NETWORK SAMPLING TO ESTIMATE DISTRIBUTION OF PAVEMENT CONDITION AND COSTS

by

J. Templeton, R. L. Lytton, and A. Garcia-Diaz

Texas Transportation Institute

Texas A&M University System

College Station, Texas

Prepared for presentation at

the Sixty-Fourth Annual

Transportation Research Board Meeting

to be held in

Washington, D.C., January 1985

August 1984

ABSTRACT

Surveying the pavement condition of a highway network by sampling is done to obtain current information that is accurate enough for the purposes of planning and funding needs estimates. Sampling is used to reduce the amount of time and manpower that is required to collect this information to an irreducible minimum. In this paper are the results of a study of data from a 1982 survey of the total mileage of three District networks in Texas to determine the answers to several questions concerning the accuracy that can be achieved with different types of sampling and sizes of sample.

For example, it is found that a Beta distribution fits the cumulative distribution of pavement scores very well; that the two parameters of the Beta distribution can be estimated with good accuracy using a 5-percent sample; and that the Beta distribution derived in this way is capable of accurately estimating the percent of pavements that are below a given score.

In addition, it is found that the size of sample required to estimate the distribution of pavement scores to a given degree of accuracy is smaller than the sample size that must be used to estimate the average cost of maintenance and rehabilitation to the same level of accuracy. The relative sizes found are presented in tables and figures and differ between classes of highway: Interstate, U.S., State, and Farm-to-Market.

Several convenient relations were found between the mean pavement

score and the variance of pavement scores, percent of pavements needing no repairs, average costs per square yard, and percent error in the estimated average costs. It is not surprising to find that the average costs are reduced and the percent error is increased as the mean pavement score increases. The mean score is easy to obtain accurately with a small 5-percent sample.

The results of this study give significant information that will be useful in planning future sampling surveys both in Texas and elsewhere.

INTRODUCTION

An essential part of pavement management at the State and District network level is to be able to estimate accurately the mean and the distribution of pavement conditions and the costs to maintain and rehabilitate the pavement network. Because of a limitation of funds, time, equipment, and manpower this kind of information can be collected efficiently by using sampling surveys of the condition of the network. This paper reports on the results of the analysis of data collected on 100 percent of the 2 mile-long pavement sections in three Districts of the Texas State Department of Highways and Public Transportation, each of which is responsible for between 2500 and 3000 miles of pavement of all functional classes. The objective of the study was to determine the answers to several questions about the minimum size of the sample and method of sampling required to obtain estimates of costs and pavement condition that are accurate enough for the purposes of planning and funding needs estimates. The major questions are:

1. Is it possible to estimate the cumulative probability distribution of pavement scores accurately with a 5 percent sample?
2. What kind of probability density function best fits the distribiution of pavement scores?
3. How accurately can the number of pavement sections with pavement scores below a minimum acceptable level be

calculated using a statistical distribution derived from a 5 percent sample?

4. What size of sample is required in order to get an accurate estimate of the distribution of pavement scores?
5. What size of sample is needed in order to accurately estimate the average cost of maintenance and rehabilitation?
6. Are there any overall relations between the mean pavement score, which can be determined accurately with a small sample, and the number of pavement sections that are not in need of any maintenance or rehabilitation?

In short, the study attempted to find some basic information and rules of thumb which could be supported by the data and which could assist in planning future condition surveys so as to minimize the effort spent and to maximize the accuracy of the resulting information as much as possible.

PAVEMENT CONDITION SURVEYS

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 on 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 included four roadways (two main roadways and two frontage roads). For the purpose of analysis of the survey data, only the main roadways were considered. One lane of each roadway within the selected segment was sampled and each of these observations were 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. Figure 1(b) shows a two lane highway segment. Only one observation is chosen from this segment.

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 shows the location of each of these districts.

For each observation, in both the 5% and 100% samples, the serviceability index was determined with the May's Ride Meter and a Visual Defects Rating was performed. In the visual rating, the following seven distress types were recorded: rutting, raveling, flushing, alligator cracking, longitudinal cracking, and failures (or "potholes"). For each, the rater noted the area covered (e.g., for alligator cracking - 0%, 1-10%, 11-25%, or >25%).

In Texas, the pavement score is defined as follows:

$$\text{Pavement Score} = (U_{ride} \times U_{distress}^{a_1}) \times 100$$

where

U_{ride} = the riding quality utility score of range 0-1

$U_{distress}$ = the visual distress utility score of range 0-1

a_1 = weighting factor for the distress utility score. It is defined as

$$a_1 = \frac{1}{ADTF \times EALF}$$

The ADTF and EALF are obtained from Tables 1 and 2 and are dependent upon the sections ADT per lane and 18-kip Equivalent Single Axle Loads.

The U_{ride} value is read from one of the curves shown in Figure 3. These curves are functions of ADT per lane x Average Vehicle Speed. Therefore the Utility Value for a high volume Interstate will be different from that for a low-volume rural highway even though both highways have the same riding quality.

