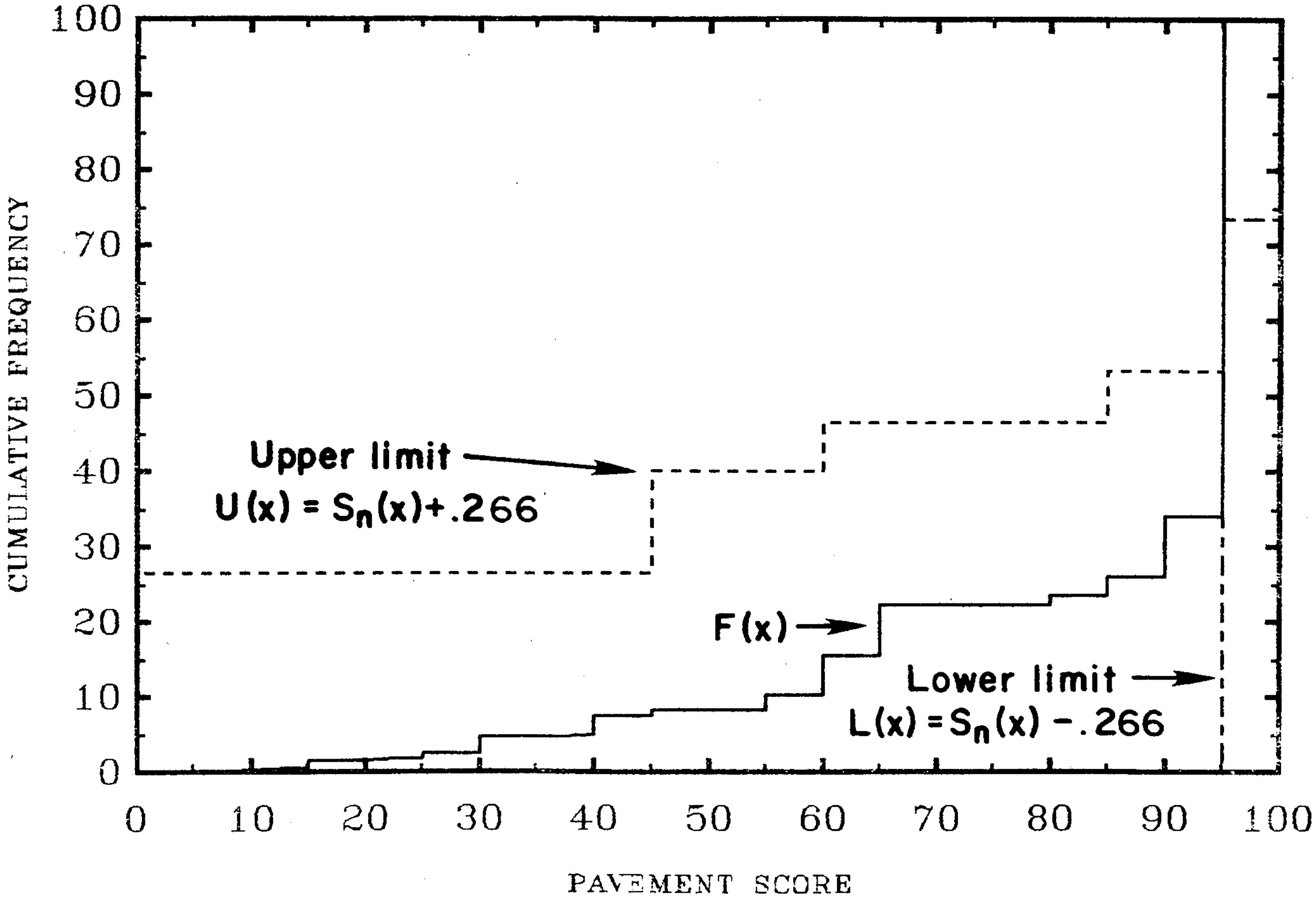


FIGURE B-8. Confidence band for cumulative distribution function of Pavement Scores F(X), District 15, System IH

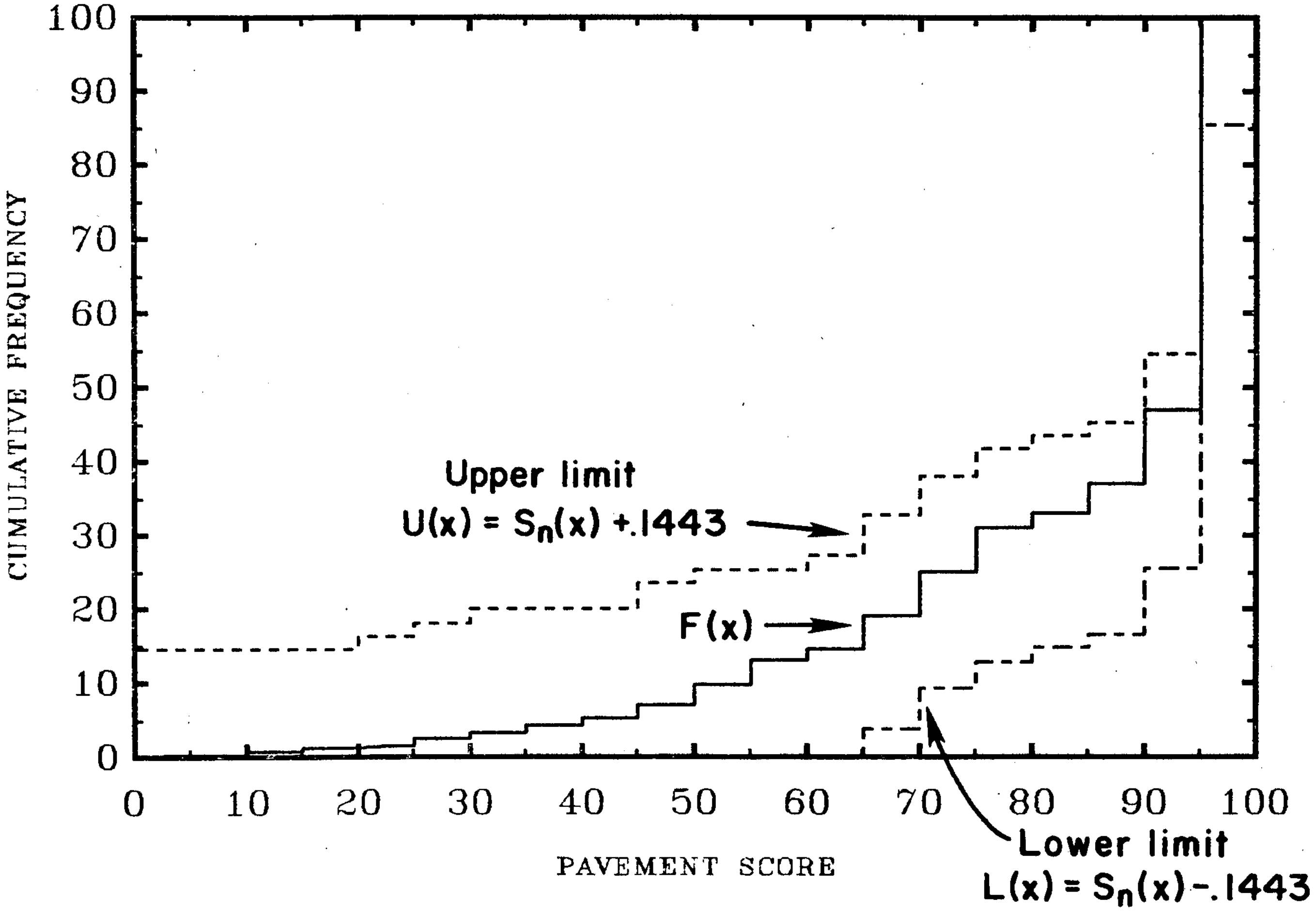


FIGURE B-9. Confidence band for cumulative distribution function of Pavement Scores F(X), District 15, System FM

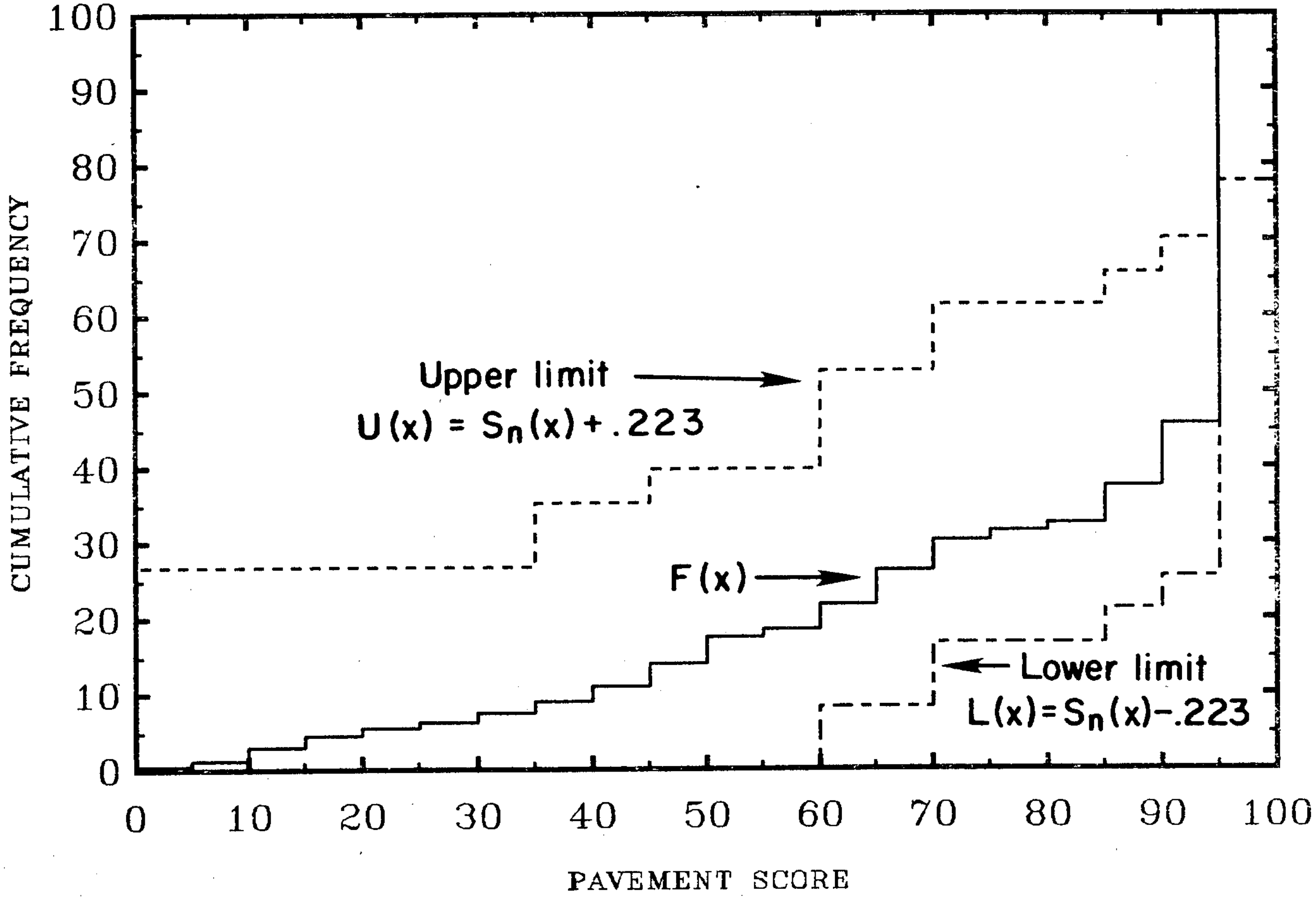


FIGURE B-10. Confidence band for cumulative distribution function of Pavement Scores F(X), District 15, System SH

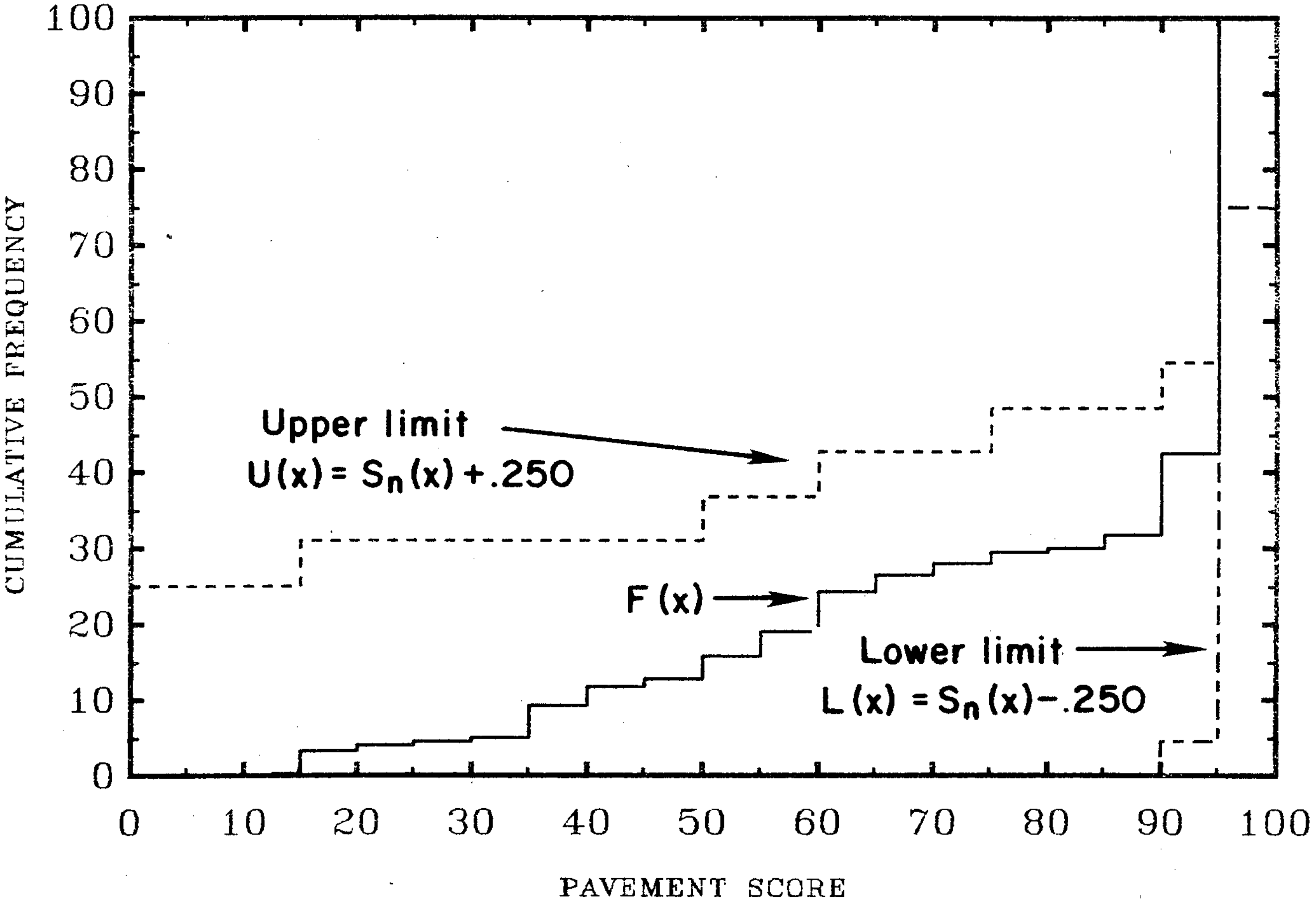


FIGURE B-11. Confidence band for cumulative distribution function of Pavement Scores F(X), District 15, System US

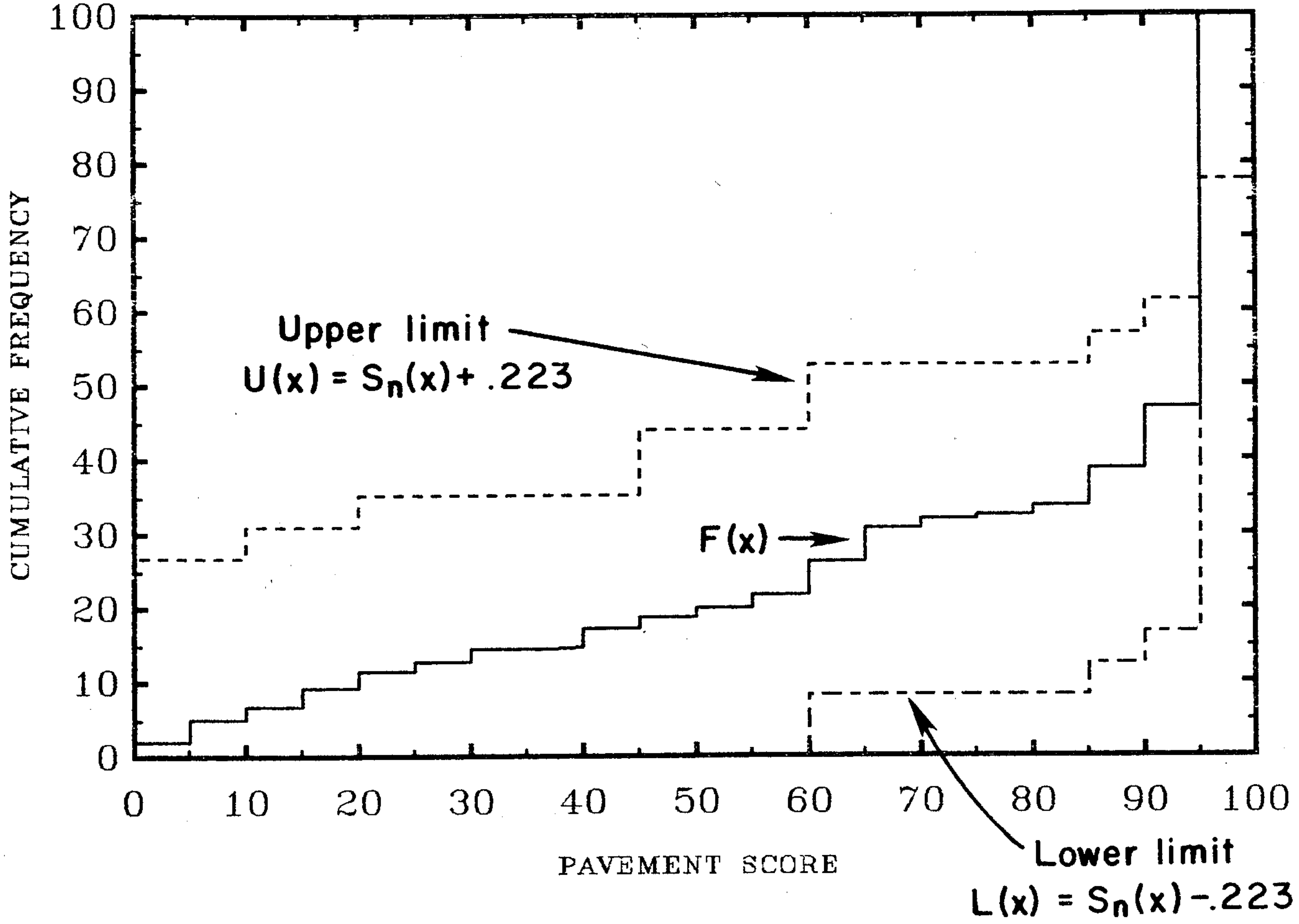


FIGURE B-11. Confidence band for cumulative distribution function of Pavement Scores F(X), Districts 8,15, System IH

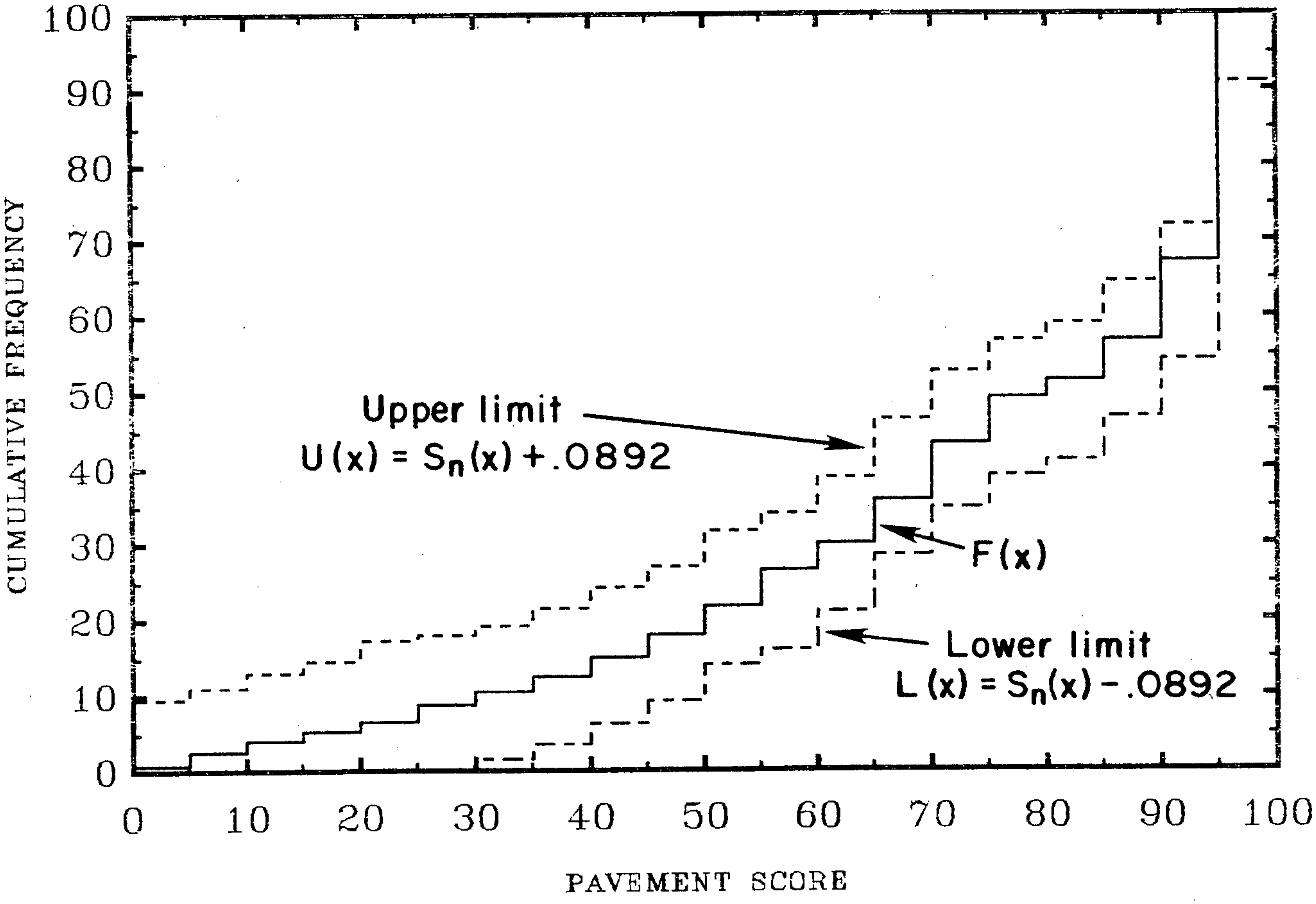


FIGURE B-13. Confidence band for cumulative distribution function of Pavement Scores F(X), Districts 8,11,15, System FM