The $U_{distress}$ is further defined as:

$$U_{distress} = (U_{rut})^{b_1} (U_{ravel})^{b_2} (U_{flush})^{b_3} (U_{failures})^{b_4} (U_{allig})^{b_5} (U_{long})^{b_6} (U_{tran})^{b_7}$$

where each U_i is based on the area of that distress type in the section. The individual U_i values are shown below.

Area* Covered	U_{rut}	U_{ravel}	U_{flush}	$U_{failures}$	U_{allig}	U_{long}	U_{tran}
Low	.840	.975	.992	.811	.890	.970	.983
Medium	.737	.781	.832	.313	.685	.720	.849
High	.721	.714	.754	.198	.400	.616	.801

*The low, medium, and high areas are defined for each distress type.

The b coefficients are environmental weighting factors, and they are determined by the following relationships with Rainfall Factor (RF) and Freeze-Thaw Factor (FF):

$$b_1 = \frac{1}{RF} \quad \text{Rutting}$$

$$b_2 = 1 \quad \text{Raveling}$$

$$b_3 = \frac{1}{RF} \quad \text{Flushing}$$

$$b_4 = b_5 = b_6 = b_7 = \frac{1}{RF \times FF} \quad \text{(all other distresses)}$$

The Rainfall and Freeze-Thaw Factors are obtained from Tables 3 and 4.

By using this definition of Pavement Score, if any single utility becomes low, then the pavement score will be low. For instance, if the highway's Ride value reaches a critical level even with no distress, the section's pavement score will drop to a failure level.

ESTIMATING COSTS OF MAJOR MAINTENANCE AND REHABILITATION

In evaluating the results of the Statewide condition survey, each pavement section with a Pavement Score less than 40 was considered to be in need of major maintenance or rehabilitation. Five funding strategies were considered in making the estimate of total costs, and one of these was selected for each pavement below the specified minimum:

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.) or its equivalent in reconstruction.

The selection of the appropriate strategy was made in the following way.

For each of the above strategies, the estimated rehabilitated Pavement Score was computed and deterioration calculations were made to determine the life expectancy. This expected life was compared to a minimum allowable expected life of 3 to 5 years depending upon the class of highway to determine which of the five strategies had the smallest allowable expected life, and that one was chosen as the strategy to be used. The selected strategy was assumed to be applied to the entire 2-mile section. Average costs were determined for each roadway class in each of the 100 percent-sampled districts. The number of pavement sections, the average costs, mean pavement scores, variance of pavement scores, and percentage of pavement sections with no costs are tabulated in Table 5. It is noted that District 11 has no Interstate highway mileage. When these results are plotted against the mean pavement score, very clear trends emerge. For example, in Figure 4, the relations between the variance of pavement score and the mean pavement score for the different highway systems is shown. Knowing this relation and the probability density function for pavement scores, it is possible to construct an accurate pavement score distribution once a good estimate of the mean score is determined. In Figure 5, the relations between the mean pavement score and the percent of pavements with zero costs is shown. In addition to the fact that the trends are again very clear, it is significant that the Farm-to-Market system curve is above the State curve which, in turn, is above the U.S. and Interstate highway curve. In Figure 6, the average costs per square yard are plotted against the mean pavement score to show that even these costs can be estimated once a good estimate of the mean pavement score is in hand. In this case, the average costs rise from the Farm-to-Market upward to the Interstate highway system.

These results are encouraging because they illustrate the clear trends that exist among the data that have been collected in the 100 percent-sampled districts. If a 5-percent sample can be used to obtain an accurate estimate of the mean pavement score, then it may be possible to use the relations in Figures 4, 5, and 6 to make accurate estimates of the condition and costs of rehabilitating pavement networks.

How large of a sample is necessary in order to make accurate estimates of these quantities? Is a five-percent sample large enough? What is the best way to organize a sample survey in order to obtain the most accurate cost information for the effort? These are the questions that were asked at the beginning of this paper. They arise naturally from considering the information in Figures 4, 5, and 6. The answers are determined by simulating different sizes and types of surveys using the data from the 100-percent sampled districts.

ACCURACY OF THE FIVE PERCENT SAMPLE IN ESTIMATING NETWORK PAVEMENT SCORE DISTRIBUTIONS

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. Cumulative Pavement Score distributions or histograms ranging from 0 to 100 were classified into twenty class intervals, each with a width of five. A comparison between the 100 percent sample histogram and the corresponding five percent histogram shows that the five percent histogram more closely resembles the 100 percent histogram in those data groups with a larger number of observations. Table 6 shows the maximum absolute difference between the two histograms for each data group along with the number of observations in the five percent sample. The maximum difference ranges between 0.0601 and 0.1769.