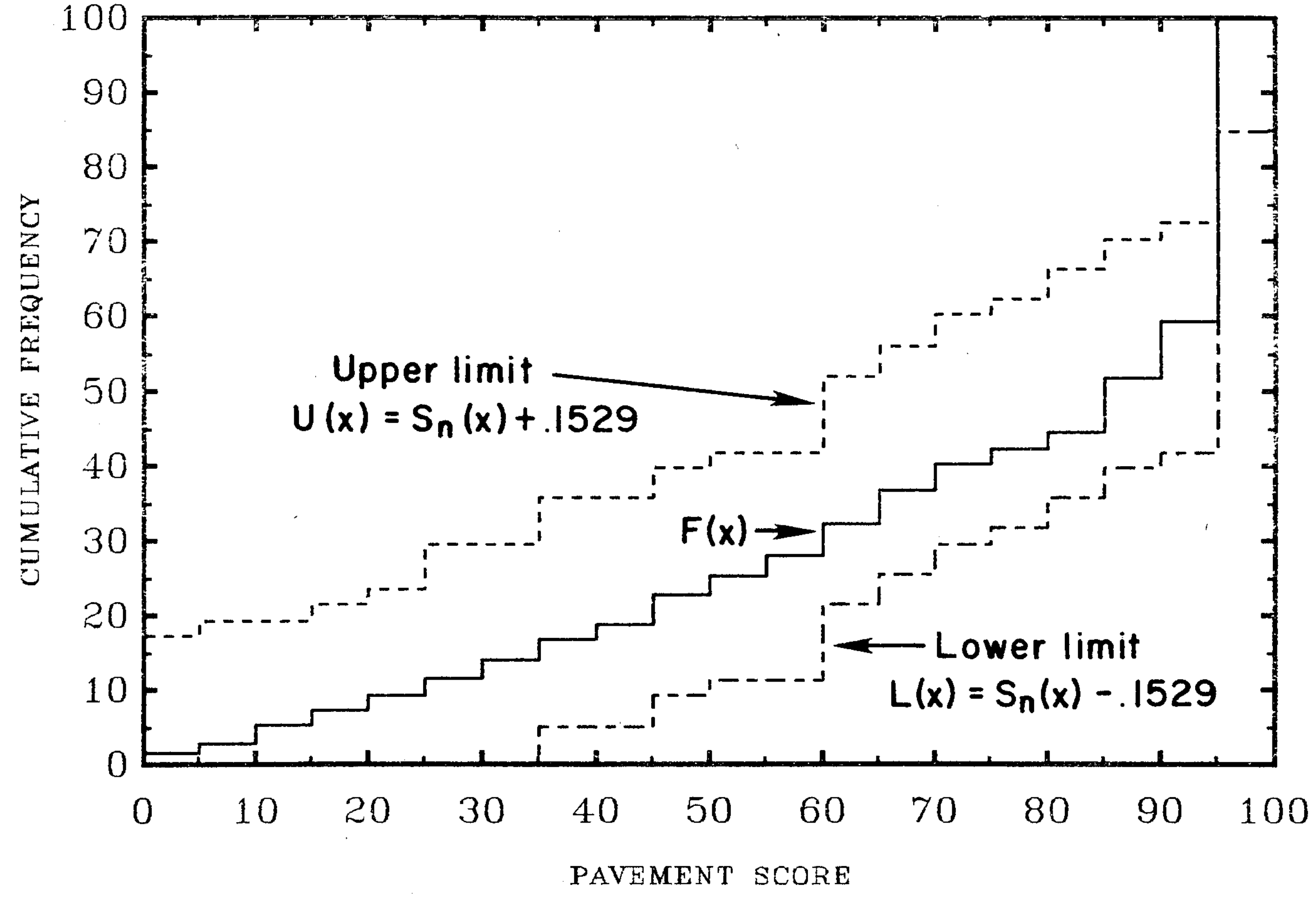


FIGURE B-14. Confidence band for cumulative distribution function of Pavement Scores F(X), Districts 8,11,15, System SH

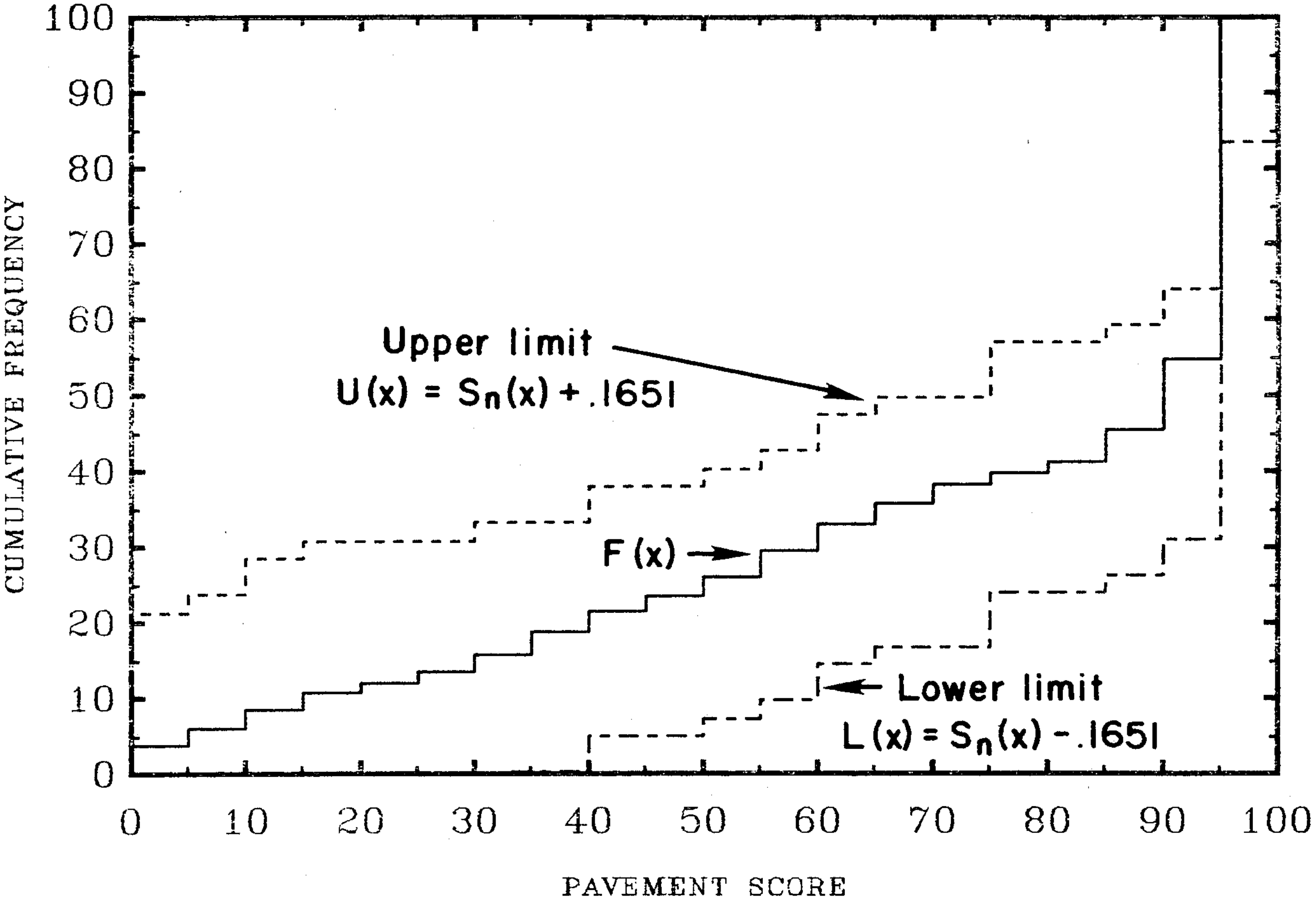


FIGURE B-15. Confidence band for cumulative distribution function of Pavement Scores F(X), Districts 8,11,15, System US

APPENDIX C

APPENDIX C

Kolmogorov-Smirnov Goodness of Fit Test

The Kolmogorov-Smirnov goodness of fit test can be used to determine if statistical evidence exists which leads to the belief that the 100 percent sample distribution comes from the beta distribution with parameters as specified in Table 6. In this procedure, if

$$\sup_{\text{all } x} \left| S_n(x) - F^*(x) \right| > K_{(1-\alpha)} \qquad \text{(C-1)}$$

there does exist statistical evidence that the true but unknown distribution function from which the elements of $S_n(x)$ came, $F(x)$, is not given by the function $F^*(x)$. In Equation (C-1)

$S_n(x)$ = 100 percent sample cumulative distribution,

$F^*(x)$ = cumulative distribution with parameters set out in Table 6,

$K_{(1-\alpha)}$ = $(1-\alpha)$ percentile of the Kolmogorov-Smirnov test statistic,

α = level of significance or the probability of saying $F(x) \neq F^*(x)$ when $F^*(x)$ is actually the true distribution,

sup = least upper bound over all x.

The difference between $s(x)$ and $F^*(x)$ is found at each Pavement score division (x=.05, .10, .15,..., 1.0). $F^*(x)$ is computed using the International Mathematical and Statistical Libraries, Inc. (IMSL) subroutine MDBETA. The maximum difference for each data group is set out in Table C-1. Also shown in Table C-1 is the 99th percentile (α=.01) of the Kolmogorov-Smirnov test statistic ($K_{.99}$).

According to Daniel (18) when the parameters of F*(X) must be estimated from the sample data, the Kolmogorov-Smirnov test no longer applies in the strict sense; but Massey (19) states that when parameters are estimated from sample data, the test is conservative in the sense that the actual α level will be smaller than the specified value.

As can be seen in Table C-1, in every case except FM's, the maximum difference is less than $K_{.99}$. Hence, there exists statistical evidence that the 100 percent sample did come from a beta distribution with parameters specified in Table 6. Because the maximum difference in FM's is only slightly larger than the corresponding critical value, the beta distribution will be assumed to be the true distribution from which the sample of FM's came.

TABLE C-1. Maximum difference between $S_{n(X)}$ and $F^*(X)$ and $K_{(.99)}$.

Data Group	Maximum Difference of $S_n(X)$ & $F^*(X)$	$K_{.99}$
IH08	0.0685	0.13
FM08	0.0729	0.05
SH08	0.0434	0.09
US08	0.0279	0.09
FM11	0.0591	0.05
SH11	0.0502	0.09
US11	0.0410	0.105
IH15	0.0775	0.09
FM15	0.0561	0.049
SH15	0.0504	0.07
US15	0.0636	0.09

APPENDIX D

APPENDIX D

Estimators of the Mean Cost Per Square Yard

As stated in the body of this report the length of segments range from 0.3 to 3 miles with an average length of two miles. Widths also vary substantially. For this reason, the following two estimators of mean cost per square yard were first considered.

$$\text{Mean of Ratio:} \quad \hat{R} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i} \qquad \text{(D-1)}$$

$$\text{Ratio of Means:} \quad \hat{r} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} \qquad \text{(D-2)}$$

where

Y_i = total cost of maintenance or rehabilitation for observation i,

X_i = total square yards in observation i, and

n = number of observations sampled.

The second estimator given above is usually referrred to as a "ratio estimator."

Scheaffer et al. (4) recommended use of the ratio estimator when the correlation coefficient of the response Y (total cost of an observation) and an auxiliary variable X (total area of an observation) is greater than 0.5. This correlation coefficient was calculated for the three 100 percent sampled districts divided by roadway systems. When only those observations with non-zero costs associated with them are considered, the following correlation coefficients result:

System	Correlation Coefficient
IH	0.436
FM	0.624
SH	0.547
US	0.586

In every case except Interstate Highways, the correlation coefficient of observation cost and observation size was greater than 0.50.

Schaeffer et al. (4) also state that use of the ratio estimator is most effective when the relationship between the response Y and an auxiliary variable X is linear through the origin. When the observations with non-zero costs associated with them were plotted with total cost of the observation on the ordinate axis and total square yards of the observation on the abscissa axis, this relationship does actually exist, as shown in Figures D-1, D-2, D-3, and D-4. Interstate Highways, however, do not exhibit this relationship as well as the other systems. The distinct lines occurring in each plot represent the various funding strategies.

Because of the above results, post stratification is considered for both the mean of ratios and the ratio of means. After the sample is taken, the observations were divided into two strata - those with

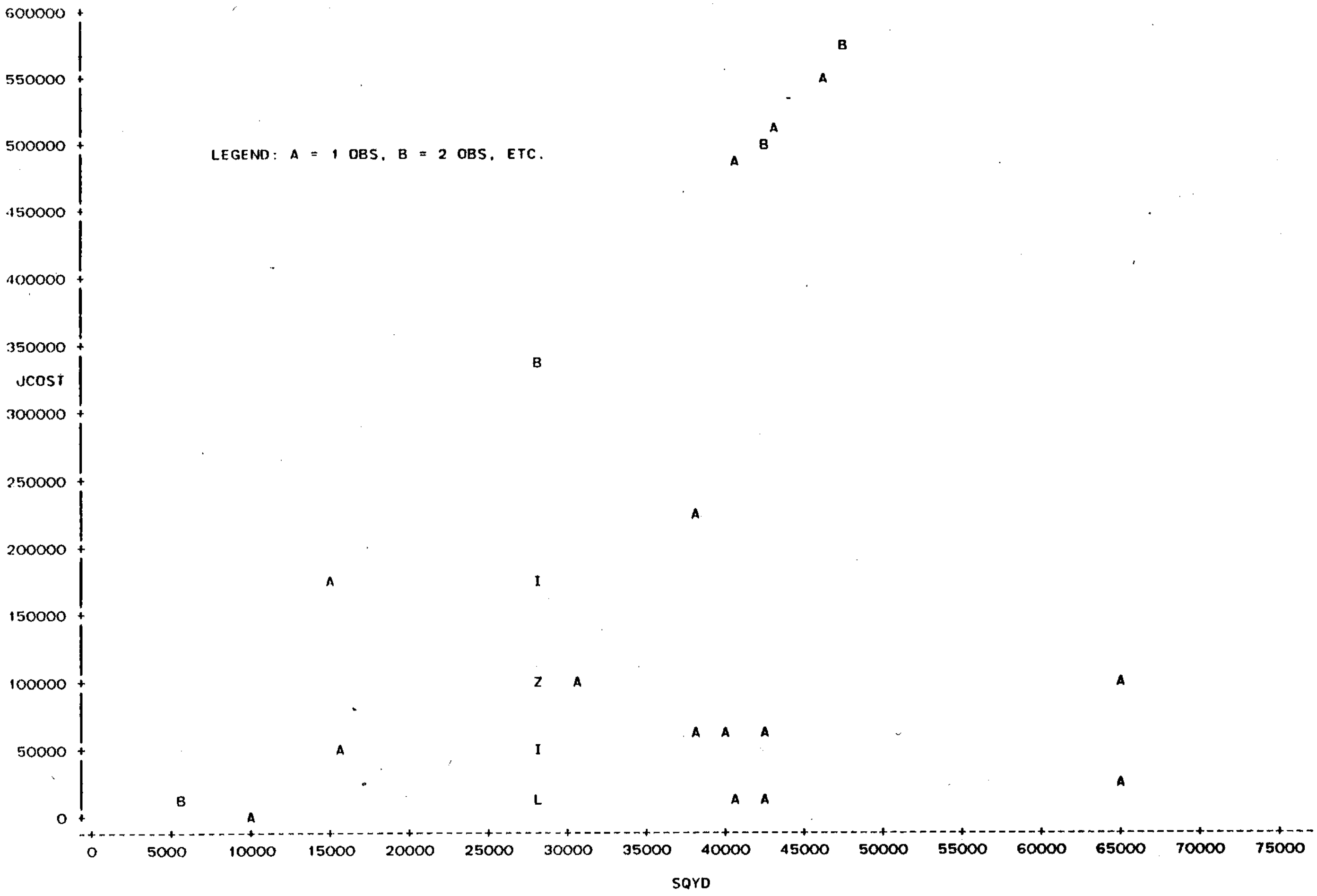


FIGURE D-1. Plot of total cost for observation i versus total square yards in observation i, Districts 8,15, System IH

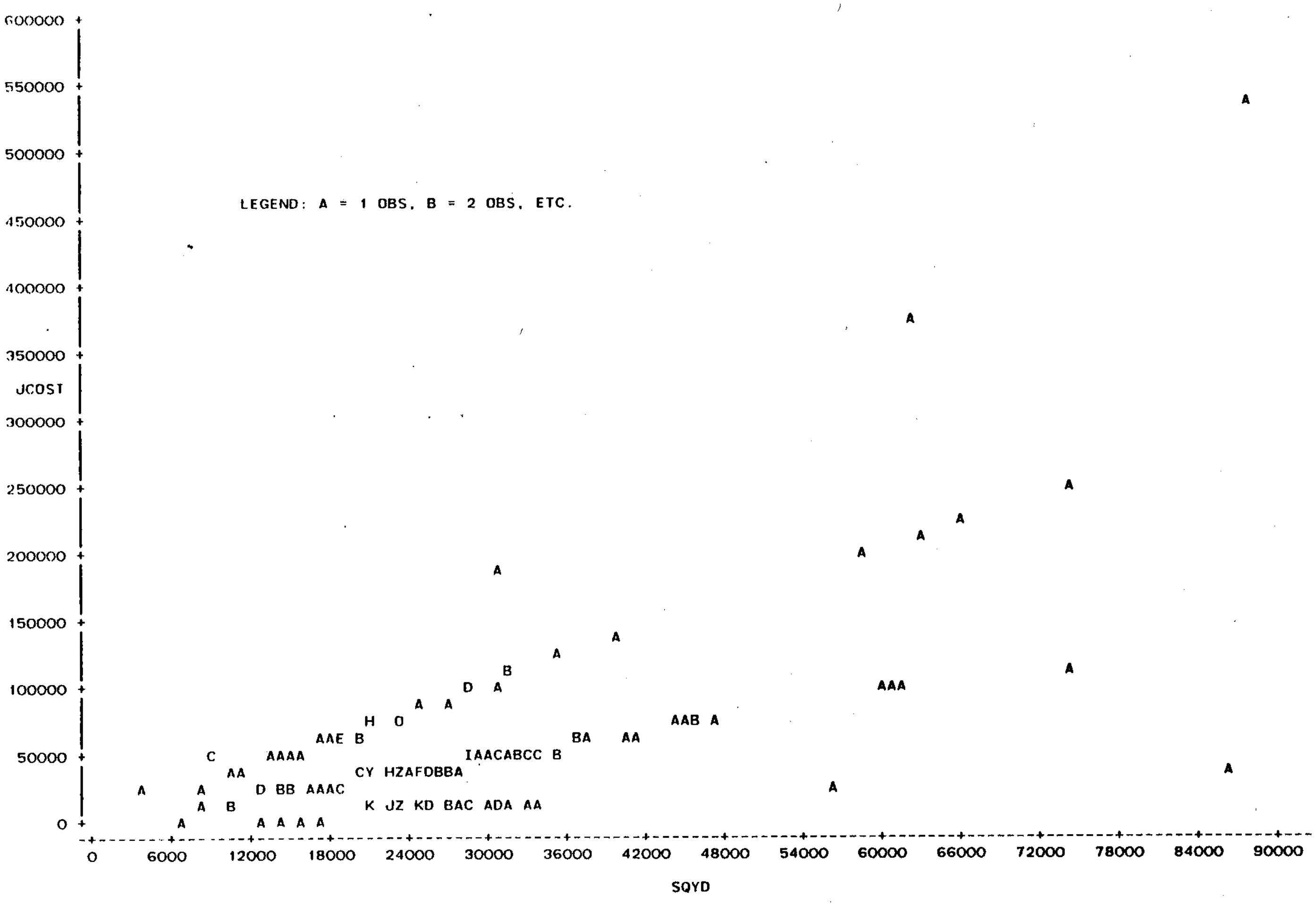


FIGURE D-2. Plot of total cost for observation i versus total square yards in observation i, Districts 8,11,15, System FM

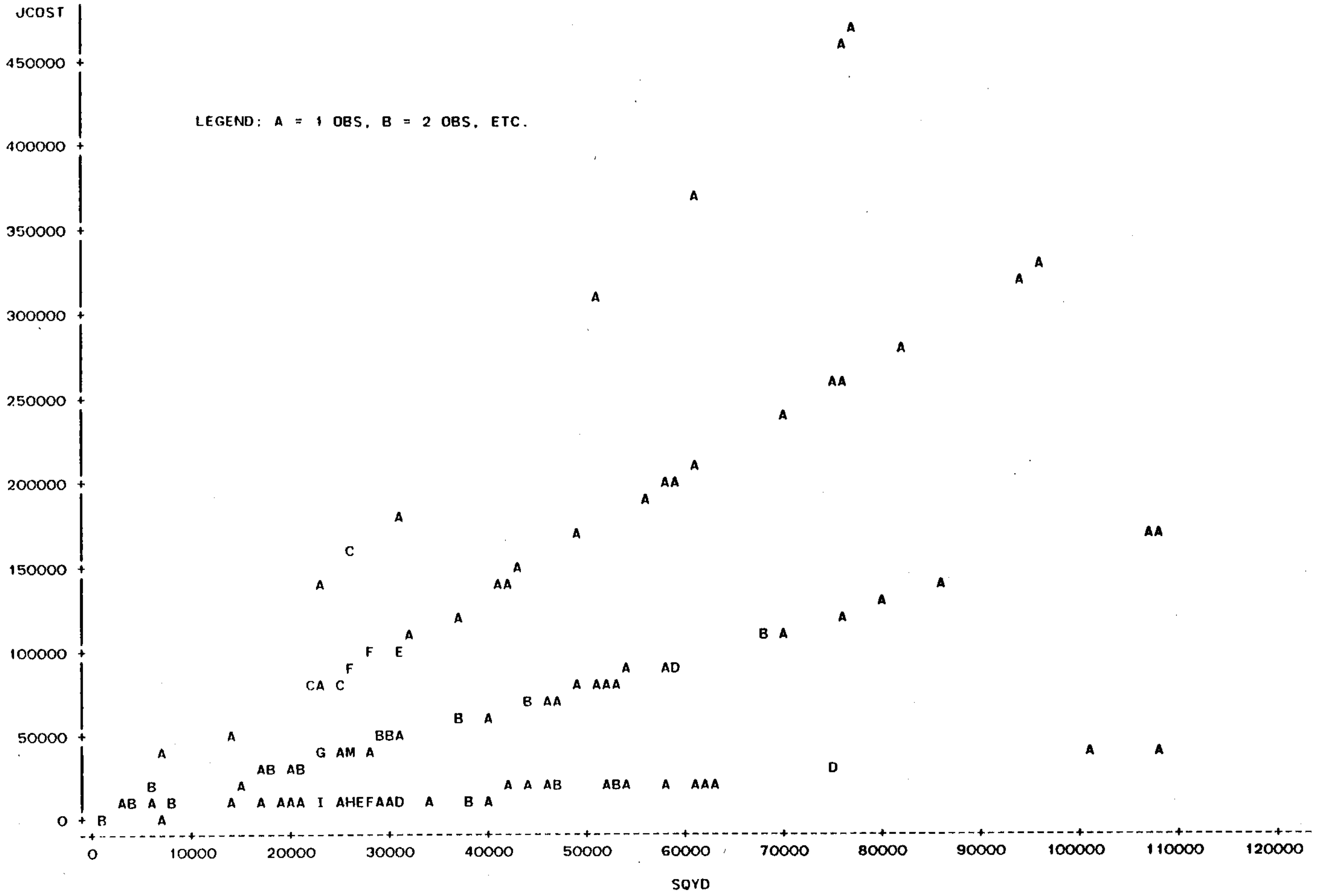


FIGURE D-3. Plot of total cost for observation i versus total square yards in observation i, Districts 8,11,15, System SH

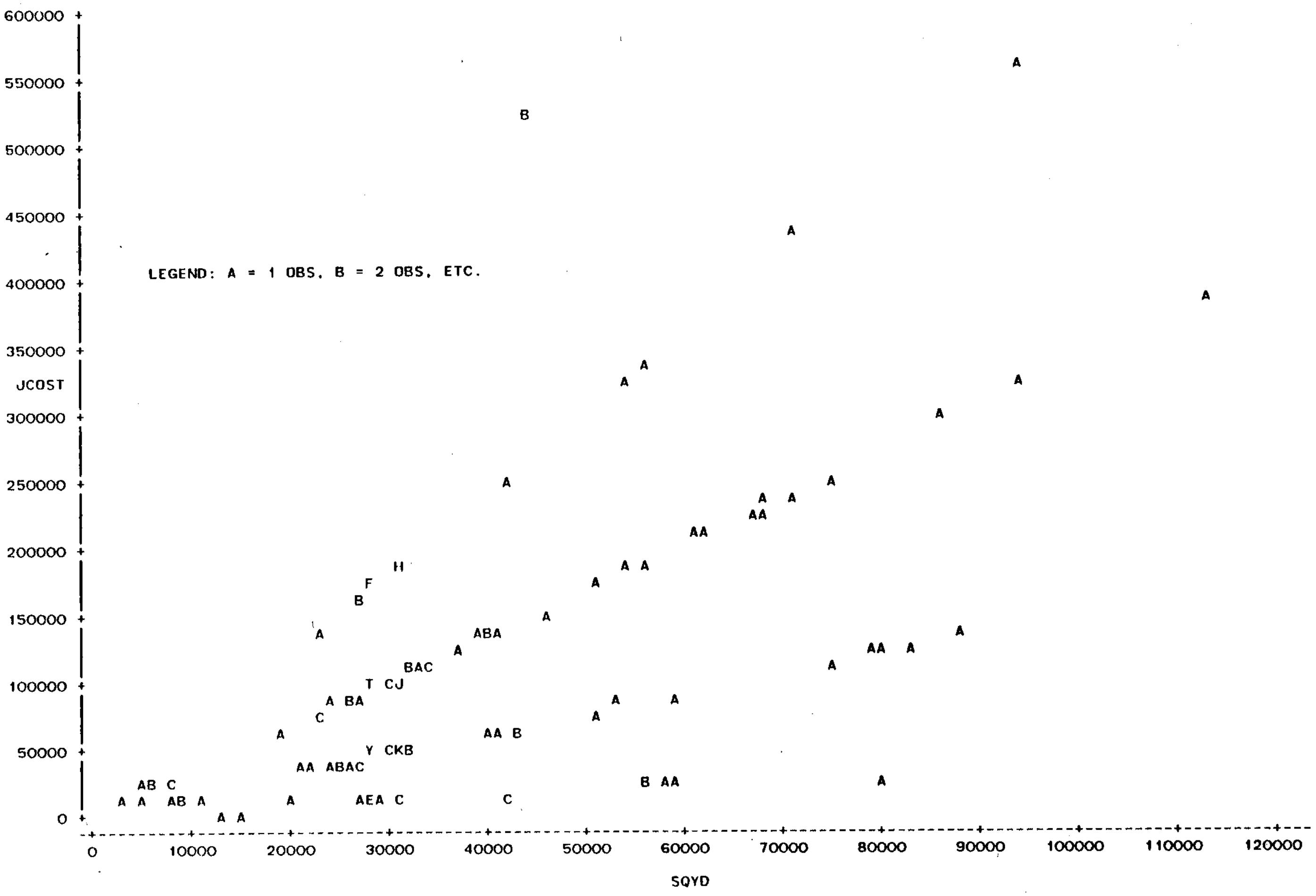


FIGURE D-4. Plot of total cost for observation i versus total square yards in observation i, Districts 8,11,15, System US non-zero costs and those with zero costs. The formulas for calculating the two statistics (ratio of means and mean of ratios) for post-stratification are the same of those for ordinary stratification. There are two different methods for constructing estimators of a ratio in stratified sampling. One is to estimate the ratio of μ_y (mean of Y) to μ_x (mean of X) within each stratum and then form a weighted average of these separate estimates as a single estimate of the population. The result of this is called a separate ratio estimator. The other method involves first estimating μ_y by $\bar{Y}_{st}$ (a weighted average of the $\bar{Y}_k$'s, where $\bar{Y}_k$ is the estimated mean of the stratum) and similarly estimating μ_x by $\bar{X}_{st}$. Then $\bar{Y}_{st}/\bar{X}_{st}$ can be used as an estimator of μ_y/μ_x. This is called a combined ratio estimator.

The equations to calculate these estimators are as follows:

(a) For

Ratio of Means:
(Separate Estimator)

$$\hat{r}_s = \frac{N_c}{N} \frac{\sum_{i=1}^{n_c} Y_{ci}}{\sum_{i=1}^{n_c} X_{ci}} + \frac{N_{nc}}{N} \frac{\sum_{i=1}^{n_{nc}} Y_{nci}}{\sum_{i=1}^{n_{nc}} X_{nci}} \qquad \text{(D-3)}$$

(b) For

Ratio of Means:
(Combined Estimator)

$$\hat{r}_c = \frac{N_c \bar{Y}_c + N_{nc} \bar{Y}_{nc}}{N_c \bar{X}_c + N_{nc} \bar{X}_{nc}} \qquad \text{(D-4)}$$

where

Y_{ci} = total non-zero cost of observation i,

X_{ci} = total square yards in i-th observation with non-zero cost,

Y_{nci} = total cost of observation i with zero costs,

X_{nci} = total square yards in i-th observation with zero costs,

$\bar{Y}_c$ = mean total cost per observation for all observaions with non-zero costs,

$\bar{X}_c$ = mean total square yards per observation for all observations with non-zero costs,

$\bar{Y}_{nc}$ = mean total cost per observation for all observations with zero costs,

$\bar{X}_{nc}$ = mean total square yards per observation for all observations with zero costs,

N = total number of observations in population,

N_c = total number of observations with non-zero costs in population, and

N_{nc} = total number of observations with zero costs in population.

Because $\sum_{i=1}^{n_{nc}} Y_{nci} = 0$ and $\bar{Y}_{nc} = 0$ ($Y_{nci}=0$ for all i)
Equation D-3 and Equation D-4 respectively become:

$$\hat{r}_s = \frac{N_c}{N} \frac{\sum_{i=1}^{n_c} Y_{ci}}{\sum_{i=1}^{n_c} X_{ci}} \qquad \text{(D-5)}$$

$$\hat{r}_c = \frac{N_c \bar{Y}_c}{N_c \bar{X}_c + N_{nc} \bar{X}_{nc}} \tag{D-6}$$

The equation to calculate the mean of ratios estimator with post stratification is as follows:

$$\hat{R}_{st} = \frac{N_c}{N} \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{Y_{ci}}{X_{ci}} + \frac{N_{nc}}{N} \frac{1}{n_{nc}} \sum_{i=1}^{n_{nc}} \frac{Y_{nci}}{X_{nci}} \tag{D-7}$$

Again, because each $Y_{nci} = 0$, Equation D-7 becomes:

$$\hat{R}_{st} = \frac{N_c}{N} \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{Y_{ci}}{X_{ci}} \tag{D-8}$$

Many times when using stratification, the number of observations within each strata are known. However, that is not the case in this situation. Even though N_c and N_{nc} are known for the three 100 percent sample districts, these are not known for the other 21 districts. The totals will change from year to year. The total number of observations with zero costs and non-zero costs will not be known for Districts 8, 11, and 15 next year. Hence, it is necessary to estimate N_c with $\frac{n_c}{n}N$ and N_{nc} with $\frac{n_{nc}}{n}N$. When these substitutions are made in Equations D-5, D-6, and D-8, Equations D-1 and D-8 become identical and Equations D-2 and D-6 become identical. Hence the three estimators which will be investigated are:

(1) Mean of Ratios ($\hat{R}$)

(2) Ratio of Means ($\hat{r}$)

(3) Ratio of Means - Post Strat., Sept. Est. ($\hat{r}_s$)

Table D-1. Estimated mean squared error of various estimators at various sample sizes for District 8.

Data Group	Sample Size %	MSE		
		Mean of Ratios	Ratio of Means	Ratio of Means (Post-Strat, Sep. Est.)
IH08	5	0.4220	0.4970	0.4621
	10	0.3038	0.3548	0.3378
	20	0.1258	0.1455	0.1395
	30	0.0615	0.0713	0.0689
FM08	5	0.0055	0.0083	0.0063
	10	0.0025	0.0039	0.0029
	20	0.0012	0.0019	0.0015
	30	0.0008	0.0013	0.0010
SH08	5	0.0538	0.0695	0.0478
	10	0.0279	0.0387	0.0293
	20	0.0130	0.0186	0.0149
	30	0.0081	0.0111	0.0085
US08	5	0.0430	0.0739	0.0424
	10	0.0242	0.0350	0.0228
	20	0.0137	0.0151	0.0130
	30	0.0080	0.0105	0.0077

Table D-2. Estimated mean squared error of various estimators at various sample sizes for District 11.

Data Group	Sample Size %	MSE		
		Mean of Ratios	Ratio of Means	Ratio of Means (Post-Strat, Sep. Est.)
FM11	5	0.0176	0.0157	0.0158
	10	0.0089	0.0087	0.0083
	20	0.0037	0.0034	0.0033
	30	0.0023	0.0019	0.0019
SH11	5	0.0847	0.1636	0.1078
	10	0.0526	0.0854	0.0611
	20	0.0263	0.0426	0.0326
	30	0.0160	0.0250	0.0183
US11	5	0.2328	0.3198	0.2620
	10	0.1089	0.1400	0.1225
	20	0.0564	0.0690	0.0609
	30	0.0302	0.0413	0.0351

Table D-3. Estimated mean squared error of various estimators at various sample sizes for District 15.

Data Group	Sample Size %	MSE		
		Mean of Ratios	Ratio of Means	Ratio of Means (Post-Strat, Sep. Est.)
IH15	5	0.2894	0.5974	0.3925
	10	0.0914	0.1664	0.1066
	20	0.0519	0.0724	0.0565
	30	0.0356	0.0465	0.0377
FM15	5	0.0040	0.0177	0.0058
	10	0.0025	0.0071	0.0031
	20	0.0017	0.0029	0.0019
	30	0.0012	0.0017	0.0011
SH15	5	0.0130	0.0278	0.0138
	10	0.0114	0.0148	0.0122
	20	0.0069	0.0058	0.0073
	30	0.0059	0.0039	0.0064
US15	5	0.1213	0.1769	0.1664
	10	0.0484	0.0663	0.0598
	20	0.0239	0.0295	0.0282
	30	0.0135	0.0170	0.0170