CONFIDENCE BANDS ON THE TRUE CUMULATIVE DISTRIBUTION

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 (2) is used. This percentile along with an empirical cumulative distribution, $S_n(X)$, can be used to form a $(1-\alpha)$ percent confidence band for a true cumulative distribution $F(X)$. As the percent of confidence is increased, the band becomes wider. In order to form this band, the $(1-\alpha)$ percentile of the Kolmogorov-Smirnov test statistic must be found. This percentile can be defined as:

$$K_{(1-\alpha)} = \frac{K_T(1-\alpha)}{\sqrt{n}} \qquad (1)$$

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:

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

and

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

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 7 shows graphically how the band is constructed. Figure 8 shows one of the better comparisons between the true 100 percent histograms $F(X)$, and the confidence bands generated with using a five percent sample, $S_n(X)$. The figure shows the comparison for Farm-to-Market roads in District 11.

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 (2), 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.

PROBABILITY DENSITY FUNCTION THAT FITS THE PAVEMENT SCORE DISTRIBUTION

If a probability density function is fitted to the histogram of a 5 percent data sample, it is possible to estimate the number of pavement sections with a Pavement Score below 40. An investigation was made into the accuracy of this procedure. The Beta distribution was chosen as the family of density functions to be used. The probability density function is defined as follows:

$$f(X) = \frac{1}{B(a,b)} X^{a-1} (1-X)^{b-1} \text{ for } a,b>0;\ 0<X<1 \qquad (4)$$

In Equation (4) B(a,b) is defined as

$$B(a,b) = \int_0^1 X^{a-1} (1-X)^{b-1}\, dx \qquad (5)$$

Since the random variable, X, must be in the interval zero to one, all of 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

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

and

$$\sigma^2 = E[(X-\mu)^2] \text{ to } S^2 = \sum_{i=1}^{n} (X_i-\bar{X})^{2/n} \qquad (7)$$

In the Beta distribution,

$$\mu = \frac{a}{a+b} \tag{8}$$

and

$$\sigma^2 = \frac{ab}{(a+b+1)(a+b)^2} \tag{9}$$

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:

$$\hat{a} = \frac{\hat{\mu}^2(1-\hat{\mu})}{\hat{\sigma}^2} - \hat{\mu} \tag{10}$$

$$\hat{b} = \frac{\hat{a} - \hat{a}\hat{\mu}}{\hat{\mu}} \tag{11}$$

Table 7 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 (3), $\hat{a}$ and $\hat{b}$ are consistent estimators of a and b.

A Kolmogorov-Smirnov goodness of fit test (2) was used to determine if the 100 percent samples do actually come from a Beta distribution with the a and b parameters given in Table 7. The tests showed that the samples did come from a Beta distribution. As a matter of interest, Figure 9 shows how the Beta distribution parameter a varies with the percent of pavements with zero costs for each of the highway systems. Figure 10 is a graph of the b parameter's relation to the percent of pavements with zero costs. Only the State highway system fails to generate a smooth curved relation.

ACCURACY OF ESTIMATING THE PERCENT OF PAVEMENTS BELOW A MINIMUM ACCEPTABLE SCORE

Two methods of predicting the percentage of pavements with a Pavement Score below 40 were used and then compared. In the first method the percentage of roads falling at or below 40 was computed directly 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 (10) and Equation (11), a and b are then estimated. Using the International Mathematical and Statistical Libraries, Inc. (IMSL) subroutine, MDBETA, 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.

Table 8 shows the mean error for each of the two methods used for each data group. As can be seen from the table, the mean error is less for every data group when the percentage is calculated using the Beta distribution. This method is, therefore, recommended for use with all such sample surveys.

SAMPLE SIZE TO GET A SPECIFIED ACCURACY IN THE ESTIMATE OF PAVEMENT SCORE DISTRIBUTION

How large of a sample size is needed to obtain a consistent accuracy in estimating the Pavement Score cumulative distribution of 5 percent, or 10 percent, or some other specified amount? This question was studied using the simulation program referred to in the previous section. Random selection of pavement sections within each county of each district was used to fill out sample sizes between 5 percent and 50 percent. A total of 300 runs were made at each sample size with each data group and the mean absolute maximum difference between the cumulative sample distribution and the true 100 percent sample cumulative distribution was calculated. Figure 11 shows the results of the study for Interstate highways and Figure 12 shows the results for Farm-to-Market roads.

The mean maximum difference between the true cumulative distribution and the sample distribution is found to be proportional to the variance of the Pavement Score for each data group. These variances are given in Table 5. The two figures represent the extreme cases of the accuracy that can be achieved with a given sample size. The least accuracy is found with the Interstate sections (Figure 11), and the greatest accuracy is with the Farm-to-Market roads (Figure 12).

Both graphs show that there is a point of diminishing returns that is reached by attempting to improve the accuracy by increasing the sample size. While there is no prescribed way of choosing the optimal sample size from these data alone (4), it is apparent that this point occurs near where the curve begins to flatten, somewhere between 20 and 35 percent, which corresponds to mean maximum differences around 3 to 10 percent.