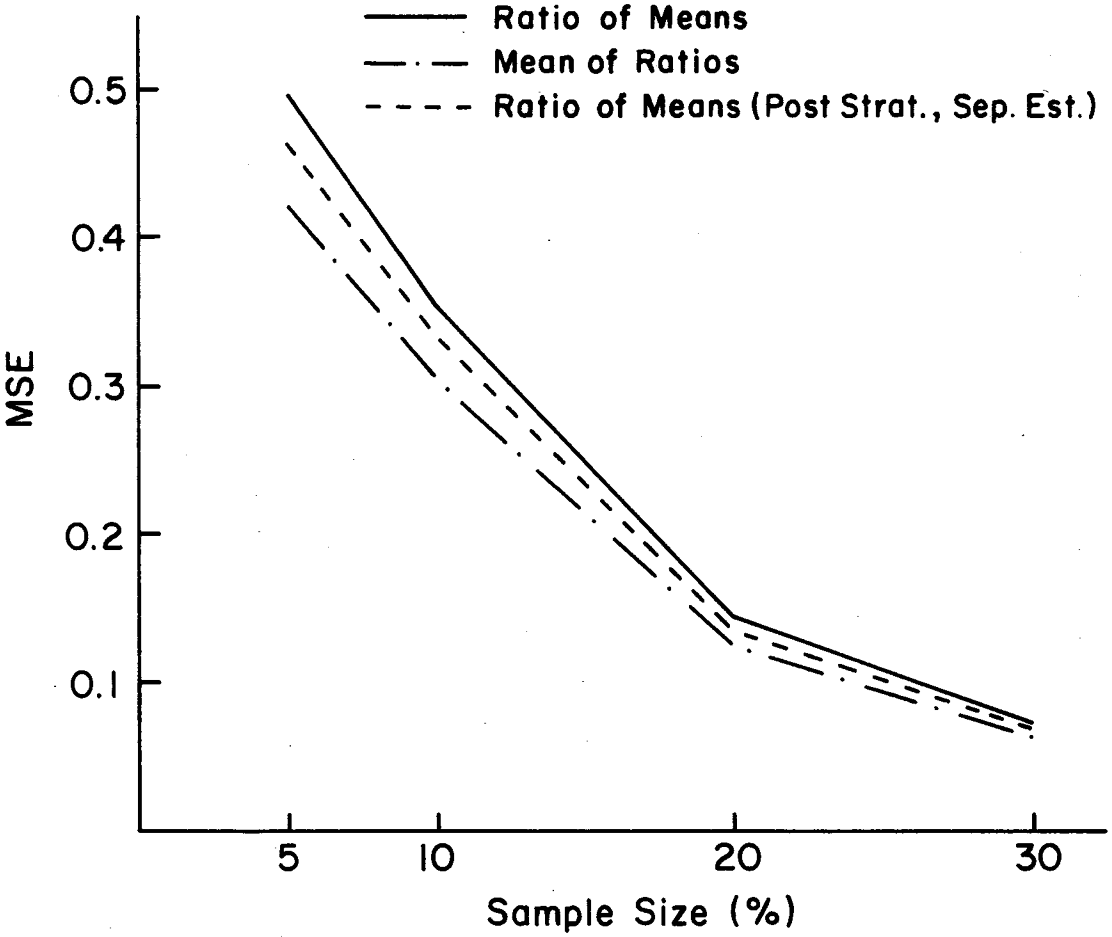


FIGURE D-5. Comparison of Estimations, District 8, System IH

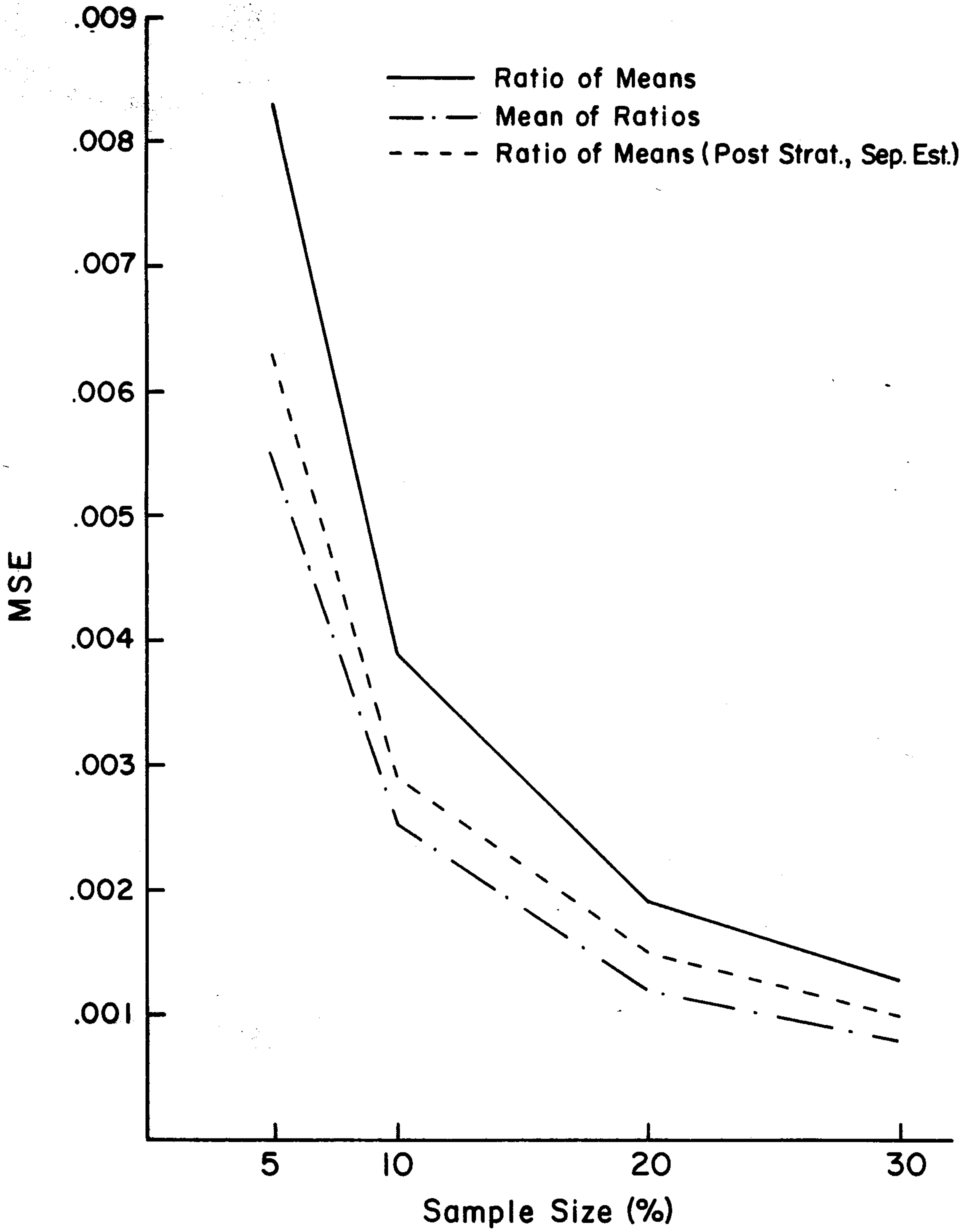


FIGURE D-6. Comparison of Estimations, District 8, System FM

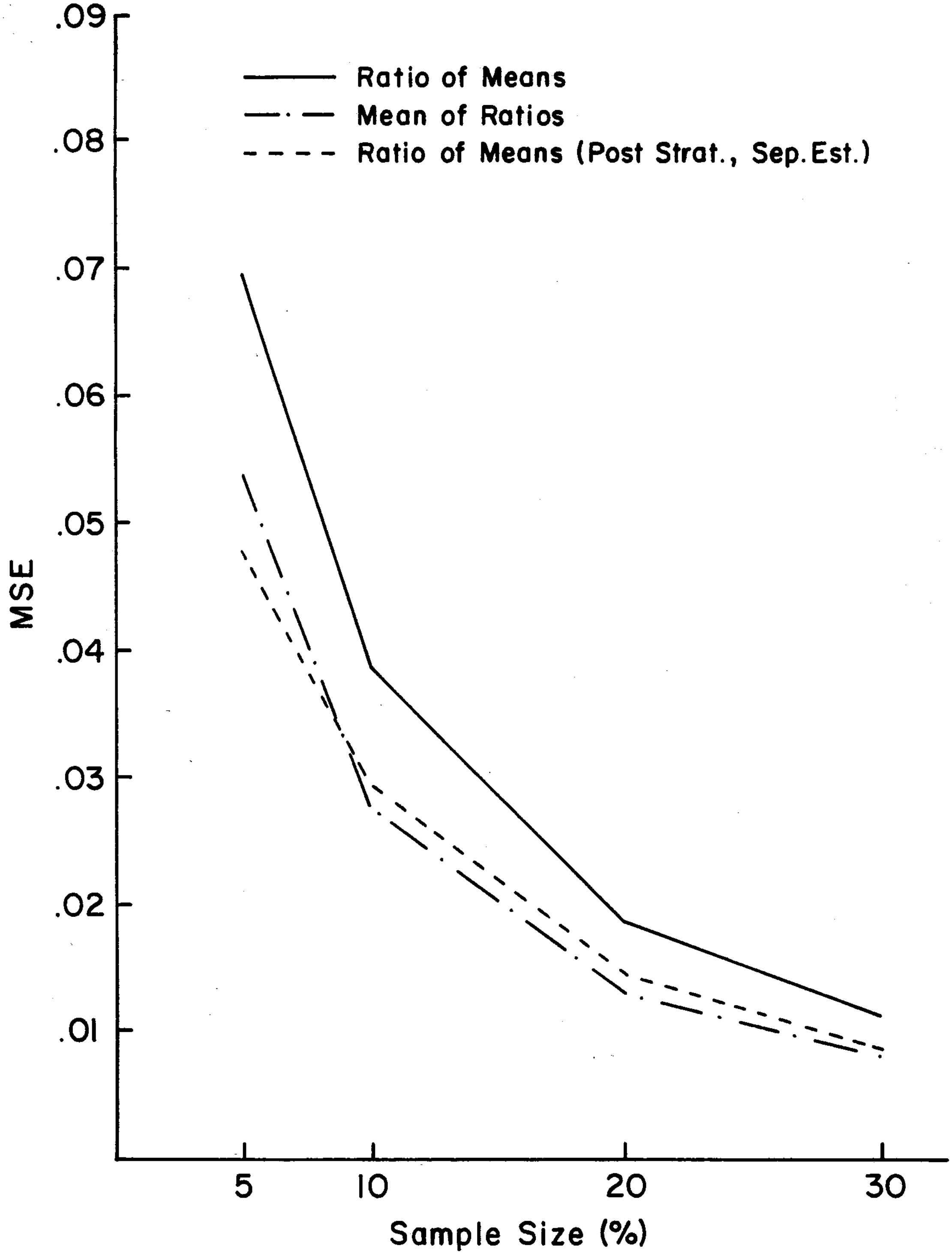


FIGURE D-7. Comparison of Estimations, District 8, System SH

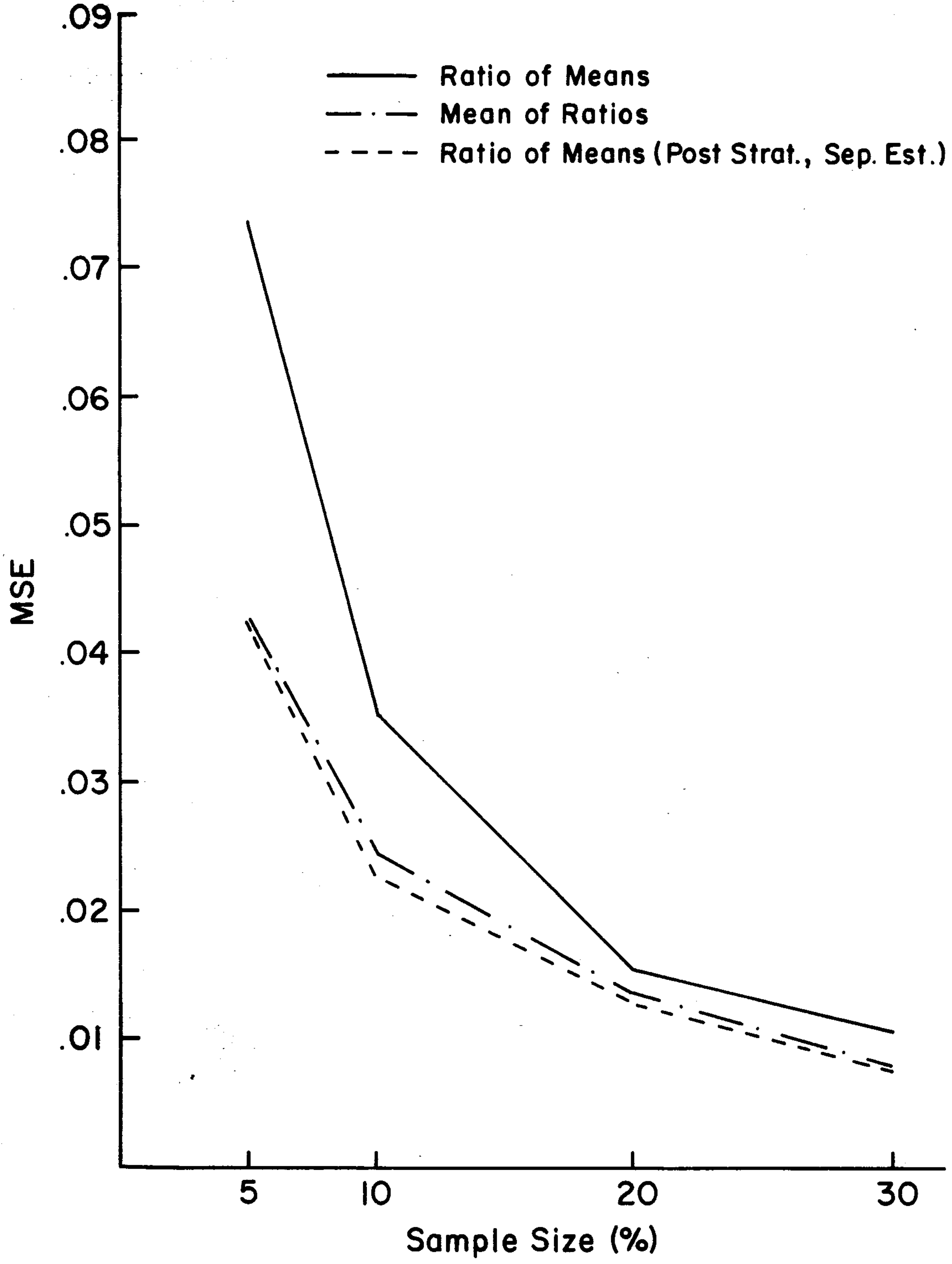


FIGURE D-8. Comparison of Estimations, District 8, System US

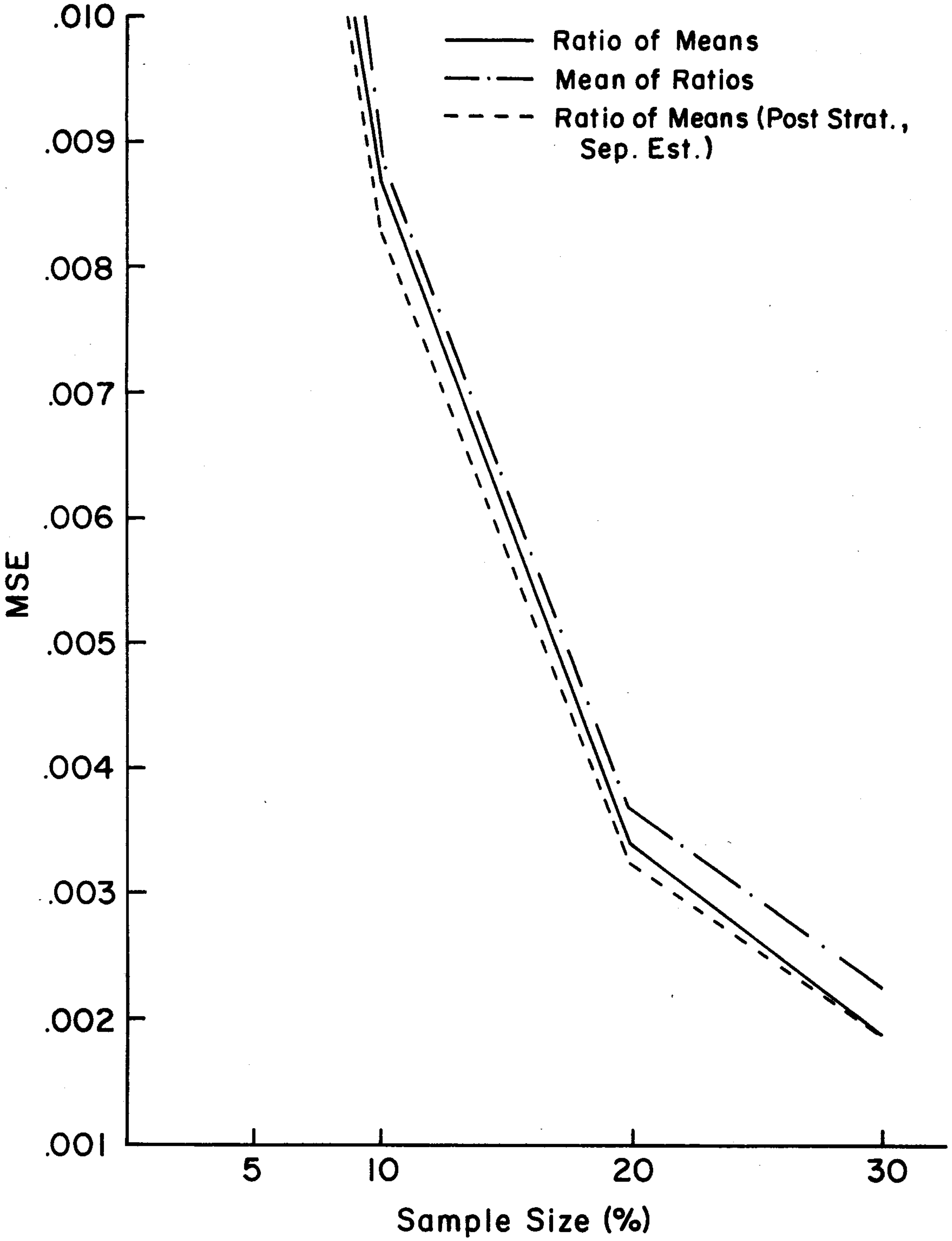


FIGURE D-9. Comparison of Estimations, District 11, System FM

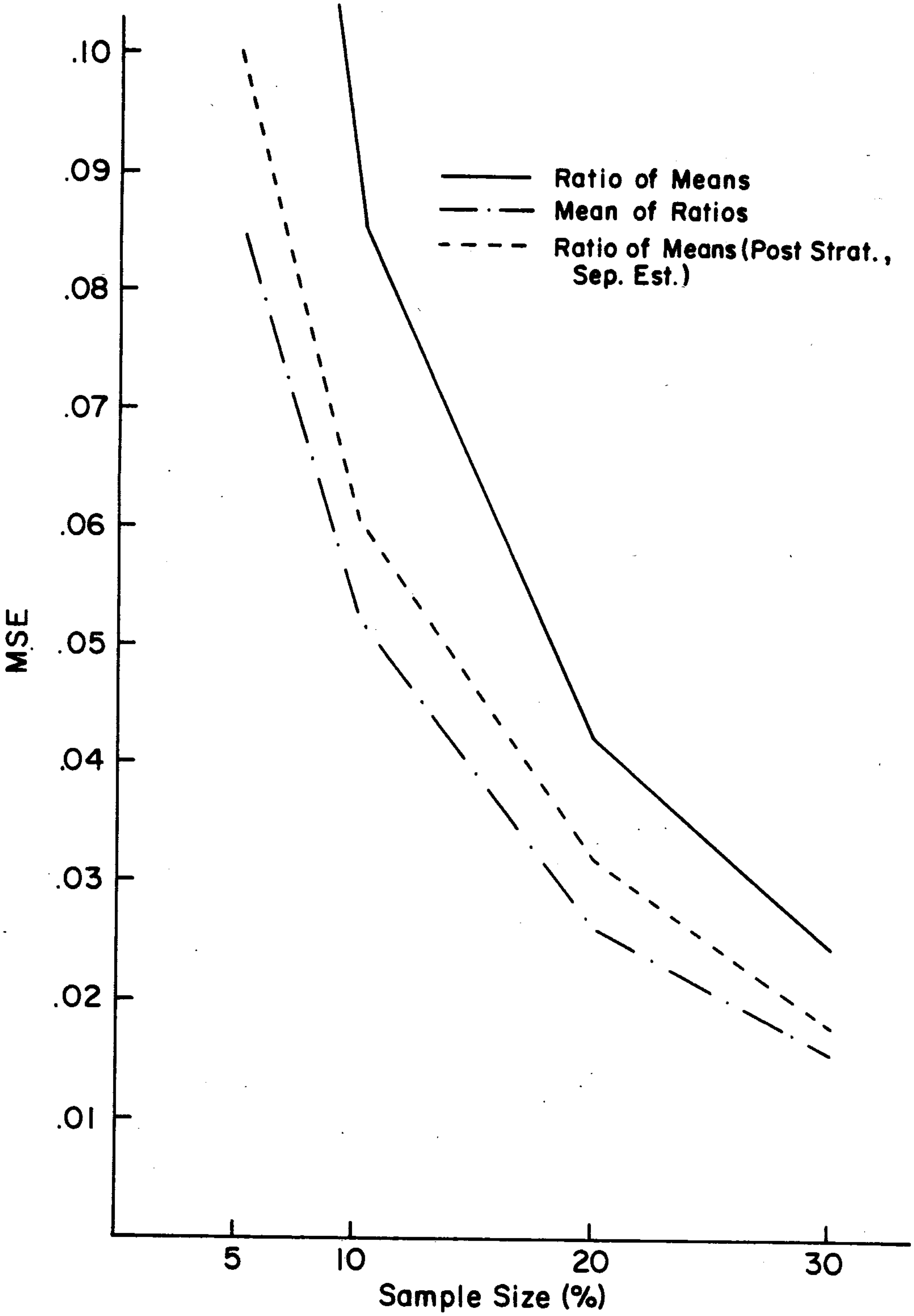


FIGURE D-10. Comparison of Estimations, District 11, System SH

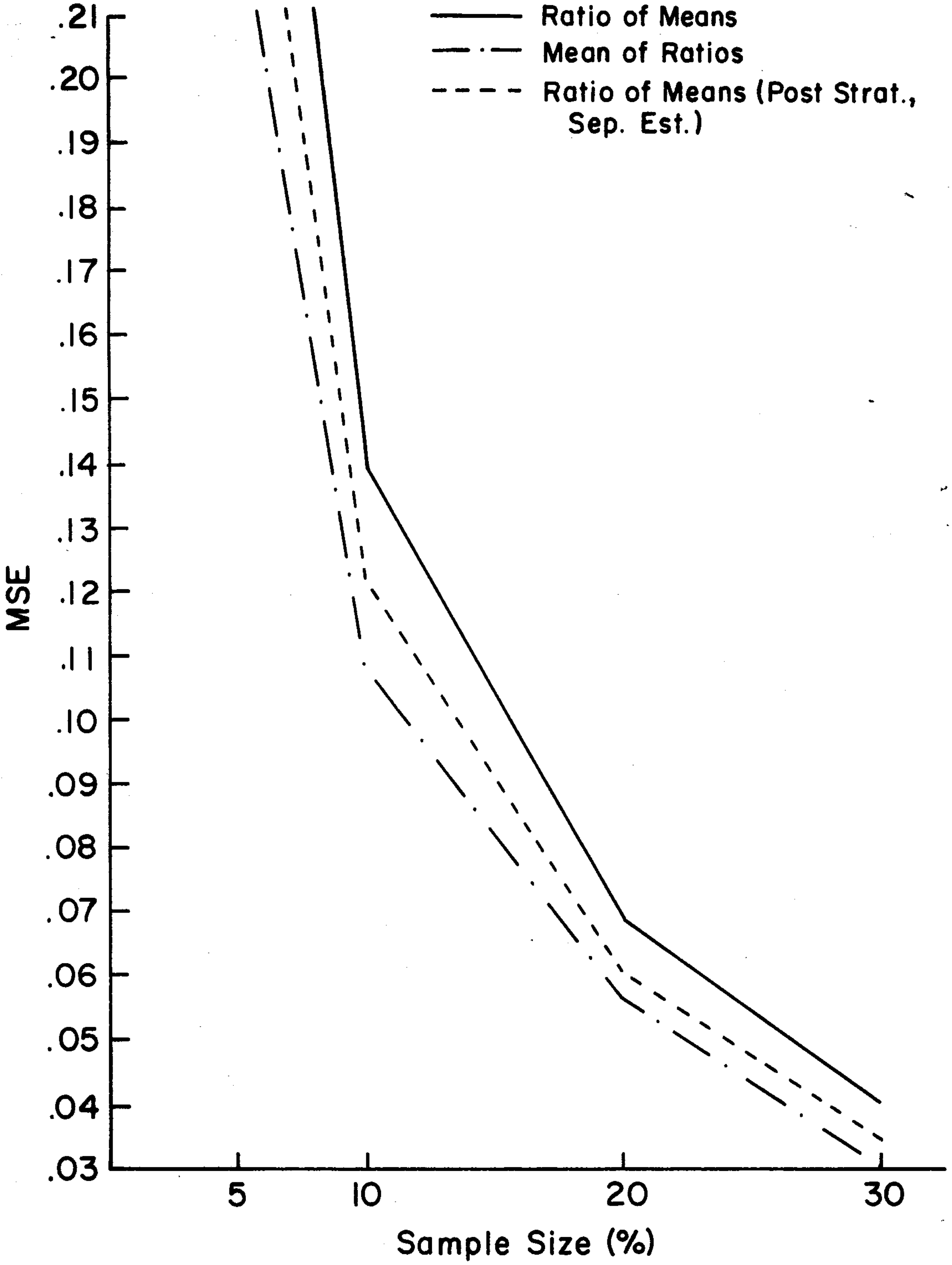


FIGURE D-11. Comparison of Estimations, District 11, System US

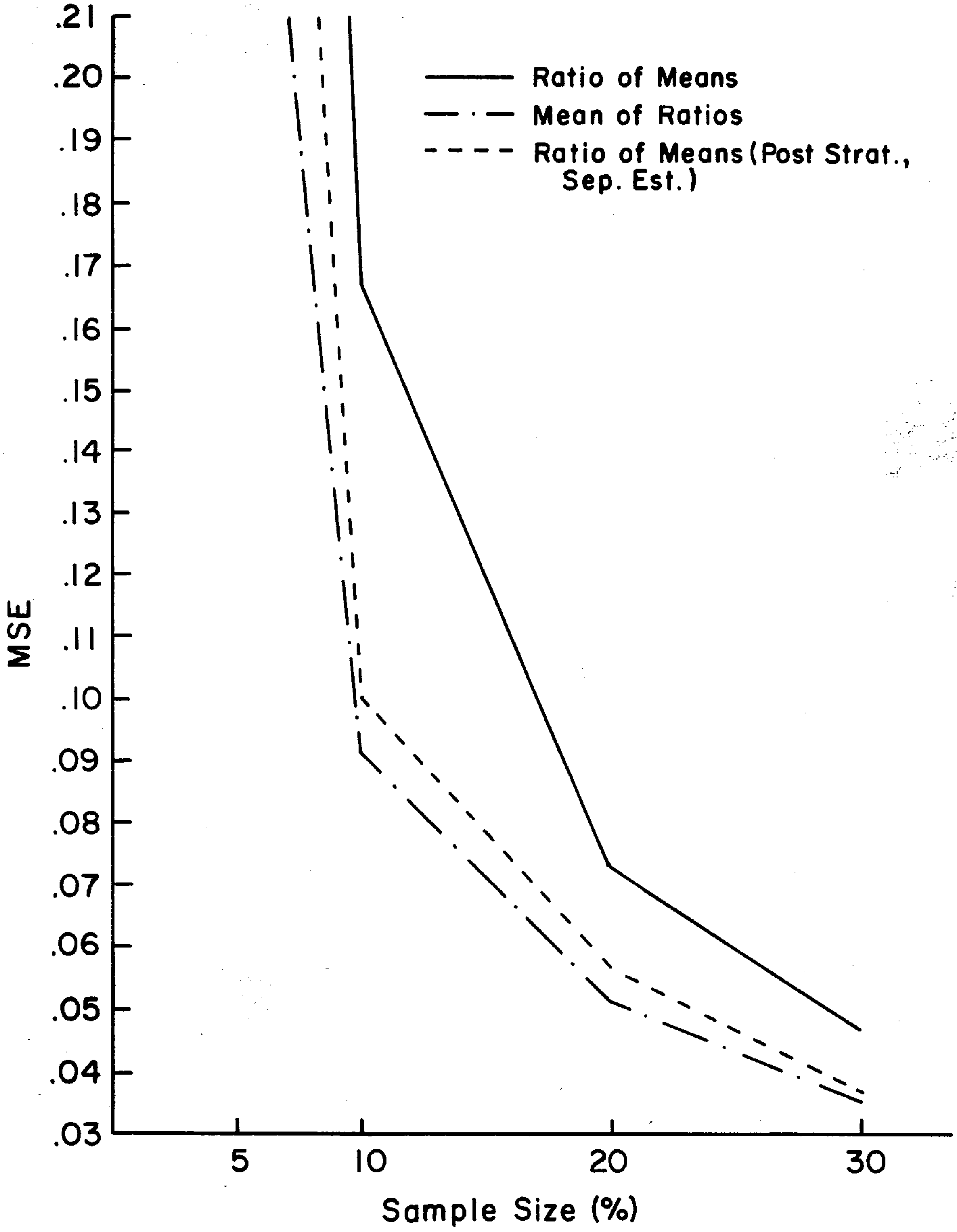


FIGURE D-12. Comparison of Estimations, District 15, System IH

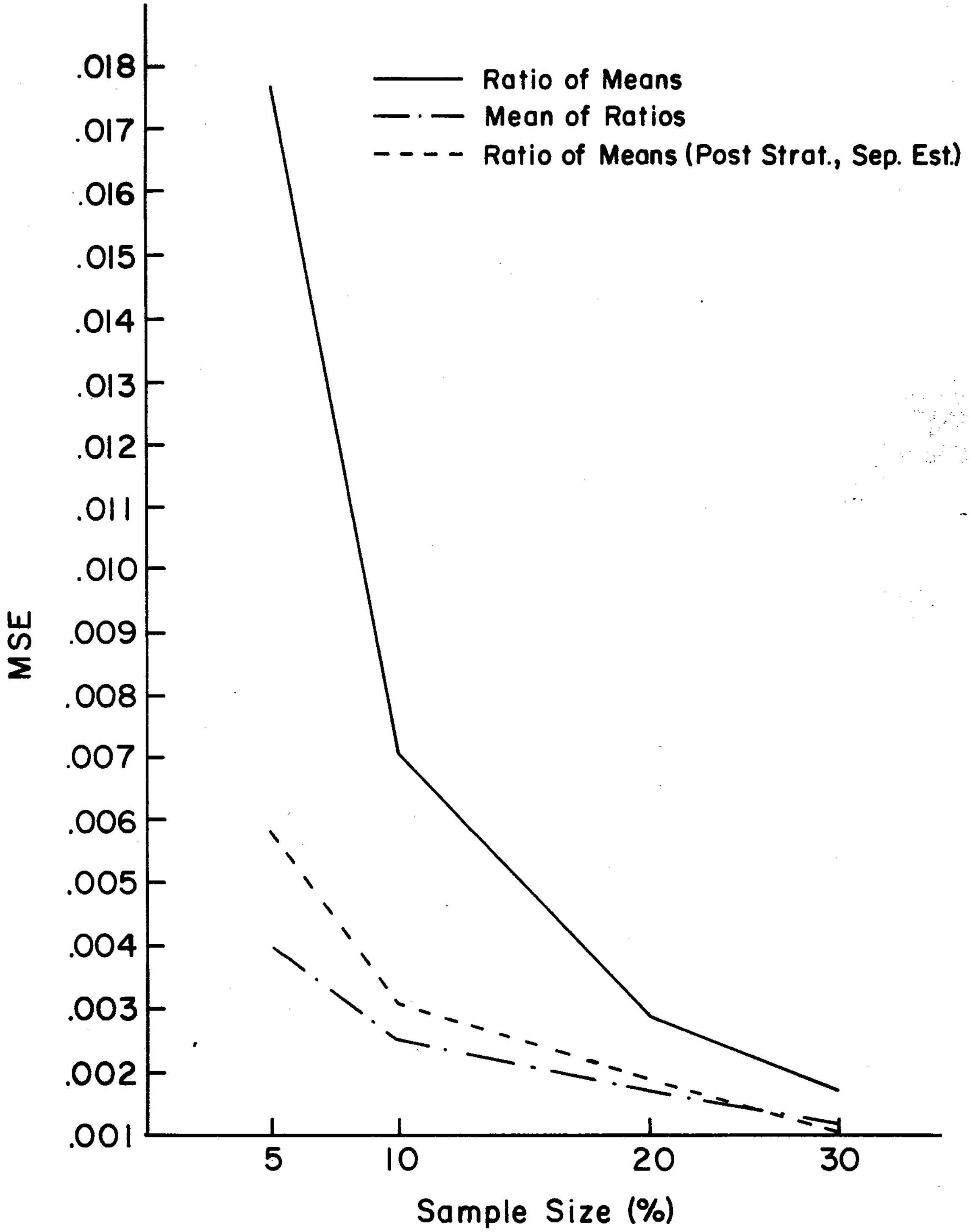


FIGURE D-13. Comparison of Estimations, District 15, System FM

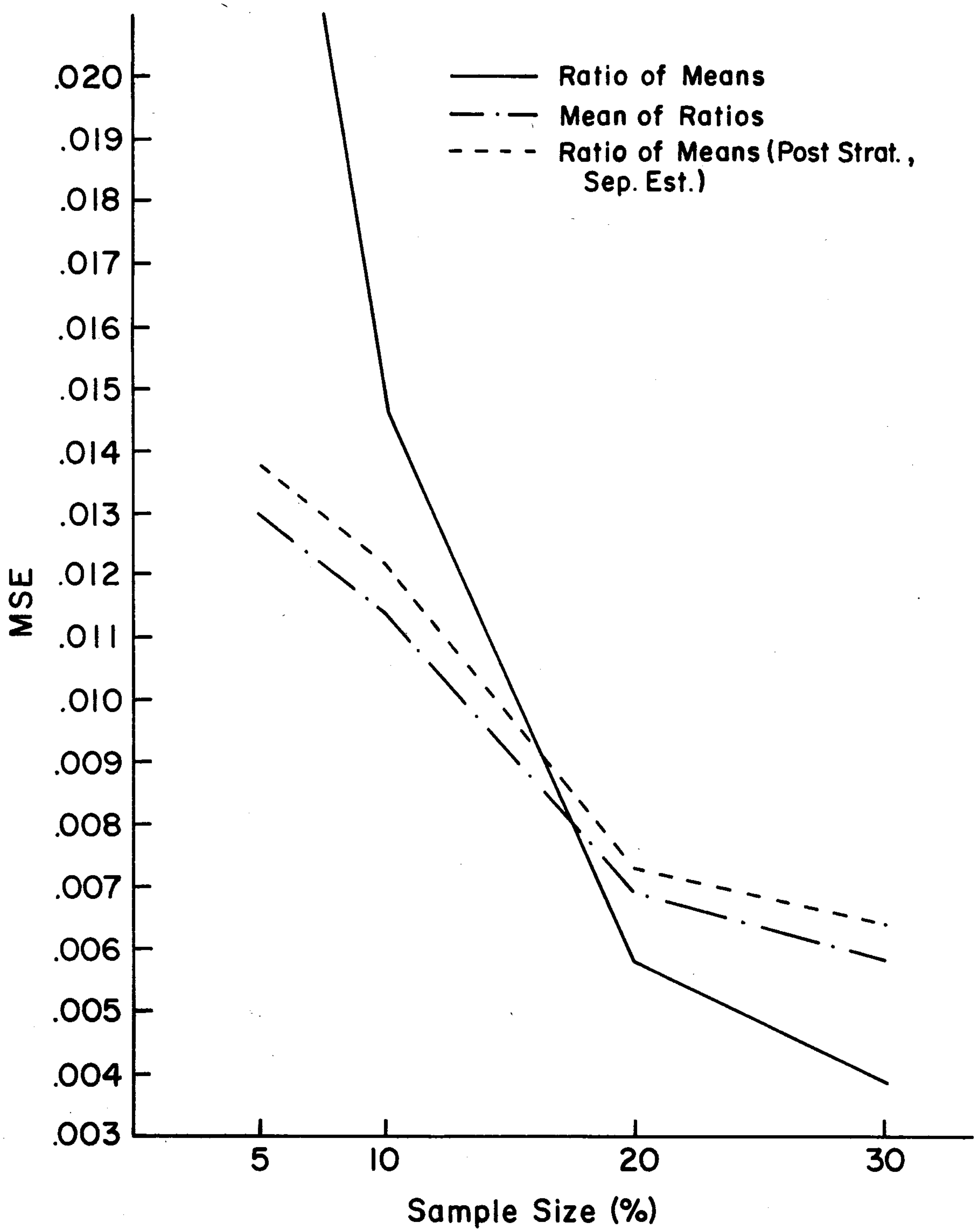


FIGURE D-14. Comparison of Estimations, District 15, System SH

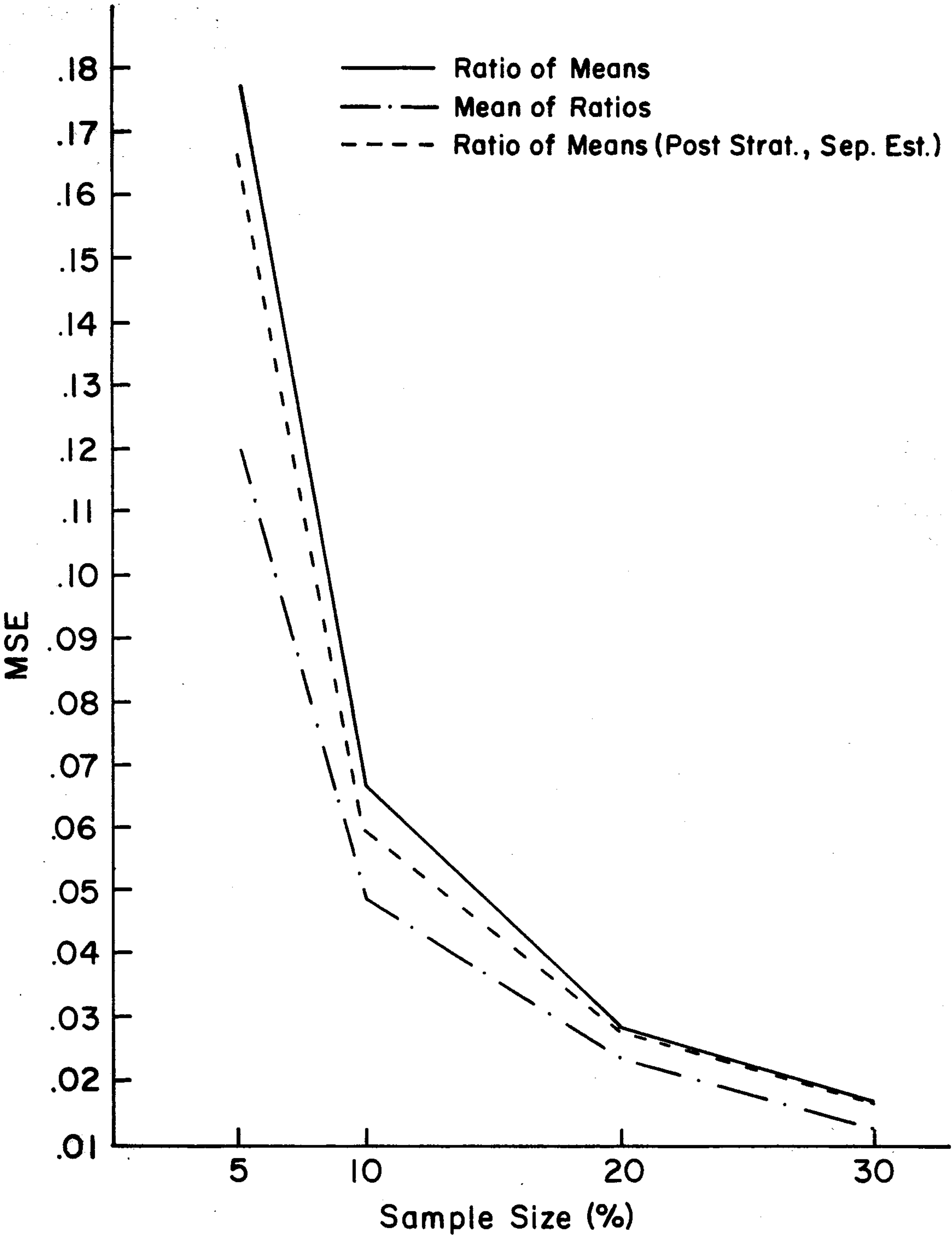


FIGURE D-15. Comparison of Estimations, District 15, System US

APPENDIX E

Table E-1. Statistics on District 8 from the district stratification sampling procedure for various sample sizes.

Data Group	Sample Size %	$\hat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
IH08	5.84	0.9272	1.3704	0.9268	0.5643	0.1904
	10.39	0.4209	1.2954	0.4179	0.3934	0.0762
	15.58	0.2303	1.3035	0.2281	0.2947	0.0495
	20.78	0.2062	1.3694	0.2058	0.2635	0.0437
	25.97	0.1309	1.3409	0.1308	0.2161	0.0251
	31.17	0.1052	1.3180	0.1041	0.1906	0.0214
	35.71	0.1003	1.3566	0.1002	0.1905	0.0176
	40.91	0.0680	1.3339	0.0677	0.1574	0.0125
	46.10	0.0608	1.3401	0.0607	0.1463	0.0119
	50.00	0.0483	1.3309	0.0480	0.1313	0.0093
SH08	5.19	0.0602	0.3793	0.0598	0.4914	0.1349
	10.00	0.0287	0.3659	0.0275	0.3492	0.0574
	15.19	0.0210	0.3808	0.0206	0.2944	0.0443
	20.00	0.0136	0.3818	0.0133	0.2360	0.0294
	25.56	0.0103	0.3946	0.0103	0.2065	0.0219
	30.00	0.0072	0.3906	0.0071	0.1709	0.0157
	35.19	0.0067	0.3804	0.0063	0.1608	0.0160
	40.00	0.0051	0.3851	0.0049	0.1416	0.0117
	45.19	0.0044	0.3803	0.0040	0.1361	0.0088
	50.00	0.0032	0.3811	0.0029	0.1142	0.0071
US08	5.56	0.0547	0.3470	0.0528	0.4797	0.1292
	10.49	0.0298	0.3373	0.0270	0.3657	0.0623
	14.51	0.0204	0.3582	0.0194	0.2894	0.0506
	20.37	0.0137	0.3394	0.0111	0.2477	0.0287
	25.00	0.0106	0.3442	0.0085	0.2156	0.0234
	30.25	0.0102	0.3348	0.0071	0.2141	0.0210
	35.19	0.0079	0.3394	0.0053	0.1860	0.0172
	40.12	0.0070	0.3376	0.0043	0.1766	0.0151
	45.37	0.0060	0.3396	0.0035	0.1633	0.0129
	50.00	0.0055	0.3359	0.0026	0.1562	0.0117
FM08	5.01	0.0055	0.1291	0.0054	0.4314	0.0933
	10.02	0.0027	0.1324	0.0026	0.2976	0.0476
	15.03	0.0016	0.1350	0.0015	0.2353	0.0249
	20.04	0.0012	0.1378	0.0012	0.1929	0.0232
	25.05	0.0009	0.1375	0.0009	0.1778	0.0132
	30.06	0.0007	0.1334	0.0006	0.1536	0.0107
	35.07	0.0006	0.1336	0.0005	0.1345	0.0101
	40.09	0.0005	0.1360	0.0005	0.1262	0.0084
	45.10	0.0004	0.1337	0.0004	0.1134	0.0077
	50.00	0.0003	0.1345	0.0003	0.1037	0.0064

Table E-2. Statistics on District 11 from the district stratification sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
SH11	5.00	0.1205	0.5749	0.1185	0.4581	0.1036
	10.00	0.0563	0.5736	0.0542	0.3097	0.0506
	15.00	0.0398	0.5734	0.0376	0.2660	0.0328
	20.00	0.0253	0.5915	0.0245	0.2036	0.0244
	25.00	0.0195	0.5828	0.0182	0.1796	0.0186
	30.00	0.0168	0.5685	0.0142	0.1695	0.0150
	35.00	0.0132	0.5787	0.0115	0.1469	0.0127
	40.00	0.0114	0.5761	0.0095	0.1389	0.0104
	45.00	0.0083	0.5928	0.0076	0.1155	0.0083
	50.00	0.0077	0.5779	0.0060	0.1150	0.0067
US11	5.39	0.2805	1.1888	0.2768	0.3416	0.0629
	10.37	0.1337	1.1526	0.1242	0.2390	0.0284
	15.35	0.0913	1.1581	0.0829	0.1938	0.0209
	20.33	0.0635	1.1514	0.0537	0.1602	0.0149
	25.31	0.0479	1.1655	0.0408	0.1375	0.0118
	30.29	0.0442	1.1448	0.0331	0.1367	0.0096
	35.27	0.0324	1.1649	0.0251	0.1161	0.0072
	40.25	0.0252	1.1650	0.0180	0.1014	0.0059
	45.23	0.0230	1.1769	0.0176	0.0959	0.0055
	50.62	0.0229	1.1624	0.0156	0.0906	0.0052
FM11	5.05	0.0161	0.3793	0.0160	0.2724	0.0433
	10.10	0.0082	0.3796	0.0081	0.1927	0.0228
	15.02	0.0061	0.3827	0.0060	0.1701	0.0159
	20.07	0.0038	0.3784	0.0037	0.1287	0.0113
	25.00	0.0031	0.3809	0.0030	0.1215	0.0079
	30.05	0.0022	0.3778	0.0021	0.1016	0.0056
	35.22	0.0018	0.3834	0.0017	0.0900	0.0053
	40.02	0.0016	0.3832	0.0014	0.0863	0.0040
	45.07	0.0013	0.3830	0.0012	0.0792	0.0034
	50.00	0.0012	0.3829	0.0011	0.0737	0.0033

Table E-3. Statistics on District 15 from the district stratification sampling procedure for various sample sizes.

Data Group	Sample Size %	$\hat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	σ^2_{error}
IH15	5.57	0.3029	0.4207	0.3014	1.0185	0.3942
	10.22	0.1643	0.3864	0.1589	0.7337	0.2381
	15.48	0.0996	0.3582	0.0893	0.5854	0.1281
	20.12	0.0631	0.3903	0.0583	0.4495	0.0962
	25.39	0.0516	0.3920	0.0470	0.4048	0.0800
	30.03	0.0465	0.3739	0.0390	0.3932	0.0650
	35.60	0.0351	0.3943	0.0308	0.3338	0.0543
	40.25	0.0301	0.3879	0.0249	0.3146	0.0434
	45.20	0.0234	0.3843	0.0177	0.2703	0.0375
	50.46	0.0217	0.3819	0.0156	0.2594	0.0354
SH15	5.25	0.0175	0.1465	0.0135	0.5360	0.1100
	10.10	0.0098	0.1541	0.0067	0.4057	0.0584
	15.15	0.0081	0.1465	0.0041	0.3640	0.0510
	20.00	0.0068	0.1499	0.0032	0.3294	0.0454
	25.25	0.0059	0.1513	0.0024	0.3047	0.0400
	30.30	0.0058	0.1459	0.0020	0.2998	0.0357
	35.15	0.0049	0.1496	0.0013	0.2922	0.0266
	40.00	0.0045	0.1513	0.0010	0.2849	0.0204
	45.05	0.0041	0.1531	0.0009	0.2718	0.0189
	50.30	0.0040	0.1511	0.0007	0.2715	0.0159
US15	5.30	0.1747	0.3679	0.1737	0.7432	0.5398
	10.28	0.0985	0.4178	0.0982	0.5925	0.2647
	15.26	0.0632	0.3849	0.0632	0.5095	0.1357
	20.25	0.0426	0.4055	0.0425	0.4313	0.0799
	25.23	0.0318	0.4105	0.0317	0.3860	0.0496
	30.22	0.0213	0.4004	0.0213	0.3047	0.0402
	35.20	0.0179	0.3914	0.0178	0.2852	0.0305
	40.50	0.0137	0.3960	0.0136	0.2467	0.0245
	45.17	0.0093	0.4025	0.0093	0.2016	0.0177
	50.16	0.0087	0.3928	0.0086	0.1896	0.0182
FM15	5.09	0.0054	0.0821	0.0039	0.5214	0.1001
	10.08	0.0034	0.0941	0.0027	0.4129	0.0663
	15.08	0.0022	0.0926	0.0015	0.3233	0.0495
	20.07	0.0016	0.0925	0.0009	0.2795	0.0349
	25.07	0.0013	0.0948	0.0007	0.2510	0.0288
	30.06	0.0013	0.0924	0.0006	0.2510	0.0278
	35.06	0.0011	0.0960	0.0005	0.2248	0.0234
	40.96	0.0010	0.0943	0.0003	0.2237	0.0199
	45.05	0.0009	0.0942	0.0003	0.2196	0.0165
	50.00	0.0009	0.0941	0.0002	0.2214	0.0142

Table E-4. Statistics on District 8 from the county stratification (Method A) sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	σ^2_{error}
IH08	6.49	0.3839	1.3653	0.3837	0.3749	0.0701
	11.69	0.2094	1.2876	0.2055	0.2699	0.0421
	16.23	0.1163	1.3431	0.1163	0.2121	0.0188
	21.43	0.0914	1.3284	0.0910	0.1773	0.0187
	25.31	0.0776	1.4041	0.0747	0.1636	0.0158
	31.82	0.0549	1.3310	0.0545	0.1401	0.0105
	37.01	0.0489	1.3627	0.0488	0.1285	0.0103
	40.91	0.0381	1.3450	0.0382	0.1182	0.0070
	46.75	0.0280	1.3413	0.0279	0.1007	0.0052
	50.00	0.0270	1.3497	0.0270	0.0990	0.0050
SH08	8.89	0.0372	0.3096	0.0291	0.4113	0.0636
	12.59	0.0251	0.3597	0.0234	0.3221	0.0529
	17.41	0.0151	0.3504	0.0126	0.2533	0.0301
	22.59	0.0118	0.3590	0.0101	0.2201	0.0253
	26.67	0.0097	0.3726	0.0089	0.1985	0.0210
	32.22	0.0078	0.3612	0.0063	0.1781	0.0170
	37.41	0.0059	0.3633	0.0046	0.1575	0.0139
	42.22	0.0050	0.3594	0.0044	0.1550	0.0131
	47.41	0.0043	0.3651	0.0031	0.1306	0.0099
	51.11	0.0040	0.3662	0.0028	0.1256	0.0091
US08	8.64	0.0295	0.3013	0.0216	0.3635	0.0615
	12.65	0.0218	0.3143	0.0160	0.3059	0.0494
	17.28	0.0171	0.3238	0.0127	0.2766	0.0356
	21.91	0.0115	0.3404	0.0090	0.2259	0.0251
	27.16	0.0113	0.3327	0.0080	0.2238	0.0230
	32.10	0.0084	0.3402	0.0059	0.1907	0.0185
	37.35	0.0069	0.3432	0.0047	0.1654	0.0181
	41.98	0.0056	0.3306	0.0039	0.1603	0.0179
	47.53	0.0055	0.3385	0.0030	0.1601	0.0117
	50.93	0.0048	0.3439	0.0027	0.1438	0.0107
FM08	5.76	0.0044	0.1275	0.0042	0.3916	0.0706
	10.66	0.0022	0.1333	0.0021	0.2658	0.0401
	15.67	0.0016	0.1345	0.0016	0.2349	0.0280
	20.47	0.0012	0.1331	0.0011	0.2040	0.0189
	25.37	0.0008	0.1347	0.0008	0.1690	0.0144
	31.02	0.0007	0.1330	0.0007	0.1564	0.0130
	35.61	0.0005	0.1355	0.0005	0.1296	0.0095
	40.51	0.0005	0.1335	0.0005	0.1277	0.0085
	45.63	0.0003	0.1352	0.0003	0.1063	0.0061
	50.32	0.0003	0.1347	0.0003	0.0950	0.0056

Table E-5. Statistics on District 11 from the county stratification (Method A) sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
SH11	7.00	0.0795	0.7352	0.0662	0.3594	0.0776
	11.67	0.0360	0.6020	0.0357	0.2415	0.0354
	16.67	0.0212	0.5790	0.0196	0.1947	0.0173
	21.67	0.0165	0.5940	0.0158	0.1666	0.0151
	26.00	0.0147	0.5877	0.0126	0.1565	0.0126
	31.67	0.0133	0.5646	0.0112	0.1514	0.0126
	36.67	0.0080	0.6021	0.0077	0.1140	0.0078
	41.67	0.0078	0.5954	0.0065	0.1075	0.0071
	46.33	0.0071	0.5743	0.0057	0.1030	0.0064
	50.67	0.0061	0.5830	0.0047	0.0981	0.0062
US11	8.71	0.1431	1.0941	0.1188	0.2478	0.0302
	12.43	0.0971	1.1275	0.0821	0.2002	0.0220
	17.01	0.0671	1.1611	0.0592	0.1655	0.0155
	21.58	0.0480	1.1586	0.0396	0.1423	0.0105
	26.97	0.0360	1.1641	0.0286	0.1216	0.0083
	31.95	0.0318	1.1699	0.0254	0.1141	0.0073
	36.51	0.0239	1.1900	0.0203	0.0976	0.0058
	41.08	0.0212	1.1746	0.0155	0.0950	0.0045
	46.47	0.0171	1.1903	0.0135	0.0858	0.0036
	51.04	0.0169	1.1730	0.0110	0.0850	0.0036
FM11	5.53	0.0161	0.3836	0.0160	0.2686	0.0457
	10.58	0.0080	0.3853	0.0077	0.1887	0.0226
	15.50	0.0046	0.3775	0.0045	0.1455	0.0122
	20.43	0.0032	0.3753	0.0032	0.1217	0.0089
	25.36	0.0026	0.3841	0.0024	0.1091	0.0071
	30.65	0.0021	0.3825	0.0019	0.0966	0.0058
	35.46	0.0021	0.3845	0.0019	0.0992	0.0052
	40.50	0.0014	0.3795	0.0013	0.0809	0.0034
	45.43	0.0011	0.3810	0.0010	0.0741	0.0028
	50.36	0.0011	0.3816	0.0010	0.0732	0.0026

Table E-6. Statistics on District 15 from the county stratification (Method A) sampling procedure for various sample sizes.

Data Group	Sample Size %	$\hat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	σ^2_{error}
IH15	7.12	0.1396	0.2739	0.1049	0.7710	0.1682
	11.15	0.0940	0.3271	0.0763	0.5662	0.1237
	16.10	0.0696	0.3601	0.0596	0.4743	0.1039
	21.05	0.0498	0.3636	0.0405	0.4007	0.0747
	26.32	0.0392	0.3510	0.0274	0.3549	0.0595
	31.27	0.0364	0.3485	0.0240	0.3499	0.0497
	36.84	0.0292	0.3601	0.0193	0.3016	0.0472
	41.18	0.0269	0.3660	0.0180	0.2974	0.0385
	46.44	0.0239	0.3561	0.0131	0.2739	0.0379
	50.77	0.0181	0.3703	0.0100	0.2402	0.0276
SH15	7.68	0.0166	0.1177	0.0081	0.5446	0.0800
	12.12	0.0110	0.1313	0.0048	0.4375	0.0584
	16.97	0.0087	0.1364	0.0032	0.3911	0.0435
	21.82	0.0070	0.1460	0.0027	0.3574	0.0394
	26.46	0.0067	0.1398	0.0020	0.3306	0.0336
	31.92	0.0059	0.1427	0.0014	0.3277	0.0274
	36.36	0.0058	0.1427	0.0013	0.3247	0.0266
	41.62	0.0054	0.1435	0.0009	0.3183	0.0200
	46.67	0.0054	0.1411	0.0006	0.3181	0.0140
	51.31	0.0052	0.1432	0.0006	0.3180	0.0136
US15	4.35	0.0510	0.2617	0.0319	0.4893	0.0796
	13.08	0.0336	0.3614	0.0301	0.3681	0.0743
	17.45	0.0298	0.3750	0.0293	0.3564	0.0594
	22.43	0.0200	0.3620	0.0185	0.2853	0.0435
	27.41	0.0157	0.3897	0.0155	0.2507	0.0350
	32.40	0.0128	0.3819	0.0125	0.2276	0.0281
	37.69	0.0094	0.3822	0.0091	0.1964	0.0202
	42.37	0.0074	0.3850	0.0071	0.1705	0.0170
	47.66	0.0061	0.3884	0.0060	0.1515	0.0152
	52.02	0.0046	0.3917	0.0046	0.1343	0.0108
FM15	5.72	0.0046	0.0933	0.0039	0.4745	0.0941
	10.63	0.0029	0.0962	0.0020	0.3759	0.0573
	15.62	0.0019	0.0966	0.0013	0.3018	0.0384
	20.53	0.0015	0.0944	0.0009	0.2744	0.0298
	25.52	0.0015	0.0921	0.0007	0.2722	0.0291
	30.52	0.0013	0.0925	0.0005	0.2584	0.0265
	35.24	0.0012	0.0914	0.0004	0.2451	0.0199
	40.60	0.0010	0.0950	0.0004	0.2284	0.0193
	45.78	0.0009	0.0945	0.0002	0.2214	0.0129
	50.32	0.0009	0.0942	0.0002	0.2186	0.0129

Table E-7. Statistics on District 8 from the county stratification (Method B) sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
IH08	6.49	0.3869	1.3779	0.3862	0.3735	0.0728
	11.69	0.2251	1.2949	0.2220	0.2794	0.0454
	16.23	0.1167	1.3570	0.1167	0.2117	0.0192
	21.43	0.0955	1.3437	0.0955	0.1808	0.0197
	25.32	0.0726	1.3868	0.0713	0.1579	0.0149
	31.82	0.0560	1.3473	0.0560	0.1417	0.0107
	37.02	0.0490	1.3553	0.0490	0.1286	0.0103
	40.91	0.0390	1.3608	0.0389	0.1201	0.0070
	46.73	0.0280	1.3390	0.0278	0.1109	0.0052
	50.00	0.0270	1.3488	0.0270	0.0987	0.0051
SH08	8.89	0.0432	0.3603	0.0416	0.4294	0.0854
	12.59	0.0264	0.3751	0.0258	0.3291	0.0569
	17.41	0.0149	0.3644	0.0136	0.2490	0.0309
	22.59	0.0116	0.3740	0.0109	0.2176	0.0252
	26.67	0.0099	0.3779	0.0095	0.2006	0.0219
	32.22	0.0077	0.3637	0.0063	0.1761	0.0169
	37.41	0.0056	0.3713	0.0048	0.1495	0.0127
	42.22	0.0057	0.3662	0.0045	0.1422	0.0124
	47.41	0.0041	0.3690	0.0032	0.1267	0.0097
	51.11	0.0039	0.3682	0.0029	0.1239	0.0090
US08	8.64	0.0415	0.3435	0.0393	0.4179	0.0981
	12.65	0.0209	0.3330	0.0177	0.3005	0.0473
	17.28	0.0163	0.3472	0.0144	0.2640	0.0371
	21.91	0.0113	0.3417	0.0090	0.2228	0.0246
	27.16	0.0107	0.3423	0.0084	0.2177	0.0229
	32.10	0.0080	0.3451	0.0060	0.1862	0.0180
	37.35	0.0068	0.3447	0.0047	0.1635	0.0179
	41.98	0.0055	0.3472	0.0037	0.1507	0.0136
	47.53	0.0055	0.3409	0.0031	0.1566	0.0114
	50.93	0.0046	0.3458	0.0027	0.1413	0.0106
FM08	5.76	0.0046	0.1276	0.0044	0.3984	0.0738
	10.66	0.0022	0.1336	0.0022	0.2686	0.0408
	15.67	0.0016	0.1336	0.0016	0.2350	0.0278
	20.47	0.0012	0.1336	0.0012	0.2048	0.0190
	25.37	0.0008	0.1335	0.0008	0.1695	0.0142
	31.02	0.0007	0.1333	0.0007	0.1560	0.0127
	35.61	0.0005	0.1353	0.0005	0.1279	0.0095
	40.51	0.0005	0.1335	0.0005	0.1278	0.0085
	45.63	0.0003	0.1348	0.0003	0.1669	0.0061
	50.32	0.0003	0.1347	0.0003	0.0952	0.0056