Use of Sampling to Estimate Average Costs

Some experimentation was done to find the best estimator of the average cost of rehabilitation and major maintenance and to determine the best sample design. It was decided at the outset that the estimator which minimizes the mean squared error between the true average cost and the estimated average cost over the varying sample sizes would be selected. The true average costs are given in Table 5.

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 an observation, was collected. The following two estimators are investigated to predict mean cost per square yard:

$$\text{Mean of Ratios:} \quad \hat{R} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i} \tag{12}$$

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

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 referred to as a "ratio estimator."

Scheaffer et al. (5) 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.

Because of the large number of pavement sections with non-zero costs, it was decided to use a third estimator of the average costs which uses the mean of ratios of those sections of pavement where costs are incurred.

$$\hat{R}_{st} = \frac{N_c}{N} \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{Y_{ci}}{X_{ci}} \tag{14}$$

where

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

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

N = total number of observations in population,

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

n_c = total number of observations with non-zero costs in the sample.

Even though N_c is known for the three 100 percent sampled districts, it is not known for the other 21 districts. The totals will also change from year to year. The total number of observations with 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, where n is the total number of observations in the sample.

Simulation runs were made to determine the mean squared error between the true and the estimated average costs over all sample sizes and the Mean of Ratios method, Equation (16), minimized this error most frequently. A typical result of these simulation runs is shown in Figure 13 for the Farm-to-Market roads in District 15.

SAMPLE SURVEY DESIGN

The accuracy of any estimate made from a sample depends not only on the method by which the estimate is calculated but also on the sample survey design. Three designs are investigated.

(1) Simple random sampling with district stratification,

(2) Simple random sampling with county stratification, and

(3) Systematic sampling with district stratification.

Each will be described briefly below.

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.

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, ..., 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:

$$\text{Method A:} \quad \hat{R}_A = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i} \tag{15}$$

$$\text{Method B:} \quad \hat{R}_B = \frac{1}{N} \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{Y_{hi}}{X_{hi}} \tag{16}$$

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.

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 consists 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 than the stratified random samples.

Simulation of Average Cost Estimates

Simulation runs were made for sample sizes ranging from 5 to 50 percent using each of the types of sampling survey described above. The mean percent error in the estimated cost per square yard was calculated after making 300 separate samplings of each highway system in each district. A typical result is shown in Figure 14.

On the basis of the simulation study, the County Stratification Method A was determined to be the sampling survey design which produced consistently the smallest mean error for any size of sample. Table 9 shows the results of this study averaged over all districts indicating the percent error in the average costs per square yard for different sizes of sample.

It is worth noting that the better condition the highway network is in, the more difficult it is to make an accurate estimate of average costs per square yard. This is illustrated in Figure 15 which shows the relation between the mean pavement score and the mean percent error in the average costs per square yard for a 25 percent sample. The curves for each highway system trend upward as the Pavement Score increases.

CONCLUSIONS

Through the use of simulation procedures described in this paper, 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 by estimating its parameters by the method of moments. This procedure leads to a better estimate of the percentage of pavements with a score below 40 than by a direct estimate from the sample.

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

(d) 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.

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

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

(g) Sampling methods A and B, which are stratified over counties, lead to similar results both in predicting pavement condition and in estimating costs.

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

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

REFERENCES

1. 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.
2. W. J. Conover. Practical Nonparametric Statistics. John Wiley & Sons, Inc., New York, 1971.
3. Robert V. Hogg and Allen T. Craig. Introduction to Mathematical Statistics. MacMillan Publishing Co., Inc., New York, 1978.
4. 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.
5. R. L. Schaeffer, W. Mendenhall, and L. Ott. Elementary Survey Sampling. Duxbury Press, Belmont, California, 1979.

LIST OF TABLES

TABLE 1. Determining Weighting Factor for ADT

TABLE 2. Determining Weighting Factor for 18-kip ESAL

TABLE 3. Determining Weighting Factor for Rainfall

TABLE 4. Determining Weighting Factor for Freeze-Thaw

TABLE 5. Pavement Scores Costs for the 100 Percent Sample Districts

TABLE 6. Maximum Absolute Difference Between 100 Percent and 5 Percent Sample Histograms

TABLE 7. Estimated Values of the Parameters a and b of the Beta Distribution

TABLE 8. Mean Error in Predicting Percentage of Roads Below a Pavement Score of 40

TABLE 9. Percent Error in Average Cost per Square Yard with County Stratification (Method A)

TABLE 1. Determining Weighting Factor for ADT

ADT	ADTF
<300	1.00
300 - 750	0.96
750 - 2,000	0.92
2,000 - 7,500	0.88
7,500 - 25,000	0.84
>25,000	0.80

TABLE 2. Determining Weighting Factor for 18-kip ESAL

20 year 18-kip (ESAL) (millions)	EALF
< 6	1.0
6-12	0.95
> 12	0.90

TABLE 3. Determining Weighting Factor for Rainfall

Rainfall (ins)	RF
< 20	1.00
20-40	0.97
> 40	0.94

TABLE 4. Determining Weighting Factor for Freeze-Thaw

Freeze-Thaw cycles/year	FF
< 10	1.00
10-30	0.973
30-50	0.967
> 50	0.960

TABLE 5. Pavement Scores Costs for the 100 Percent Sample Districts

Data Group	District	Number of Pavement Sections	Average Pavement Score	Variance of Pavement Score	Percentage of Pavements with Zero Costs	Mean Cost per sq.yd.
Interstate	8	154	59	1392	61.7	$1.35
	11	--	--	--	--	--
	15	323	88	429	92.3	0.46
U.S. Highways	8	324	79	795	85.5	0.39
	11	241	57	1343	58.5	1.25
	15	331	83	581	88.5	0.40
State Highways	8	270	69	975	75.2	0.40
	11	300	68	889	78.3	0.62
	15	499	82	634	89.6	0.21
Farm-to-Market Roads	8	938	73	639	88.4	0.14
	11	832	63	816	78.1	0.37
	15	1101	86	386	96.2	0.12

TABLE 6. Maximum Absolute Difference Between 100 Percent and 5 Percent Sample Histograms

Data Group	District	Maximum Absolute Difference Between 100% & 5% Histograms	Number of Observations in the 5% Sample
Interstate	8	0.1560	8
	11	--	--
	15	0.0744	15
U.S. Highways	8	0.0698	15
	11	0.1769	10
	15	0.1319	17
State Highways	8	0.1350	12
	11	0.1389	14
	15	0.0889	23
Farm-to-Market Roads	8	0.0601	48
	11	0.1129	41
	15	0.0703	55

TABLE 7. 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

TABLE 8. Mean Error in Predicting Percentage of Roads Below a Pavement Score of 40

Data Group	District	Mean Error Calculated Directly from Sample	Mean Error Calculated from Beta Distribution
Interstate	8	0.175	0.154
	11	--	--
	15	0.051	0.039
U.S. Highways	8	0.071	0.058
	11	0.125	0.109
	15	0.057	0.039
State Highways	8	0.098	0.080
	11	0.085	0.069
	15	0.051	0.044
Farm-to-Market Roads	8	0.036	0.031
	11	0.051	0.045
	15	0.022	0.017

TABLE 9. Percent Error in Average Cost per Square Yard 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	0.2997
15	0.3897	0.23464	0.3009	0.2753
20	0.3286	0.2071	0.2688	0.2247
25	0.2931	0.1912	0.2481	0.2057
30	0.28214	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

LIST OF FIGURES

Figure 1(a). Observation From a Two-mile Segment of a Divided Highway
(b). Observation From a Two-mile Segment of a Two-lane Highway

Figure 2. Location of 100 Percent Sampled Districts

Figure 3. Serviceability Index Utility Curve

Figure 4. Relation of the Variance of Pavement Score to the Mean Pavement Score

Figure 5. Relations Between Mean Pavement Score and the Percentage of Pavements with Zero Costs

Figure 6. Relations Between Mean Pavement Score and the Average Cost per Square Yard

Figure 7. Confidence Bands for F(X)

Figure 8. Confidence Band for Cumulative Distribution Function of Pavement Scores F(X), District 11, System FM

Figure 9. Relation Between the Percent of Pavements with Zero Costs and the Beta Distribution Parameter, a

Figure 10. Relation Between the Percent of Pavements with Zero Costs and the Beta Distribution Parameter, b

Figure 11. Comparison of District Results of the Mean Maximum Difference Using County Stratification Method A, System IH

Figure 12. Comparison of District Results of the Mean Maximum Difference Using County Stratification Method A, System FM

Figure 13. Comparison of Estimations, District 15, System FM

Figure 14. Comparison of Sample Designs Used to Estimate Costs, System IH

Figure 15. Relation Between Mean Pavement Score and Mean Percent Error in Estimated Average Cost per Square Yard