Table E-8. Statistics on District 11 from the county stratification (Method B) sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
SH11	7.00	0.0604	0.5870	0.0593	0.3198	0.0548
	11.67	0.0373	0.5619	0.0339	0.2462	0.0364
	16.67	0.0213	0.5849	0.0200	0.1945	0.0175
	21.67	0.0171	0.5854	0.0159	0.1696	0.0156
	26.00	0.0151	0.5683	0.0125	0.1594	0.0140
	31.67	0.0141	0.5653	0.0111	0.1557	0.0125
	36.67	0.0097	0.5730	0.0075	0.1264	0.0092
	41.67	0.0080	0.5809	0.0065	0.1144	0.0077
	46.33	0.0076	0.5759	0.0057	0.1119	0.0074
	50.67	0.0066	0.5753	0.0047	0.1031	0.0067
US11	8.71	0.1853	1.2566	0.1853	0.2783	0.0411
	12.45	0.1034	1.2030	0.1012	0.2078	0.0230
	17.01	0.0661	1.1899	0.0625	0.1648	0.0152
	21.58	0.0454	1.1968	0.0426	0.1368	0.0104
	26.97	0.0343	1.1795	0.0294	0.1172	0.0082
	31.95	0.0293	1.1986	0.0266	0.1084	0.0070
	36.51	0.0240	1.1912	0.0206	0.0980	0.0058
	41.08	0.0199	1.1866	0.0159	0.0918	0.0043
	46.47	0.0171	1.1924	0.0138	0.0858	0.0036
	51.04	0.0158	1.1809	0.0110	0.0819	0.0034
FM11	5.53	0.0170	0.3846	0.0168	0.2750	0.0485
	10.58	0.0082	0.3871	0.0079	0.1916	0.0234
	15.50	0.0046	0.3772	0.0045	0.1455	0.0123
	20.43	0.0033	0.3769	0.0033	0.1225	0.0091
	25.36	0.0026	0.3844	0.0024	0.1092	0.0071
	30.65	0.0021	0.3825	0.0019	0.0468	0.0058
	35.46	0.0020	0.3838	0.0018	0.0987	0.0051
	40.50	0.0014	0.3800	0.0013	0.0814	0.0034
	45.43	0.0011	0.3810	0.0010	0.0741	0.0027
	50.36	0.0011	0.3819	0.0010	0.0734	0.0026

Table E-9. Statistics on District 15 from the county stratification (method B) sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
IH15	7.12	0.1952	0.3556	0.1843	0.8521	0.1963
	11.15	0.1047	0.3632	0.0954	0.5809	0.1576
	16.10	0.0721	0.3774	0.0653	0.4845	0.1058
	21.05	0.0507	0.3817	0.0446	0.3999	0.0798
	26.32	0.0387	0.3602	0.0287	0.3510	0.0596
	31.27	0.0356	0.3610	0.0258	0.3444	0.0496
	36.84	0.0284	0.3707	0.0204	0.2957	0.0466
	41.18	0.0261	0.3748	0.0189	0.2919	0.0383
	46.44	0.0229	0.3635	0.0136	0.2667	0.0373
	50.77	0.0178	0.3722	0.0101	0.2384	0.0275
SH15	7.68	0.0159	0.1519	0.0125	0.5202	0.0898
	12.12	0.0099	0.1461	0.0058	0.4026	0.0627
	16.97	0.0079	0.1438	0.0035	0.3675	0.0431
	21.82	0.0065	0.1479	0.0026	0.3295	0.0385
	26.46	0.0065	0.1440	0.0021	0.3254	0.0344
	31.92	0.0055	0.1464	0.0015	0.3129	0.0278
	36.36	0.0054	0.1459	0.0013	0.3111	0.0266
	41.62	0.0051	0.1458	0.0010	0.3080	0.0201
	46.67	0.0051	0.1428	0.0006	0.3060	0.0139
	51.31	0.0049	0.1453	0.0006	0.3002	0.0161
US15	9.35	0.0803	0.3771	0.0798	0.5476	0.2021
	13.08	0.0412	0.3986	0.0412	0.3958	0.1011
	17.45	0.0343	0.4100	0.0342	0.3756	0.0734
	22.43	0.0214	0.3816	0.0211	0.2925	0.0484
	27.41	0.0164	0.3956	0.0164	0.2573	0.0364
	32.40	0.0134	0.3970	0.0134	0.2343	0.0289
	37.69	0.0100	0.3948	0.0100	0.2028	0.0215
	42.37	0.0076	0.3888	0.0074	0.1736	0.0172
	47.66	0.0063	0.3995	0.0063	0.1554	0.0153
	52.02	0.0048	0.3965	0.0047	0.1366	0.0110
FM15	5.72	0.0046	0.0953	0.0040	0.4732	0.0970
	10.63	0.0028	0.0918	0.0020	0.3725	0.0586
	15.62	0.0019	0.0970	0.0013	0.3002	0.0383
	20.53	0.0015	0.0940	0.0009	0.2764	0.0298
	25.52	0.0015	0.0909	0.0007	0.2740	0.0292
	30.52	0.0013	0.0924	0.0005	0.2485	0.0265
	35.24	0.0011	0.0941	0.0004	0.2397	0.0200
	40.60	0.0010	0.0948	0.0004	0.2294	0.0194
	45.78	0.0009	0.0944	0.0002	0.2220	0.0130
	50.32	0.0009	0.0939	0.0002	0.2209	0.0149

Table E-10. Statistics on District 8 from the systematic sampling procedure for various sample sizes.

Data Group	Sample Size %	$\widehat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
IH08	5.84	0.7699	1.3352	0.7696	0.5163	0.1558
	10.39	0.1381	1.3506	0.1381	0.2202	0.0273
	15.58	0.1112	1.3578	0.1112	0.1867	0.0262
	20.78	0.0526	1.2856	0.0484	0.1557	0.0046
	25.97	0.0833	1.3583	0.0832	0.1771	0.0143
	31.17	0.0826	1.3272	0.0821	0.2046	0.0035
	35.71	0.0337	1.2904	0.0301	0.1285	0.0020
	40.91	0.0295	1.2869	0.0255	0.1183	0.0022
	46.10	0.0051	1.3119	0.0037	0.0450	0.0008
	50.00	0.0112	1.3179	0.0112	0.0784	0.0006
SH08	5.18	0.0522	0.3744	0.0516	0.4694	0.1060
	10.00	0.0324	0.3702	0.0315	0.3772	0.0601
	15.19	0.0129	0.3941	0.0129	0.2297	0.0279
	20.00	0.0016	0.3702	0.0007	0.0811	0.0035
	25.19	0.0060	0.0490	0.0059	0.1623	0.0112
	30.00	0.0107	0.3989	0.0107	0.2346	0.0116
	35.19	0.0020	0.3592	0.0004	0.1020	0.0023
	40.00	0.0009	0.3722	0.0001	0.0696	0.0006
	45.19	0.0100	0.3554	0.0080	0.2239	0.0125
	50.00	0.0052	0.3702	0.0043	0.1633	0.0055
US08	5.25	0.0254	0.3451	0.0234	0.3378	0.0529
	10.19	0.0225	0.3152	0.0169	0.3373	0.0343
	14.51	0.0209	0.3366	0.0181	0.3165	0.0374
	20.06	0.0273	0.3955	0.0272	0.3903	0.0268
	25.00	0.0058	0.3440	0.0037	0.1811	0.0054
	30.25	0.0011	0.3630	0.0004	0.0705	0.0024
	35.19	0.0013	0.3593	0.0003	0.0786	0.0022
	40.12	0.0091	0.3222	0.0044	0.1740	0.0292
	45.06	0.0031	0.3529	0.0017	0.1061	0.0090
	50.00	0.0037	0.3440	0.0016	0.1179	0.0102
FM08	5.01	0.0074	0.1347	0.0074	0.4857	0.1440
	10.02	0.0027	0.1373	0.0027	0.3511	0.0147
	15.03	0.0024	0.1411	0.0024	0.2873	0.0396
	20.04	0.0005	0.1311	0.0004	0.1230	0.0105
	25.05	0.0003	0.1245	0.0000	0.1106	0.0019
	29.96	0.0007	0.1359	0.0006	0.1634	0.0070
	34.97	0.0004	0.1278	0.0003	0.1153	0.0075
	39.98	0.0008	0.1350	0.0008	0.1994	0.0013
	45.10	0.0010	0.1168	0.0005	0.1655	0.0233
	50.00	0.0001	0.1340	0.0001	0.0517	0.0019

Table E-11. Statistics on District 11 from the systematic sampling procedure for various sample sizes.

Data Group	Sample Size %	$\hat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}^2_{\hat{R}}$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
SH11	5.00	0.0741	0.5768	0.0723	0.3556	0.0664
	10.00	0.0438	0.5768	0.0420	0.2650	0.0438
	15.00	0.0243	0.5708	0.0218	0.1877	0.0279
	20.00	0.0404	0.5768	0.0386	0.2650	0.0350
	25.00	0.0043	0.5768	0.0024	0.0822	0.0044
	30.00	0.0079	0.5383	0.0012	0.1317	0.0032
	35.00	0.0068	0.5391	0.0003	0.1304	0.0007
	40.00	0.0006	0.6094	0.0005	0.0358	0.0003
	45.00	0.0060	0.5566	0.0020	0.1023	0.0051
	50.00	0.0032	0.5768	0.0013	0.0697	0.0034
US11	5.39	0.3080	1.2279	0.3075	0.3670	0.0624
	10.37	0.6390	1.2073	0.0372	0.1288	0.0084
	15.35	0.0589	1.1984	0.0563	0.1425	0.0174
	20.33	0.0112	1.1951	0.0082	0.0845	0.0000
	25.31	0.0392	1.2179	0.0382	0.1611	0.0024
	30.29	0.0241	1.2364	0.0239	0.1123	0.0028
	35.27	0.0051	1.1784	0.0000	0.0573	0.0000
	40.25	0.0033	1.1956	0.0004	0.0436	0.0002
	45.23	0.0041	1.2087	0.0024	0.0393	0.0011
	49.79	0.0119	1.1987	0.0093	0.0771	0.0017
FM11	5.05	0.0156	0.3791	0.0215	0.2757	0.0377
	10.10	0.0090	0.3810	0.0112	0.2093	0.0221
	15.02	0.0010	0.3592	0.0033	0.0773	0.0016
	20.07	0.0013	0.3958	0.0014	0.0698	0.0046
	25.00	0.0010	0.3818	0.0008	0.0724	0.0018
	30.05	0.0015	0.3750	0.0023	0.0930	0.0024
	35.10	0.0013	0.3922	0.0034	0.0771	0.0036
	40.02	0.0008	0.3792	0.0013	0.0769	0.0006
	45.07	0.0026	0.3781	0.0007	0.1367	0.0005
	50.00	0.0009	0.3818	0.0007	0.0724	0.0000

Table E-12. Statistics on District 15 from the systematic sampling procedure for various sample sizes.

Data Group	Sample Size %	$\hat{MSE}$	$\hat{\mu}_{\hat{R}}$	$\hat{\sigma}_{\hat{R}}^2$	$\hat{\mu}_{error}$	$\hat{\sigma}^2_{error}$
IH15	5.26	0.1318	0.3679	0.1234	0.7554	0.0524
	10.22	0.0834	0.3989	0.0797	0.5438	0.0985
	15.17	0.0664	0.4042	0.0633	0.5069	0.0567
	20.12	0.0569	0.3518	0.0452	0.4109	0.1001
	25.08	0.0717	0.3450	0.0585	0.4848	0.1038
	30.03	0.0450	0.3197	0.0254	0.4200	0.0365
	35.29	0.0131	0.3472	0.0004	0.2451	0.0019
	40.25	0.0079	0.3990	0.0042	0.1411	0.0176
	45.20	0.0207	0.3199	0.0010	0.3046	0.0050
	49.85	0.0381	0.3691	0.0298	0.3754	0.0391
SH15	5.05	0.0191	0.1479	0.0153	0.5730	0.1053
	10.75	0.0086	0.1499	0.0050	0.3660	0.0619
	15.15	0.0077	0.1518	0.0043	0.3675	0.0401
	21.29	0.0069	0.1458	0.0027	0.3055	0.0623
	25.05	0.0076	0.1647	0.0056	0.3939	0.0175
	30.10	0.0055	0.1412	0.0007	0.3276	0.0163
	35.15	0.0031	0.1549	0.0000	0.2626	0.0003
	40.00	0.0063	0.1308	0.0001	0.3770	0.0018
	45.05	0.0052	0.1524	0.0019	0.2745	0.0429
	49.90	0.0047	0.1461	0.0007	0.3041	0.0150
US15	5.30	0.0734	0.4048	0.0733	0.4630	0.2441
	10.28	0.0409	0.3914	0.0409	0.4140	0.0844
	15.27	0.0060	0.4069	0.0059	0.1851	0.0030
	20.25	0.0167	0.3840	0.0165	0.3180	0.0033
	25.23	0.0024	0.3975	0.0024	0.1029	0.0046
	30.22	0.0023	0.3904	0.0022	0.1003	0.0042
	35.20	0.0017	0.4267	0.0010	0.0796	0.0045
	40.19	0.0003	0.3856	0.0001	0.0360	0.0004
	45.17	0.0028	0.3878	0.0027	0.1294	0.0009
	49.84	0.0000	0.3977	0.0000	0.0056	0.0000
FM15	5.09	0.0038	0.0966	0.0033	0.4445	0.0687
	10.08	0.0018	0.0899	0.0009	0.3062	0.0302
	15.08	0.0013	0.1022	0.0009	0.2230	0.0402
	20.07	0.0025	0.0852	0.0012	0.3910	0.0178
	25.07	0.0011	0.0958	0.0005	0.2327	0.0207
	30.06	0.0013	0.1007	0.0009	0.2256	0.0362
	35.06	0.0002	0.1048	0.0000	0.1267	0.0003
	40.05	0.0011	0.0893	0.0002	0.2561	0.0135
	45.05	0.0003	0.1024	0.0000	0.1467	0.0015
	49.96	0.0008	0.0938	0.0001	0.2185	0.0099

Table E-13. Mean and variance of the maximum difference between Pavement Score distributions from the district stratification sampling procedure in District 8.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
IH08	5.84	0.2984	0.0121
	10.39	0.2064	0.0055
	15.58	0.1645	0.0035
	20.78	0.1374	0.0037
	25.97	0.1171	0.0018
	31.17	0.1042	0.0018
	35.71	0.0971	0.0015
	40.91	0.0834	0.0011
	46.10	0.0761	0.0009
	50.00	0.0676	0.0008
SH08	5.19	0.1924	0.0049
	10.00	0.1305	0.0020
	15.19	0.1029	0.0014
	20.00	0.0883	0.0010
	25.56	0.0777	0.0010
	30.00	0.0654	0.0007
	35.19	0.0601	0.0005
	40.00	0.0548	0.0004
	45.19	0.0485	0.0003
	50.00	0.0424	0.0003
US08	5.56	0.1565	0.0039
	10.49	0.1126	0.0019
	14.51	0.0944	0.0016
	20.37	0.0760	0.0008
	25.00	0.0652	0.0006
	30.25	0.0585	0.0005
	35.19	0.0522	0.0004
	40.12	0.0459	0.0004
	45.37	0.0413	0.0003
	50.00	0.0382	0.0002
FM08	5.01	0.1028	0.0014
	10.02	0.0677	0.0006
	15.03	0.0529	0.0004
	20.04	0.0450	0.0003
	25.05	0.0404	0.0002
	30.06	0.0351	0.0002
	35.07	0.0323	0.0002
	40.09	0.0280	0.0001
	45.10	0.0262	0.0001
	50.00	0.0224	0.0001

Table E-14. Mean and variance of the maximum difference between Pavement Score distributions from the district stratification sampling procedure in District 11.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
SH11	5.00	0.1846	0.0047
	10.00	0.1248	0.0020
	15.00	0.1002	0.0012
	20.00	0.0869	0.0011
	25.00	0.0723	0.0007
	30.00	0.0644	0.0005
	35.00	0.0593	0.0004
	40.00	0.0518	0.0004
	45.00	0.0462	0.0003
	50.00	0.0415	0.0003
US11	5.39	0.2194	0.0065
	10.37	0.1564	0.0034
	15.35	0.1266	0.0023
	20.33	0.1026	0.0015
	25.31	0.0893	0.0013
	30.29	0.0798	0.0010
	35.27	0.0722	0.0008
	40.25	0.0652	0.0006
	45.23	0.0576	0.0004
	50.62	0.0537	0.0004
FM11	5.05	0.1095	0.0015
	10.10	0.0751	0.0007
	15.02	0.0627	0.0005
	20.07	0.0516	0.0004
	25.00	0.0452	0.0002
	30.05	0.0399	0.0002
	35.22	0.0336	0.0002
	40.02	0.0319	0.0001
	45.07	0.0284	0.0001
	50.00	0.0250	0.0001

Table E-15. Mean and variance of the maximum difference between Pavement Score distributions from the district stratification sampling procedure in District 15.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
IH15	5.57	0.1523	0.0043
	10.22	0.1113	0.0025
	15.48	0.0875	0.0016
	20.12	0.0733	0.0010
	25.39	0.0636	0.0009
	30.03	0.0562	0.0007
	35.60	0.0519	0.0007
	40.25	0.0465	0.0005
	45.20	0.0418	0.0004
	50.46	0.0370	0.0003
SH15	5.25	0.1325	0.0033
	10.10	0.0883	0.0014
	15.15	0.0695	0.0008
	20.00	0.0593	0.0006
	25.25	0.0518	0.0004
	30.30	0.0472	0.0004
	35.15	0.0399	0.0003
	40.00	0.0380	0.0002
	45.05	0.0325	0.0002
	50.30	0.0285	0.0002
US15	5.30	0.1523	0.0038
	10.28	0.1063	0.0017
	15.26	0.0850	0.0013
	20.25	0.0749	0.0010
	25.23	0.0605	0.0007
	30.22	0.0549	0.0005
	35.20	0.0526	0.0005
	40.50	0.0445	0.0004
	45.17	0.0403	0.0003
	50.16	0.0368	0.0003
FM15	5.09	0.0799	0.0011
	10.08	0.0550	0.0005
	15.08	0.0443	0.0003
	20.07	0.0367	0.0002
	25.07	0.0325	0.0002
	30.06	0.0286	0.0001
	35.06	0.0261	0.0001
	40.96	0.0222	0.0001
	45.05	0.0201	0.0001
	50.00	0.0196	0.0001

Table E-16. Mean and variance of the maximum difference between Pavement Score distributions from the county stratification (method A) sampling procedure in District 8.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
IH08	6.49	0.1965	0.0039
	11.69	0.1414	0.0023
	16.23	0.1179	0.0017
	21.43	0.0966	0.0013
	25.32	0.0920	0.0010
	31.82	0.0780	0.0008
	37.01	0.0696	0.0006
	40.91	0.0640	0.0005
	46.73	0.0559	0.0004
	50.00	0.0522	0.0003
SH08	8.89	0.1360	0.0023
	12.59	0.1145	0.0020
	17.41	0.0924	0.0010
	22.59	0.0763	0.0008
	26.67	0.0698	0.0007
	32.22	0.0613	0.0005
	37.41	0.0538	0.0004
	42.22	0.0481	0.0003
	47.41	0.0428	0.0003
	51.11	0.0414	0.0002
US08	8.64	0.1122	0.0017
	12.65	0.0898	0.0010
	17.28	0.0763	0.0008
	21.91	0.0661	0.0007
	27.16	0.0580	0.0005
	32.10	0.0524	0.0004
	37.35	0.0457	0.0003
	41.98	0.0437	0.0003
	47.53	0.0371	0.0002
	50.93	0.0345	0.0002
FM08	5.76	0.0903	0.0012
	10.66	0.1653	0.0006
	15.67	0.0523	0.0003
	20.47	0.0439	0.0003
	25.37	0.0381	0.0002
	31.02	0.0343	0.0002
	35.61	0.0296	0.0001
	40.51	0.0270	0.0001
	45.63	0.0245	0.0001
	50.32	0.0221	0.0001

Table E-17. Mean and variance of the maximum difference between Pavement Score distributions from the county stratification (method A) sampling procedure in District 11.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
SH11	7.00	0.1543	0.0029
	11.67	0.1052	0.0013
	16.67	0.0861	0.0009
	21.67	0.0737	0.0007
	26.00	0.0687	0.0007
	31.67	0.0571	0.0004
	36.67	0.0511	0.0003
	41.67	0.0450	0.0002
	46.33	0.0425	0.0002
	50.67	0.0381	0.0002
US11	8.71	0.1469	0.0025
	12.45	0.1232	0.0020
	17.01	0.1035	0.0013
	21.58	0.0855	0.0009
	26.97	0.0733	0.0007
	31.95	0.0659	0.0006
	36.51	0.0608	0.0005
	41.08	0.0535	0.0003
	46.47	0.0499	0.0003
	51.04	0.0439	0.0003
FM11	5.53	0.1020	0.0013
	10.58	0.0728	0.0006
	15.50	0.0577	0.0004
	20.43	0.0474	0.0003
	25.36	0.0430	0.0003
	30.65	0.0369	0.0002
	35.46	0.0342	0.0001
	40.50	0.0298	0.0001
	45.43	0.0269	0.0001
	50.36	0.0252	0.0001

Table E-18. Mean and variance of the maximum difference between Pavement Score distributions from the county stratification (method A) sampling procedure in District 15.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
IH15	7.12	0.1055	0.0019
	11.15	0.0811	0.0012
	16.10	0.0680	0.0010
	21.05	0.0537	0.0005
	26.32	0.0488	0.0005
	31.27	0.0428	0.0003
	36.84	0.0379	0.0003
	41.18	0.0354	0.0002
	46.44	0.0306	0.0002
	50.77	0.0289	0.0002
SH15	7.68	0.0945	0.0015
	12.12	0.0717	0.0008
	16.97	0.0581	0.0005
	21.82	0.0512	0.0004
	26.46	0.0462	0.0003
	31.92	0.0382	0.0003
	36.36	0.0364	0.0003
	41.62	0.0314	0.0002
	46.67	0.0282	0.0002
	51.31	0.0270	0.0001
US15	9.35	0.0996	0.0015
	13.08	0.0810	0.0011
	17.45	0.0700	0.0010
	22.43	0.0611	0.0006
	27.41	0.0533	0.0005
	32.40	0.0463	0.0004
	37.69	0.0423	0.0003
	42.37	0.0369	0.0002
	47.66	0.0335	0.0002
	52.02	0.0312	0.0002
FM15	5.72	0.0748	0.0010
	10.63	0.0528	0.0006
	15.62	0.0430	0.0003
	20.53	0.0365	0.0002
	25.52	0.0308	0.0002
	30.52	0.0277	0.0001
	35.24	0.0253	0.0001
	40.60	0.0225	0.0001
	45.78	0.0202	0.0001
	50.32	0.0186	0.0001

Table E-19. Mean and variance of the maximum difference between Pavement Score distributions from the county stratification (method B) sampling procedure in District 8.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
IH08	6.49	0.1971	0.0040
	11.69	0.1434	0.0024
	16.23	0.1180	0.0017
	21.43	0.0968	0.0012
	25.32	0.0913	0.0010
	31.82	0.0778	0.0007
	37.02	0.0697	0.0006
	40.91	0.0635	0.0005
	46.73	0.0560	0.0004
	50.00	0.0523	0.0003
SH08	8.89	0.1456	0.0026
	12.59	0.1135	0.0019
	17.41	0.0934	0.0011
	22.59	0.0770	0.0009
	26.67	0.0693	0.0007
	32.22	0.0618	0.0005
	37.41	0.0541	0.0004
	42.22	0.0483	0.0003
	47.41	0.0430	0.0003
	51.11	0.0415	0.0002
US08	8.64	0.1421	0.0030
	12.65	0.0942	0.0012
	17.28	0.0770	0.0008
	21.91	0.0673	0.0007
	27.16	0.0596	0.0005
	32.10	0.0528	0.0004
	37.35	0.0463	0.0003
	41.98	0.0423	0.0003
	47.53	0.0377	0.0002
	50.93	0.0348	0.0002
FM08	5.76	0.0899	0.0012
	10.66	0.0655	0.0006
	15.67	0.0523	0.0003
	20.47	0.0439	0.0003
	25.37	0.0380	0.0002
	31.02	0.0343	0.0002
	35.61	0.0295	0.0001
	40.51	0.0270	0.0001
	45.63	0.0244	0.0001
	50.32	0.0221	0.0001

Table E-20. Mean and variance of the maximum difference between Pavement Score distributions from the county stratification (method B) sampling procedure in District 11.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
SH11	7.00	0.1530	0.0030
	11.67	0.1053	0.0013
	16.67	0.0861	0.0009
	21.67	0.0740	0.0007
	26.00	0.0690	0.0007
	31.67	0.0571	0.0004
	36.67	0.0511	0.0003
	41.67	0.0451	0.0002
	46.33	0.0425	0.0002
	50.67	0.0380	0.0002
US11	8.71	0.1711	0.0037
	12.45	0.1289	0.0021
	17.01	0.1040	0.0013
	21.58	0.0862	0.0009
	26.97	0.0732	0.0007
	31.95	0.0659	0.0006
	36.51	0.0613	0.0005
	41.08	0.0535	0.0003
	46.47	0.0500	0.0003
	51.04	0.0438	0.0003
FM11	5.53	0.1023	0.0013
	10.58	0.0729	0.0006
	15.50	0.0577	0.0004
	20.43	0.0473	0.0003
	25.36	0.0430	0.0003
	30.65	0.0369	0.0002
	35.46	0.0341	0.0001
	40.50	0.0298	0.0001
	45.43	0.0269	0.0001
	50.36	0.0252	0.0001

Table E-21. Mean and variance of the maximum difference between Pavement Score distributions from the county stratification (method B) sampling procedure in District 15.