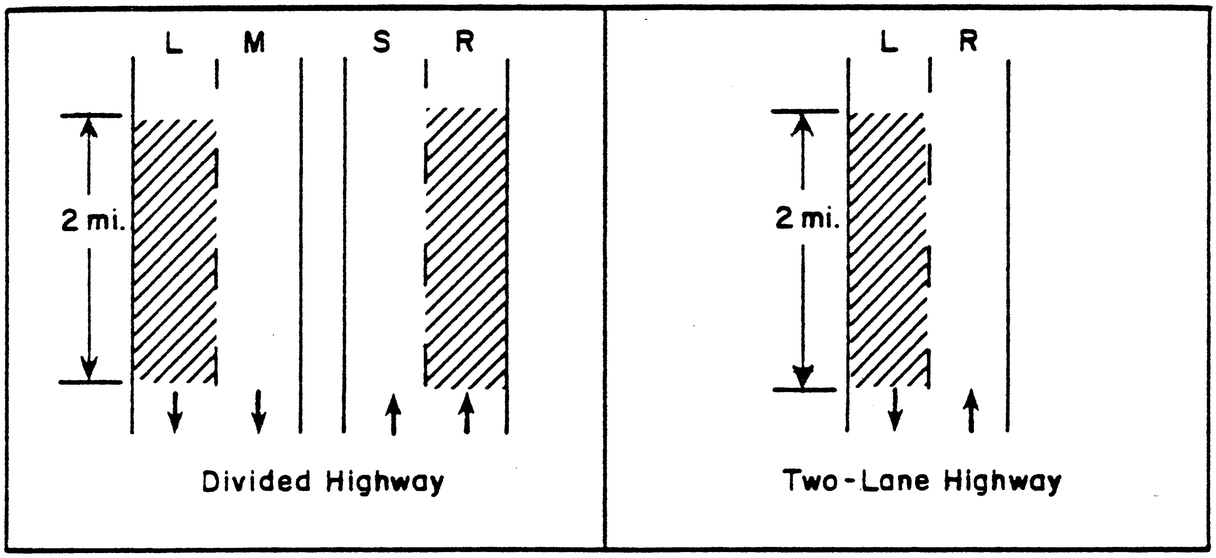


Figure 1(a). Observation From a Two-mile Segment of a Divided Highway.

Figure 1(b). Observation From a Two-mile Segment of a Two-lane Highway.

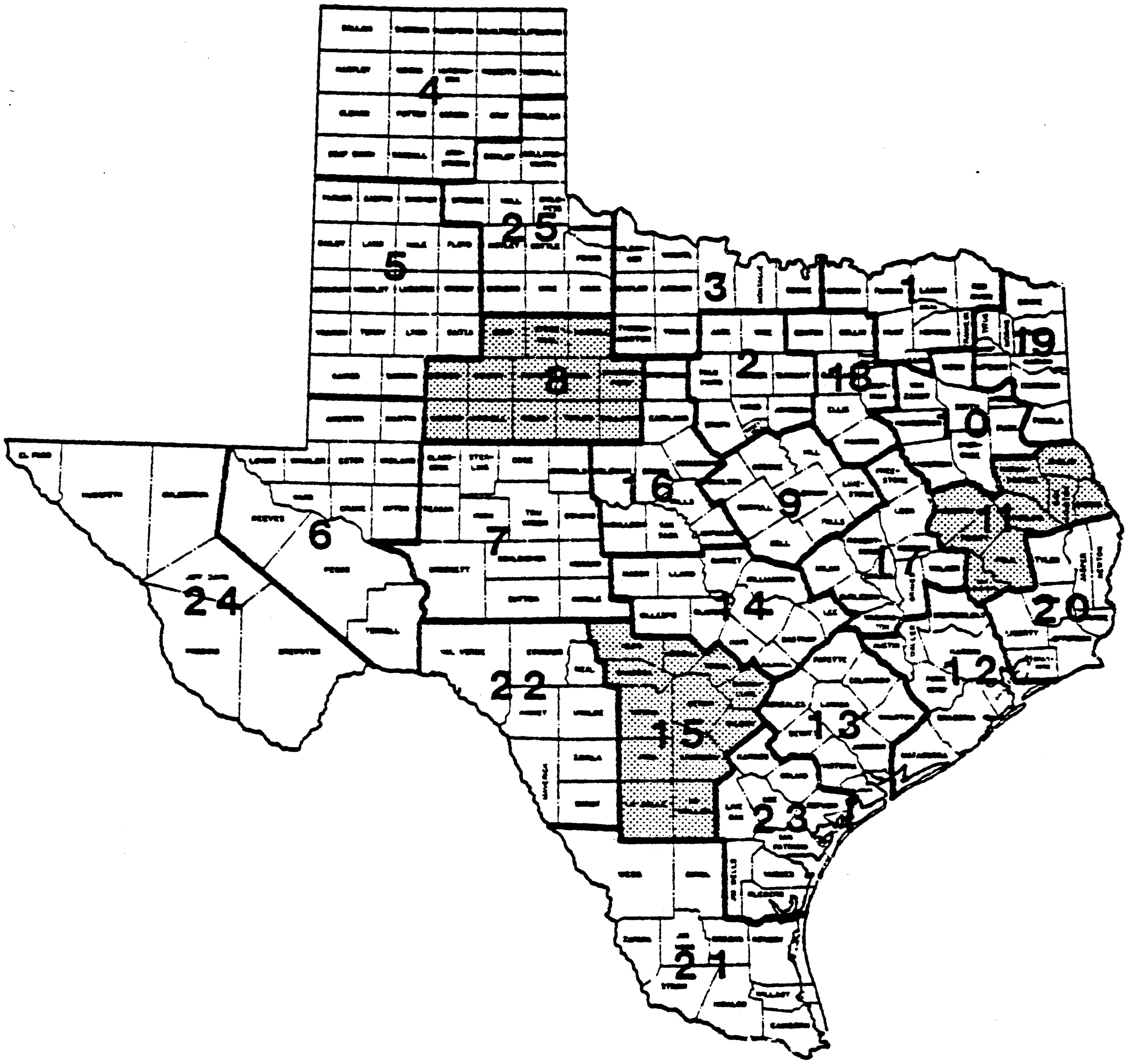
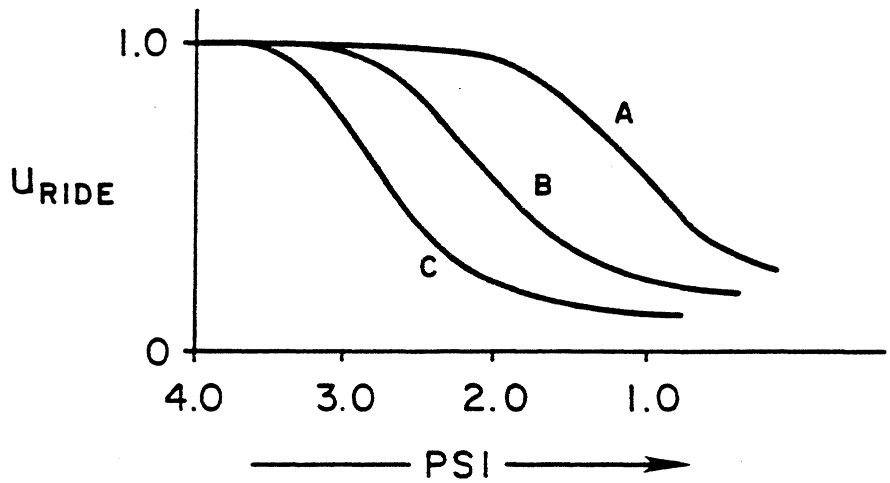