Data Group	Sample Size (%)	Mean Max. Difference	Variance of Difference
IH15	7.12	0.1170	0.0024
	11.15	0.0854	0.0014
	16.10	0.0691	0.0011
	21.05	0.0545	0.0005
	26.32	0.0494	0.0005
	31.27	0.0430	0.0003
	36.84	0.0382	0.0003
	41.18	0.0356	0.0002
	46.44	0.0308	0.0001
	50.77	0.0287	0.0001
SH15	7.68	0.1046	0.0018
	12.12	0.0744	0.0009
	16.97	0.0596	0.0005
	21.82	0.0522	0.0004
	26.46	0.0467	0.0003
	31.92	0.0388	0.0003
	36.36	0.0366	0.0003
	41.62	0.0315	0.0002
	46.67	0.0284	0.0001
	51.31	0.0271	0.0001
US15	9.35	0.1162	0.0025
	13.08	0.0869	0.0012
	17.45	0.0718	0.0010
	22.43	0.0623	0.0006
	27.41	0.0541	0.0005
	32.40	0.0470	0.0004
	37.69	0.0428	0.0003
	42.37	0.0373	0.0002
	47.66	0.0337	0.0002
	52.02	0.0314	0.0602
FM15	5.72	0.0762	0.0010
	10.63	0.0530	0.0006
	15.62	0.0430	0.0003
	20.53	0.0365	0.0002
	25.52	0.0308	0.0002
	30.52	0.0277	0.0001
	35.24	0.0254	0.0001
	40.60	0.0225	0.0001
	45.78	0.0202	0.0001
	50.32	0.0186	0.0001