Figure 2. Location of 100 Percent Sampled Districts



CURVE	$\frac{ADT}{LANE}$ x SPEED
A	< 27,500
B	27,500 - 165,000
C	> 165,000

Figure 3. Serviceability Index Utility Curve

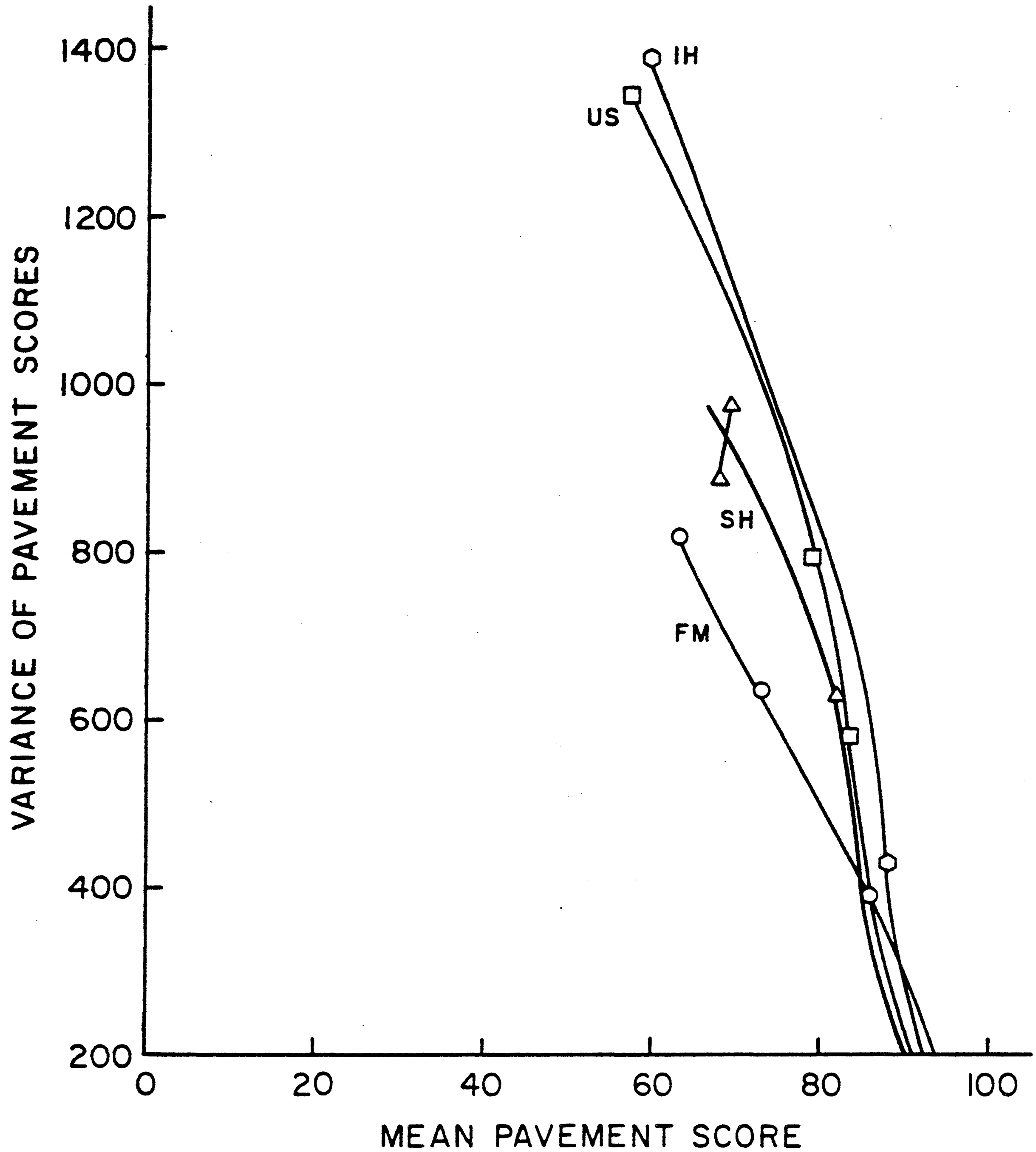