APPENDIX F

Computer Programs for Simulation Runs

```
C
C
C    THIS PROGRAM RANDOMLY SELECTS A 5% SAMPLE AND CALCULATES THE
C    ERROR IN ESTIMATING THE PERCENTAGE OF ROADS BELOW A SCORE OF
C    40 USING TWO ESTIMATION METHODS.
C    METHOD 1:  THE PERCENTAGE IS CALCULATED DIRECTLY FROM THE
C               SAMPLE
C    METHOD 2:  THE BETA DISTRIBUTION IS USED IN THE ESTIMATION
C               PROCESS
C    THIS ENTIRE PROCEDURE IS REPEATED 300 TIMES, HENCE A MEAN ERROR
C    CAN BE CALCULATED FOR EACH METHOD.
C
C
      DIMENSION SCORE(1200),DESIGN(1200),SAMP(1200),PLACE1(1200)
      DIMENSION PLACE2(1200),NUMOBS(11),DIF1(300),DIF2(300),TRPCT(11)
      DIMENSION NS(11),IPER(1200),NARRAY(1200)
      REAL MSCORE
      REAL*8 DSEED,PLACE1,PLACE2
      INTEGER DESIGN
      READ(5,1)(NUMOBS(I),I=1,11)
    1 FORMAT(11I4)
      READ(5,3)(NS(I),I=1,11)
    3 FORMAT(11I3)
      READ(5,4)(TRPCT(I),I=1,11)
    4 FORMAT(11F6.4)
      DO 500 K=1,11
      KOUNT=0
      TOT4=0.0
      TOT5=0.0
      DSEED=123457.0D0
      NSAMP=NS(K)
      NUM=NUMOBS(K)
      L=K+9
      DO 23 I=1,NUM
      READ(L,2)PLACE1(I),PLACE2(I),DESIGN(I),SAMP(I)
    2 FORMAT(T13,A8,A6,19X,I1,3X,F3.0)
   23 CONTINUE
      DO 300 J=1,300
      NN=0
C
C
C    GGPER IS AN IMSL SUBROUTINE WHICH GENERATES A RANDOM PERMUTATION
C    OF NUM INTEGERS
C
C
      CALL GGPER(DSEED,NUM,IPER)
C
C
C    THIS SECTION GENERATES A RANDOM SAMPLE OF SIZE NN
C
C
   80 DO 100 I=1,NUM
      IF(I.EQ.1) GO TO 90
      DO 110 KIS=1,NN
      IF(IPER(I).EQ.NARRAY(KIS))GO TO 100
  110 CONTINUE
   90 IF(DESIGN(IPER(I)).LT.3)GO TO 3000
      IF(IPER(I).EQ.1)GO TO 20
      IF(IPER(I).EQ.NUM)GO TO 40
      JJ=IPER(I)-1
```

```
      IF(PLACE1(IPER(I)).EQ.PLACE1(JJ))GO TO 60
      KK=IPER(I)+1
      IF(PLACE1(IPER(I)).EQ.PLACE1(KK))GO TO 70
 3000 NN=NN+1
      NARRAY(NN)=IPER(I)
      SCORE(NN)=SAMP(IPER(I))
      IF(NN.GE.NSAMP)GO TO 1000
      GO TO 100
   20 KK=IPER(I)+1
      IF(PLACE1(IPER(I)).EQ.PLACE1(KK))GO TO 30
      GO TO 3000
   30 IF(PLACE2(IPER(I)).NE.PLACE2(KK))GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      SCORE(NN)=SAMP(IPER(I))
      NN=NN+1
      NARRAY(NN)=KK
      SCORE(NN)=SAMP(KK)
      IF(NN.GE.NSAMP)GO TO 1000
      GO TO 100
   40 JJ=IPER(I)-1
      IF(PLACE1(IPER(I)).EQ.PLACE1(JJ))GO TO 50
      GO TO 3000
   50 IF(PLACE2(IPER(I)).NE.PLACE2(JJ))GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      SCORE(NN)=SAMP(IPER(I))
      NN=NN+1
      NARRAY(NN)=JJ
      SCORE(NN)=SAMP(JJ)
      IF(NN.GE.NSAMP)GO TO 1000
      GO TO 100
   60 IF(PLACE2(IPER(I)).NE.PLACE2(JJ))GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      SCORE(NN)=SAMP(IPER(I))
      NN=NN+1
      NARRAY(NN)=JJ
      SCORE(NN)=SAMP(JJ)
      IF(NN.GE.NSAMP)GO TO 1000
      GO TO 100
   70 IF(PLACE2(IPER(I)).NE.PLACE2(KK))GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      SCORE(NN)=SAMP(IPER(I))
      NN=NN+1
      NARRAY(NN)=KK
      SCORE(NN)=SAMP(KK)
      IF(NN.GE.NSAMP)GO TO 1000
  100 CONTINUE
C
C
C   THIS SECTION CALCULATES THE ERROR OF THE TWO METHODS AT EACH
C   ITERATION
C
C
 1000 TOT1=0.0
      TOT2=0.0
      TOT3=0.0
      DO 120 I=1,NN
```

```
      IF(SCORE(I).LE.40.)TOT1=TOT1+1.
      TOT2=TOT2+SCORE(I)
  120 CONTINUE
      PCT=TOT1/NN
      MSCORE=TOT2/(NN*100.)
      DO 130 I=1,NN
  130 TOT3=TOT3+(((SCORE(I)/100.)-MSCORE)**2)
      VSCORE=TOT3/NN
      A=((1.-MSCORE)*(MSCORE**2)/VSCORE)-MSCORE
      IF(A.GT.0.)GO TO 111
      GO TO 300
  111 KOUNT=KOUNT+1
      B=(A-(MSCORE*A))/MSCORE
      X=0.40
C
C
C    MDBETA IS AN IMSL SUBROUTINE WICH RETURNS A CUMULATIVE PERCENTAGE,
C    P, BELOW X FROM A BETA DISTRIBUTION WITH PARAMETERS A AND B
C
C
      CALL MDBETA(X,A,B,P,IER)
      DIF1(J)=ABS(PCT-TRPCT(K))
      DIF2(J)=ABS(P-TRPCT(K))
  300 CONTINUE
C
C
C    THIS SECTION CALCULATES THE MEAN ERROR OF THE TWO METHODS OVER
C    THE 300 ITERATIONS
C
C
      DO 140 J=1,KOUNT
      TOT4=TOT4+DIF1(J)
      TOT5=TOT5+DIF2(J)
  140 CONTINUE
      ERROR1=TOT4/KOUNT
      ERROR2=TOT5/KOUNT
      WRITE(6,150)KOUNT,ERROR1,ERROR2
  150 FORMAT(1X,I3,5X,F8.6,5X,F8.6)
  500 CONTINUE
      STOP
      END
```

```
C
C
C   THIS PROGRAM RANDOMLY SELECTS 300 SAMPLES AND CALCULATES
C   MSE, R EST., MEAN OF R, VAR. OF R, MEAN ERROR, VAR. OF
C   ERROR, MEAN MAXIMUM DIFFERENCE, AND VARIANCE OF THESE
C   MAXIMUM DIFFERENCES.  THIS PROCESS IS REPEATED AT SAMPLE
C   SIZES RANGING FROM 5% TO 50% IN 5% INCREMENTS.
C
C
      DIMENSION R(300),CSQYD(1000),SAMP1(1200),SAMP3(1200),PLACE1(1200),
     1PLACE2(1200),ERROR(300),IPER(1200),NARRAY(1000),NS(10)
      DIMENSION SAMP2(1200),S(20),SCORE(1000),TRPCT(20),PCT(20),
     1XMAX(300),DIF(20)
      REAL MEANC,MSE,MEANER,MEANMX
      REAL*8 DSEED,PLACE1,PLACE2
      INTEGER SAMP1,SAMP2,SCORE
      DSEED=123457.0D0
      READ(5,583)MEANC,NUMOBS,NS
  583 FORMAT(F4.0,I4,10I3)
      READ(5,593)(TRPCT(I),I=1,20)
  593 FORMAT(13F6.4/7F6.4)
      K=300
      DO 1002 I=1,NUMOBS
      READ(1,1001) PLACE1(I),PLACE2(I),SAMP1(I),SAMP2(I),SAMP3(I)
 1001 FORMAT(T13,A8,A6,19X,I1,3X,I3,9X,F5.2)
 1002 CONTINUE
      DO 500 IN=1,10
      N=NS(IN)
      TOT2=0.0
      TOT3=0.0
      TOT1=0.0
      TOT4=0.0
      TOT5=0.0
      TOT6=0.0
      TOT7=0.0
      DO 2000 J=1,K
      DO 329 I=1,20
  329 S(I)=0.0
      SUM1=0.0
      NN=0
C
C
C   GGPER IS AN IMSL SUBROUTINE WHICH GENERATES A RANDOM
C   PERMUTATION OF NUMOBS INTEGERS
C
C
      CALL GGPER (DSEED,NUMOBS,IPER)
C
C
C   THIS SECTION GENERATES A RANDOM SAMPLE OF SIZE NN
C
C
   80 DO 100 I=1,NUMOBS
      IF (I.EQ.1)  GO TO 90
      DO 110 KIS=1,NN
      IF (IPER(I) .EQ. NARRAY(KIS)) GO TO 100
  110 CONTINUE
   90 IF (SAMP1(IPER(I)) .LT. 3) GO TO 3000
      IF (IPER(I) .EQ. 1) GO TO 20
      IF (IPER(I) .EQ. NUMOBS) GO TO 40
```

```
      JJ=IPER(I)-1
      IF(PLACE1(IPER(I)) .EQ. PLACE1(JJ)) GO TO 60
      KK=IPER(I)+1
      IF(PLACE1(IPER(I)) .EQ. PLACE1(KK)) GO TO 70
 3000 NN=NN+1
      NARRAY(NN)=IPER(I)
      CSQYD(NN)=SAMP3(IPER(I))
      SCORE(NN)=SAMP2(IPER(I))
      IF(NN .GE. N) GO TO 1000
      GO TO 100
   20 KK=IPER(I)+1
      IF (PLACE1(IPER(I)) .EQ. PLACE1(KK)) GO TO 30
      GO TO 3000
   30 IF(PLACE2(IPER(I)).NE.PLACE2(KK)) GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      CSQYD(NN)=SAMP3(IPER(I))
      SCORE(NN)=SAMP2(IPER(I))
      NN=NN+1
      NARRAY(NN)=KK
      CSQYD(NN)=SAMP3(KK)
      SCORE(NN)=SAMP2(KK)
      IF(NN .GE.N) GO TO 1000
      GO TO 100
   40 JJ=IPER(I)-1
      IF(PLACE1(IPER(I)).EQ.PLACE1(JJ)) GO TO 50
      GO TO 3000
   50 IF (PLACE2(IPER(I)).NE.PLACE2(JJ)) GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      CSQYD(NN)=SAMP3(IPER(I))
      SCORE(NN)=SAMP2(IPER(I))
      NN=NN+1
      NARRAY(NN)=JJ
      CSQYD(NN)=SAMP3(JJ)
      SCORE(NN)=SAMP2(JJ)
      IF (NN.GE.N) GO TO 1000
      GO TO 100
   60 IF(PLACE2(IPER(I)).NE.PLACE2(JJ)) GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      CSQYD(NN)=SAMP3(IPER(I))
      SCORE(NN)=SAMP2(IPER(I))
      NN=NN+1
      NARRAY(NN)=JJ
      CSQYD(NN)=SAMP3(JJ)
      SCORE(NN)=SAMP2(JJ)
      IF(NN.GE.N) GO TO 1000
      GO TO 100
   70 IF(PLACE2(IPER(I)).NE.PLACE2(KK))GO TO 3000
      NN=NN+1
      NARRAY(NN)=IPER(I)
      CSQYD(NN)=SAMP3(IPER(I))
      SCORE(NN)=SAMP2(IPER(I))
      NN=NN+1
      NARRAY(NN)=KK
      CSQYD(NN)=SAMP3(KK)
      SCORE(NN)=SAMP2(KK)
      IF(NN .GE. N) GO TO 1000
  100 CONTINUE
```

```
C
C
C   THIS SECTION CALCULATES THE ABOVE NAMED STATISTICS AND
C   WRITES THEM TO A WYLBUR FILE
C
C
 1000 DO 2001 I=1,NN
      IF(SCORE(I).LE.5)S(1)=S(1)+1.
      IF(SCORE(I).GT.5 .AND. SCORE(I).LE.10)S(2)=S(2)+1.
      IF(SCORE(I).GT.10 .AND. SCORE(I).LE.15)S(3)=S(3)+1.
      IF(SCORE(I).GT.15 .AND. SCORE(I).LE.20)S(4)=S(4)+1.
      IF(SCORE(I).GT.20 .AND. SCORE(I).LE.25)S(5)=S(5)+1.
      IF(SCORE(I).GT.25 .AND. SCORE(I).LE.30)S(6)=S(6)+1.
      IF(SCORE(I).GT.30 .AND. SCORE(I).LE.35)S(7)=S(7)+1.
      IF(SCORE(I).GT.35 .AND. SCORE(I).LE.40)S(8)=S(8)+1.
      IF(SCORE(I).GT.40 .AND. SCORE(I).LE.45)S(9)=S(9)+1.
      IF(SCORE(I).GT.45 .AND. SCORE(I).LE.50)S(10)=S(10)+1.
      IF(SCORE(I).GT.50 .AND. SCORE(I).LE.55)S(11)=S(11)+1.
      IF(SCORE(I).GT.55 .AND. SCORE(I).LE.60)S(12)=S(12)+1.
      IF(SCORE(I).GT.60 .AND. SCORE(I).LE.65)S(13)=S(13)+1.
      IF(SCORE(I).GT.65 .AND. SCORE(I).LE.70)S(14)=S(14)+1.
      IF(SCORE(I).GT.70 .AND. SCORE(I).LE.75)S(15)=S(15)+1.
      IF(SCORE(I).GT.75 .AND. SCORE(I).LE.80)S(16)=S(16)+1.
      IF(SCORE(I).GT.80 .AND. SCORE(I).LE.85)S(17)=S(17)+1.
      IF(SCORE(I).GT.85 .AND. SCORE(I).LE.90)S(18)=S(18)+1.
      IF(SCORE(I).GT.90 .AND. SCORE(I).LE.95)S(19)=S(19)+1.
      IF(SCORE(I).GT.95)S(20)=S(20)+1.
      SUM1=SUM1+CSQYD(I)
 2001 CONTINUE
      R(J)=SUM1/NN
      TOT1=TOT1+R(J)
      TOT2=TOT2+((R(J)-MEANC)**2)
      ERROR(J)=(ABS(R(J)-MEANC))/MEANC
      PCT(1)=S(1)/NN
      DO 729 I=1,19
      KUT=I+1
      PCT(KUT)=S(KUT)/NN+PCT(I)
  729 CONTINUE
      DO 730 I=1,20
      DIF(I)=ABS(TRPCT(I)-PCT(I))
  730 CONTINUE
      XXMAX=-999.
      DO 731 I=1,20
      IF(DIF(I).GT.XXMAX)XXMAX=DIF(I)
  731 CONTINUE
      XMAX(J)=XXMAX
 2000 CONTINUE
      MSE=TOT2/K
      RMEAN=TOT1/K
      DO 2002 J=1,K
      TOT3=TOT3+((R(J)-RMEAN)**2)
      TOT4=TOT4+ERROR(J)
      TOT6=TOT6+XMAX(J)
 2002 CONTINUE
      VAR=TOT3/K
      MEANER=TOT4/K
      MEANMX=TOT6/K
      DO 2004 J=1,K
      TOT5=TOT5+((ERROR(J)-MEANER)**2)
      TOT7=TOT7+((XMAX(J)-MEANMX)**2)
```

```
2004 CONTINUE
     VARER=TOT5/K
     VARMAX=TOT7/K
     WRITE(3,2005)NN,MSE,RMEAN,VAR,MEANER,VARER,MEANMX,VARMAX
2005 FORMAT(I3,2X,7F10.4)
 500 CONTINUE
     STOP
     END
```

```
C
C
C   THIS PROGRAM USES THE STRATIFED BY DISTRICTS AND COUNTIES
C   RANDOM SAMPLING PLAN.  AFTER A RANDOM SAMPLE IS CHOSEN
C   FROM EVERY COUNTY, R EST. IS CALCULATED AS IF THE SAMPLE
C   WAS A SIMPLE RANDOM SAMPLE RATHER THAN STRATIFIED OVER
C   COUNTIES.  THIS PROCEDURE IS REPEATED 300 TIMES FOR ALL
C   SAMPLE SIZES (5%,10%,...50%).  THE STATISTICS COMPUTED
C   ARE MSE, R EST., MEAN OF R, VAR. OF R, MEAN ERROR, VAR.
C   OF ERROR, MEAN MAXIMUM DIFFERENCE, VAR. OF MAXIMUM DIFFERENCE.
C
C
      DIMENSION NUMOBS(20),PCT(20),TRPCT(20),PLACE1(20,100),
     1PLACE2(20,100),SAMP2(20,100),IPER(300),NARRAY(20,100),
     2NN(20),NC(10,20),CSQYD(20,100),R(300),ERROR(300),DIF(20),
     3XMAX(300)
      INTEGER SCORE(20,100),SAMP1(20,100),DESIGN(20,100)
      REAL MEANC,MSE,MEANER,MEANMX
      REAL*8 DSEED,PLACE1,PLACE2
      DSEED=123457.0D0
      KRUNS=300
      NUMCT=13
      READ (5,11)MEANC,(NUMOBS(K),K=1,NUMCT)
   11 FORMAT (F4.0,20I3)
      DO 101 NP=1,10
      READ (5,31)(NC(NP,K),K=1,NUMCT)
   31 FORMAT(20I3)
  101 CONTINUE
      READ (5,12)(TRPCT(I),I=1,20)
   12 FORMAT (13F6.4/7F6.4)
      DO 102 K=1,NUMCT
      NOBS=NUMOBS(K)
      DO 201 I=1,NOBS
      READ (1,21)PLACE1(K,I),PLACE2(K,I),DESIGN(K,I),SAMP1(K,I),
     1SAMP2(K,I)
   21 FORMAT(T13,A8,A6,19X,I1,3X,I3,9X,F5.2)
  201 CONTINUE
  102 CONTINUE
      DO 500 NP=1,10
      TOT1=0.0
      TOT2=0.0
      TOT3=0.0
      TOT4=0.0
      TOT5=0.0
      TOT6=0.0
      TOT7=0.0
      DO 400 KR=1,KRUNS
      ISUM2=0
      DO 254 KD=1,20
      NN(KD)=0
  254 CONTINUE
      SUM1=0.0
      KOUNT=0
      S1=0.0
      S2=0.0
      S3=0.0
      S4=0.0
      S5=0.0
      S6=0.0
      S7=0.0
```

```
      S8=0.0
      S9=0.0
      S10=0.0
      S11=0.0
      S12=0.0
      S13=0.0
      S14=0.0
      S15=0.0
      S16=0.0
      S17=0.0
      S18=0.0
      S19=0.0
      S20=0.0
      DO 300 K=1,NUMCT
C
C
C   GGPER IS AN IMSL SUBROUTINE WHICH GENERATES A RANDOM SAMPLE
C   OF NUMOBS(K) INTEGERS
C
C
      CALL GGPER (DSEED,NUMOBS(K),IPER)
C
C
C   THIS SECTION GENERATES A RANDOM SAMPLE OF SIZE NN FROM COUNTY K
C
C
      NOBS=NUMOBS(K)
   80 DO 100 I=1,NOBS
      IF (I.EQ.1) GO TO 90
      NNN=NN(K)
      DO 110 KIS=1,NNN
      IF (IPER(I).EQ.NARRAY(K,KIS))GO TO 100
  110 CONTINUE
   90 IF(DESIGN(K,IPER(I)).LT.3) GO TO 3000
      IF(IPER(I) .EQ. 1) GO TO 20
      IF(IPER(I) .EQ.NUMOBS(K)) GO TO 40
      JJ=IPER(I)-1
      IF(PLACE1(K,IPER(I)) .EQ. PLACE2(K,JJ)) GO TO 60
      KK=IPER(I)+1
      IF(PLACE1(K,IPER(I)) .EQ. PLACE2(K,KK)) GO TO 70
 3000 NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      IF (NN(K).GE.NC(NP,K)) GO TO 300
  100 CONTINUE
   20 KK=IPER(I)+1
      IF(PLACE1(K,IPER(I)).EQ.PLACE1(K,KK)) GO TO 30
      GO TO 3000
   30 IF(PLACE2(K,IPER(I)).NE.PLACE2(K,KK)) GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=KK
      CSQYD(K,NN(K))=SAMP2(K,KK)
```

```
      SCORE(K,NN(K))=SAMP1(K,KK)
      IF(NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   40 JJ=IPER(I)-1
      IF(PLACE1(K,IPER(I)).EQ.PLACE1(K,JJ)) GO TO 50
      GO TO 3000
   50 IF(PLACE2(K,IPER(I)).NE.PLACE2(K,JJ)) GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=JJ
      CSQYD(K,NN(K))=SAMP2(K,JJ)
      SCORE(K,NN(K))=SAMP1(K,JJ)
      IF(NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   60 IF(PLACE2(K,IPER(I)).NE.PLACE2(K,JJ))GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=JJ
      CSQYD(K,NN(K))=SAMP2(K,JJ)
      SCORE(K,NN(K))=SAMP1(K,JJ)
      IF (NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   70 IF (PLACE2(K,IPER(I)).NE.PLACE2(K,KK)) GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=KK
      CSQYD(K,NN(K))=SAMP2(K,KK)
      SCORE(K,NN(K))=SAMP1(K,KK)
      IF(NN(K) .GE. NC(NP,K))GO TO 300
      GO TO 100
  300 CONTINUE
C
C
C   THIS SECTION CALCULATES THE ABOVE NAMED STATISTICS AND
C   WRITES THEM TO A WYLBUR FILE
C
C
      DO 2001 K=1,NUMCT
      NNN=NN(K)
      DO 2002 I=1,NNN
      SUM1=SUM1+CSQYD(K,I)
      IF(SCORE(K,I).LE.5) GO TO 900
      IF(SCORE(K,I).GT.5 .AND. SCORE(K,I).LE.10) GO TO 901
      IF(SCORE(K,I).GT.10 .AND. SCORE(K,I).LE.15) GO TO 902
      IF(SCORE(K,I).GT.15 .AND. SCORE(K,I).LE.20) GO TO 903
```

```
     IF(SCORE(K,I).GT.20 .AND. SCORE(K,I).LE.25) GO TO 904
     IF(SCORE(K,I).GT.25 .AND. SCORE(K,I).LE.30) GO TO 905
     IF(SCORE(K,I).GT.30 .AND. SCORE(K,I).LE.35) GO TO 906
     IF(SCORE(K,I).GT.35 .AND. SCORE(K,I).LE.40) GO TO 907
     IF(SCORE(K,I).GT.40 .AND. SCORE(K,I).LE.45) GO TO 908
     IF(SCORE(K,I).GT.45 .AND. SCORE(K,I).LE.50) GO TO 909
     IF(SCORE(K,I).GT.50 .AND. SCORE(K,I).LE.55) GO TO 910
     IF(SCORE(K,I).GT.55 .AND. SCORE(K,I).LE.60) GO TO 911
     IF(SCORE(K,I).GT.60 .AND. SCORE(K,I).LE.65) GO TO 912
     IF(SCORE(K,I).GT.65 .AND. SCORE(K,I).LE.70) GO TO 913
     IF(SCORE(K,I).GT.70 .AND. SCORE(K,I).LE.75) GO TO 914
     IF(SCORE(K,I).GT.75 .AND. SCORE(K,I).LE.80) GO TO 915
     IF(SCORE(K,I).GT.80 .AND. SCORE(K,I).LE.85) GO TO 916
     IF(SCORE(K,I).GT.85 .AND. SCORE(K,I).LE.90) GO TO 917
     IF(SCORE(K,I).GT.90 .AND. SCORE(K,I).LE.95) GO TO 918
     IF(SCORE(K,I).GT.95) GO TO 919
 900 S1=S1+1.
     GO TO 2002
 901 S2=S2+1.
     GO TO 2002
 902 S3=S3+1.
     GO TO 2002
 903 S4=S4+1.
     GO TO 2002
 904 S5=S5+1.
     GO TO 2002
 905 S6=S6+1.
     GO TO 2002
 906 S7=S7+1.
     GO TO 2002
 907 S8=S8+1.
     GO TO 2002
 908 S9=S9+1.
     GO TO 2002
 909 S10=S10+1.
     GO TO 2002
 910 S11=S11+1.
     GO TO 2002
 911 S12=S12+1.
     GO TO 2002
 912 S13=S13+1.
     GO TO 2002
 913 S14=S14+1.
     GO TO 2002
 914 S15=S15+1.
     GO TO 2002
 915 S16=S16+1.
     GO TO 2002
 916 S17=S17+1.
     GO TO 2002
 917 S18=S18+1.
     GO TO 2002
 918 S19=S19+1.
     GO TO 2002
 919 S20=S20+1.
2002 CONTINUE
2001 CONTINUE
     R(KR)=SUM1/KOUNT
     TOT1=TOT1+R(KR)
     TOT2=TOT2+((R(KR)-MEANC)**2)
```

```
      ERROR(KR)=(ABS(R(KR)-MEANC))/MEANC
      IF (NP .NE.10) GO TO 920
      WRITE(2,2003)ERROR(KR)
 2003 FORMAT(F8.4)
  920 PCT(1)=S1/KOUNT
      PCT(2)=S2/KOUNT+PCT(1)
      PCT(3)=S3/KOUNT+PCT(2)
      PCT(4)=S4/KOUNT+PCT(3)
      PCT(5)=S5/KOUNT+PCT(4)
      PCT(6)=S6/KOUNT+PCT(5)
      PCT(7)=S7/KOUNT+PCT(6)
      PCT(8)=S8/KOUNT+PCT(7)
      PCT(9)=S9/KOUNT+PCT(8)
      PCT(10)=S10/KOUNT+PCT(9)
      PCT(11)=S11/KOUNT+PCT(10)
      PCT(12)=S12/KOUNT+PCT(11)
      PCT(13)=S13/KOUNT+PCT(12)
      PCT(14)=S14/KOUNT+PCT(13)
      PCT(15)=S15/KOUNT+PCT(14)
      PCT(16)=S16/KOUNT+PCT(15)
      PCT(17)=S17/KOUNT+PCT(16)
      PCT(18)=S18/KOUNT+PCT(17)
      PCT(19)=S19/KOUNT+PCT(18)
      PCT(20)=S20/KOUNT+PCT(19)
      DO 930 I=1,20
      DIF(I)=ABS(TRPCT(I)-PCT(I))
  930 CONTINUE
      XXMAX=-999.
      DO 940 I=1,20
      IF(DIF(I).GT.XXMAX)XXMAX=DIF(I)
  940 CONTINUE
      XMAX(KR)=XXMAX
      IF (NP .NE.10) GO TO 400
      WRITE(3,2004)XMAX(KR)
 2004 FORMAT(F8.4)
  400 CONTINUE
      MSE=TOT2/KRUNS
      RMEAN=TOT1/KRUNS
      DO 941 KR=1,KRUNS
      TOT3=TOT3+((R(KR)-RMEAN)**2)
      TOT4=TOT4+ERROR(KR)
      TOT6=TOT6+XMAX(KR)
  941 CONTINUE
      VAR=TOT3/KRUNS
      MEANER=TOT4/KRUNS
      MEANMX=TOT6/KRUNS
      DO 942 KR=1,KRUNS
      TOT5=TOT5+((ERROR(KR)-MEANER)**2)
      TOT7=TOT7+((XMAX(KR)-MEANMX)**2)
  942 CONTINUE
      VARER=TOT5/KRUNS
      VARMAX=TOT7/KRUNS
      WRITE(4,943)KOUNT,MSE,RMEAN,VAR,MEANER,VARER,MEANMX,VARMAX
  943 FORMAT(I3,2X,7F10.5)
  500 CONTINUE
      STOP
      END
```

```
C
C
C   THIS PROGRAM USES THE STRATIFIED BY DISTRICTS AND COUNTIES
C   RANDOM SAMPLING PLAN.  AFTER A RANDOM SAMPLE IS CHOSEN
C   FROM EVERY COUNTY, A WEIGHTED AVERAGE OF THE COUNTY MEANS IS
C   CALCULATED.  (ESTIMATE OF COST AND SCORE DISTRIBUTION)  THIS
C   PROCEDURE IS REPEATED 300 TIMES FOR ALL SAMPLE SIZES (5%,10%,...
C   50%).  THE MSE, R EST., MEAN OF R, VAR. OF R, MEAN ERROR,
C   VAR. OF ERROR, MEAN MAXIMUM DIFFERENCE AND VAR. OF MAXIMUM
C   DIFFERENCE IS THEN CALCULATED.
C
C
      DIMENSION NUMOBS(20),PCT(20),TRPCT(20),PLACE1(20,100),
     1PLACE2(20,100),SAMP2(20,100),IPER(300),NARRAY(20,100),
     2NN(20),NC(10,20),CSQYD(20,100),R(300),ERROR(300),DIF(20),
     3XMAX(300)
      DIMENSION SS(20),S(20,20),P(20,20),RAT(20),SUM1(20)
      INTEGER SCORE(20,100),SAMP1(20,100),DESIGN(20,100)
      REAL MEANC,MSE,MEANER,MEANMX
      REAL*8 DSEED,PLACE1,PLACE2
      DSEED=123457.0D0
      KRUNS=300
      NDIS=270
      NUMCT=12
      READ (5,11)MEANC,(NUMOBS(K),K=1,NUMCT)
   11 FORMAT (F4.0,20I3)
      DO 101 NP=1,10
      READ (5,31)(NC(NP,K),K=1,NUMCT)
   31 FORMAT(20I3)
  101 CONTINUE
      READ (5,12)(TRPCT(I),I=1,20)
   12 FORMAT (13F6.4/7F6.4)
      DO 102 K=1,NUMCT
      NOBS=NUMOBS(K)
      DO 201 I=1,NOBS
      READ (1,21)PLACE1(K,I),PLACE2(K,I),DESIGN(K,I),SAMP1(K,I),
     1SAMP2(K,I)
   21 FORMAT(T13,A8,A6,19X,I1,3X,I3,9X,F5.2)
  201 CONTINUE
  102 CONTINUE
      DO 500 NP=1,10
      TOT1=0.0
      TOT2=0.0
      TOT3=0.0
      TOT4=0.0
      TOT5=0.0
      TOT6=0.0
      TOT7=0.0
      DO 400 KR=1,KRUNS
      DO 254 KD=1,20
      NN(KD)=0
  254 CONTINUE
      SUM2=0.0
      KOUNT=0
      DO 910 I=1,20
      DO 911 K=1,NUMCT
      S(I,K)=0.0
  911 CONTINUE
  910 CONTINUE
      DO 912 K=1,NUMCT
```

```
  912 SUM1(K)=0.0
      DO 913 I=1,20
  913 SS(I)=0.0
      DO 300 K=1,NUMCT
C
C
C   GGPER IS AN IMSL SUBROUTINE WHICH GENERATES A RANDOM SAMPLE OF
C   NUMOBS(K) INTEGERS
C
C
      CALL GGPER (DSEED,NUMOBS(K),IPER)
C
C
C   THIS SECTION GENERATES A RANDOM SAMPLE OF SIZE NN FROM COUNTY K
C
C
      NOBS=NUMOBS(K)
   80 DO 100 I=1,NOBS
      IF (I.EQ.1) GO TO 90
      NNN=NN(K)
      DO 110 KIS=1,NNN
      IF (IPER(I).EQ.NARRAY(K,KIS))GO TO 100
  110 CONTINUE
   90 IF(DESIGN(K,IPER(I)).LT.3) GO TO 3000
      IF(IPER(I) .EQ. 1) GO TO 20
      IF(IPER(I) .EQ.NUMOBS(K)) GO TO 40
      JJ=IPER(I)-1
      IF(PLACE1(K,IPER(I)) .EQ. PLACE2(K,JJ)) GO TO 60
      KK=IPER(I)+1
      IF(PLACE1(K,IPER(I)) .EQ. PLACE2(K,KK)) GO TO 70
 3000 NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      IF (NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   20 KK=IPER(I)+1
      IF(PLACE1(K,IPER(I)).EQ.PLACE1(K,KK)) GO TO 30
      GO TO 3000
   30 IF(PLACE2(K,IPER(I)).NE.PLACE2(K,KK)) GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=KK
      CSQYD(K,NN(K))=SAMP2(K,KK)
      SCORE(K,NN(K))=SAMP1(K,KK)
      IF(NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   40 JJ=IPER(I)-1
      IF(PLACE1(K,IPER(I)).EQ.PLACE1(K,JJ)) GO TO 50
      GO TO 3000
   50 IF(PLACE2(K,IPER(I)).NE.PLACE2(K,JJ)) GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
```

```
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=JJ
      CSQYD(K,NN(K))=SAMP2(K,JJ)
      SCORE(K,NN(K))=SAMP1(K,JJ)
      IF(NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   60 IF(PLACE2(K,IPER(I)).NE.PLACE2(K,JJ))GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=JJ
      CSQYD(K,NN(K))=SAMP2(K,JJ)
      SCORE(K,NN(K))=SAMP1(K,JJ)
      IF (NN(K).GE.NC(NP,K)) GO TO 300
      GO TO 100
   70 IF (PLACE2(K,IPER(I)).NE.PLACE2(K,KK)) GO TO 3000
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=IPER(I)
      CSQYD(K,NN(K))=SAMP2(K,IPER(I))
      SCORE(K,NN(K))=SAMP1(K,IPER(I))
      NN(K)=NN(K)+1
      KOUNT=KOUNT+1
      NARRAY(K,NN(K))=KK
      CSQYD(K,NN(K))=SAMP2(K,KK)
      SCORE(K,NN(K))=SAMP1(K,KK)
      IF(NN(K) .GE. NC(NP,K))GO TO 300
  100 CONTINUE
  300 CONTINUE
C
C
C   THIS SECTION CALCULATES THE ABOVE NAMED STATISTICS AND WRITES
C   THEM TO A WYLBUR FILE
C
C
      DO 2001 K=1,NUMCT
      NNN=NN(K)
      DO 2002 I=1,NNN
      IF(SCORE(K,I).LE.5) S(1,K)=S(1,K)+1.
      IF(SCORE(K,I).GT.5 .AND. SCORE(K,I).LE.10) S(2,K)=S(2,K)+1.
      IF(SCORE(K,I).GT.10 .AND. SCORE(K,I).LE.15) S(3,K)=S(3,K)+1.
      IF(SCORE(K,I).GT.15 .AND. SCORE(K,I).LE.20) S(4,K)=S(4,K)+1.
      IF(SCORE(K,I).GT.20 .AND. SCORE(K,I).LE.25) S(5,K)=S(5,K)+1.
      IF(SCORE(K,I).GT.25 .AND. SCORE(K,I).LE.30) S(6,K)=S(6,K)+1.
      IF(SCORE(K,I).GT.30 .AND. SCORE(K,I).LE.35) S(7,K)=S(7,K)+1.
      IF(SCORE(K,I).GT.35 .AND. SCORE(K,I).LE.40) S(8,K)=S(8,K)+1.
      IF(SCORE(K,I).GT.40 .AND. SCORE(K,I).LE.45) S(9,K)=S(9,K)+1.
      IF(SCORE(K,I).GT.45 .AND. SCORE(K,I).LE.50) S(10,K)=S(10,K)+1.
      IF(SCORE(K,I).GT.50 .AND. SCORE(K,I).LE.55) S(11,K)=S(11,K)+1.
      IF(SCORE(K,I).GT.55 .AND. SCORE(K,I).LE.60) S(12,K)=S(12,K)+1.
      IF(SCORE(K,I).GT.60 .AND. SCORE(K,I).LE.65) S(13,K)=S(13,K)+1.
      IF(SCORE(K,I).GT.65 .AND. SCORE(K,I).LE.70) S(14,K)=S(14,K)+1.
      IF(SCORE(K,I).GT.70 .AND. SCORE(K,I).LE.75) S(15,K)=S(15,K)+1.
```

```
      IF(SCORE(K,I).GT.75 .AND. SCORE(K,I).LE.80) S(16,K)=S(16,K)+1.
      IF(SCORE(K,I).GT.80 .AND. SCORE(K,I).LE.85) S(17,K)=S(17,K)+1.
      IF(SCORE(K,I).GT.85 .AND. SCORE(K,I).LE.90) S(18,K)=S(18,K)+1.
      IF(SCORE(K,I).GT.90 .AND. SCORE(K,I).LE.95) S(19,K)=S(19,K)+1.
      IF(SCORE(K,I).GT.95) S(20,K)=S(20,K)+1.
      SUM1(K)=SUM1(K)+CSQYD(K,I)
 2002 CONTINUE
      P(1,K)=S(1,K)/NN(K)
      DO 901 KUT=1,19
      LLL=KUT+1
      P(LLL,K)=S(LLL,K)/NN(K)+P(KUT,K)
  901 CONTINUE
      RAT(K)=SUM1(K)/NN(K)
 2001 CONTINUE
      DO 905 K=1,NUMCT
  905 SUM2=SUM2+NUMOBS(K)*RAT(K)
      DO 902 I=1,20
      DO 903 K=1,NUMCT
      SS(I)=SS(I)+(NUMOBS(K)*P(I,K))
  903 CONTINUE
  902 CONTINUE
      R(KR)=SUM2/NDIS
      DO 904 I=1,20
  904 PCT(I)=SS(I)/NDIS
      TOT1=TOT1+R(KR)
      TOT2=TOT2+((R(KR)-MEANC)**2)
      ERROR(KR)=(ABS(R(KR)-MEANC))/MEANC
      DO 930 I=1,20
      DIF(I)=ABS(TRPCT(I)-PCT(I))
  930 CONTINUE
      XXMAX=-999.
      DO 940 I=1,20
      IF(DIF(I).GT.XXMAX)XXMAX=DIF(I)
  940 CONTINUE
      XMAX(KR)=XXMAX
  400 CONTINUE
      MSE=TOT2/KRUNS
      RMEAN=TOT1/KRUNS
      DO 941 KR=1,KRUNS
      TOT3=TOT3+((R(KR)-RMEAN)**2)
      TOT4=TOT4+ERROR(KR)
      TOT6=TOT6+XMAX(KR)
  941 CONTINUE
      VAR=TOT3/KRUNS
      MEANER=TOT4/KRUNS
      MEANMX=TOT6/KRUNS
      DO 942 KR=1,KRUNS
      TOT5=TOT5+((ERROR(KR)-MEANER)**2)
      TOT7=TOT7+((XMAX(KR)-MEANMX)**2)
  942 CONTINUE
      VARER=TOT5/KRUNS
      VARMAX=TOT7/KRUNS
      WRITE(4,943)KOUNT,MSE,RMEAN,VAR,MEANER,VARER,MEANMX,VARMAX
  943 FORMAT(I3,2X,7F10.5)
  500 CONTINUE
      STOP
      END
```

```
C
C
C    THIS PROGRAM USES SYSTEMATIC SAMPLING.  EVERY POSSIBLE SAMPLE IS
C    DRAWN AND THE MSE, R EST., MEAN OF R, VAR. OF R, MEAN ERROR, AND
C    VAR. OF ERROR IS CALCULATED.  THIS PROCESS IS REPEATED AT SAMPLE
C    SIZES RANGING FROM 5% TO 50% AT 5% INCREMENTS.
C
C
      DIMENSION NUMOBS(11),MEANC(11),CSQYD(1200),NSAMP(10)
      DIMENSION IN(10),SAMP(1200),R(20),ERROR(20)
      REAL MEANC,MSE,MEANER,IN
      READ(5,1)(NUMOBS(I),I=1,11)
    1 FORMAT(11I4)
      READ(5,2)(MEANC(I),I=1,11)
    2 FORMAT(11F4.2)
      DO 500 K=1,11
      L=K+9
      NUM=NUMOBS(K)
      READ(L,3)(SAMP(I),I=1,NUM)
    3 FORMAT(T62,F5.2)
      READ(5,4)(NSAMP(I),I=1,10)
    4 FORMAT(10I3)
      READ(5,5)(IN(I),I=1,10)
    5 FORMAT(10F5.2)
      DO 300 I=1,10
      TOT1=0.0
      TOT2=0.0
      TOT3=0.0
      TOT4=0.0
      TOT5=0.0
      INT=IN(I)
      AA=IN(I)
      NS=NSAMP(I)
C
C
C    THIS SECTION SELECTS EVERY POSSIBLE SAMPLE
C
C
      DO 20 KIS=1,INT
      SUM1=0.0
      DO 10 NN=1,NS
      KK=NN-1
      II=KIS+(KK*AA)
      CSQYD(NN)=SAMP(II)
   10 CONTINUE
      DO 2000 M=1,NS
 2000 SUM1=SUM1+CSQYD(M)
      R(KIS)=SUM1/NS
      TOT1=TOT1+R(KIS)
      TOT2=TOT2+((R(KIS)-MEANC(K))**2)
      ERROR(KIS)=(ABS(R(KIS)-MEANC(K)))/MEANC(K)
   20 CONTINUE
C
C
C    THIS SECTION CALCULATES THE ABOVE NAMED STATISTICS
C
C
      MSE=TOT2/INT
      RMEAN=TOT1/INT
      DO 2001 M=1,INT
      TOT3=TOT3+((R(M)-RMEAN)**2)
```

```
      TOT4=TOT4+ERROR(M)
 2001 CONTINUE
      VAR=TOT3/INT
      MEANER=TOT4/INT
      DO 2002 M=1,INT
      TOT5=TOT5+((ERROR(M)-MEANER)**2)
 2002 CONTINUE
      VARER=TOT5/INT
      WRITE(6,2003)NS,MSE,RMEAN,VAR,MEANER,VARER
 2003 FORMAT(1X,I3,2X,5F13.4)
  300 CONTINUE
  500 CONTINUE
      STOP
      END
```