Figure 4. Relation of the Variance of Pavement Score to the Mean Pavement Score

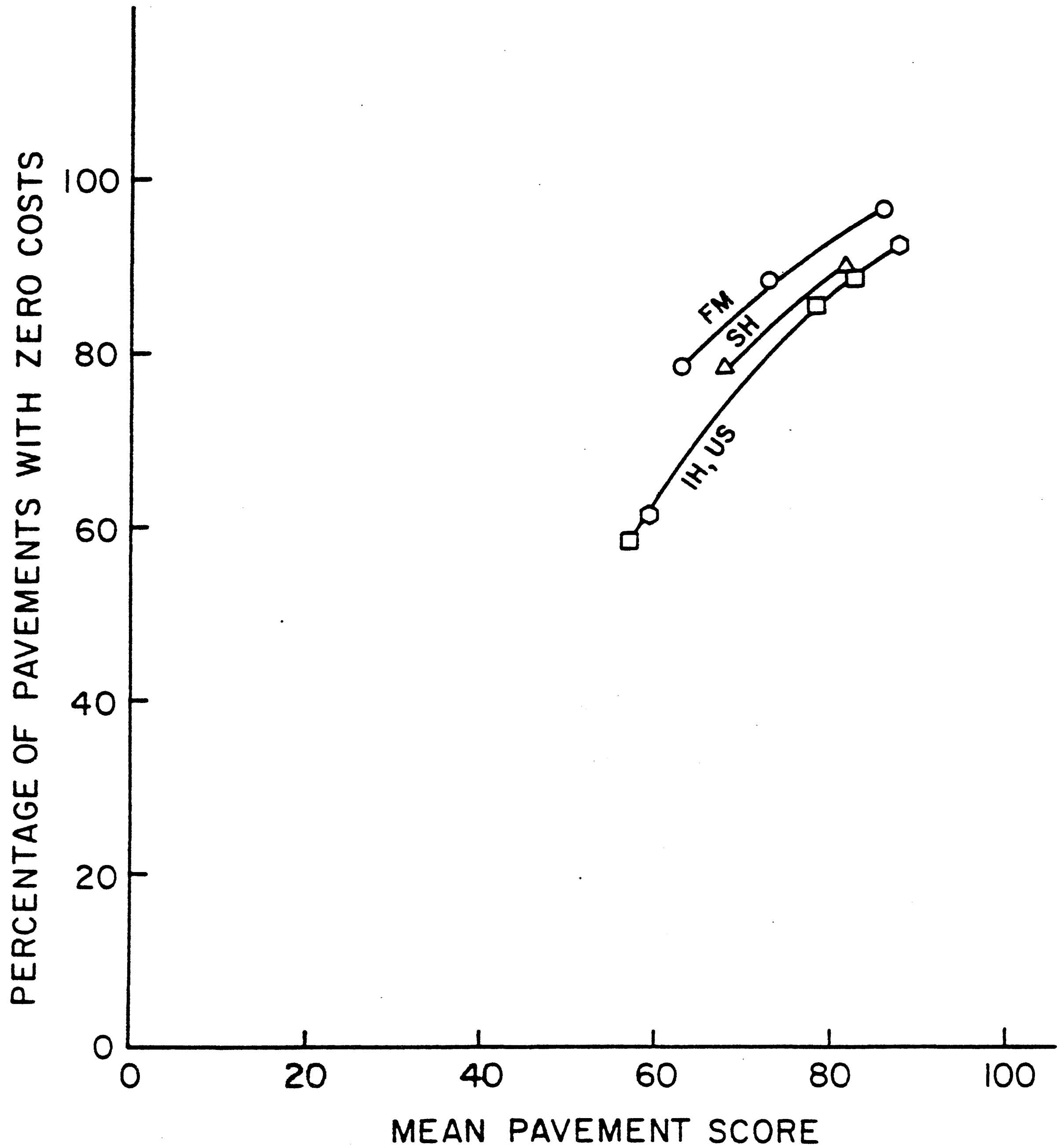


Figure 5. Relations Between Mean Pavement Score and the Percentage of Pavements with Zero Costs

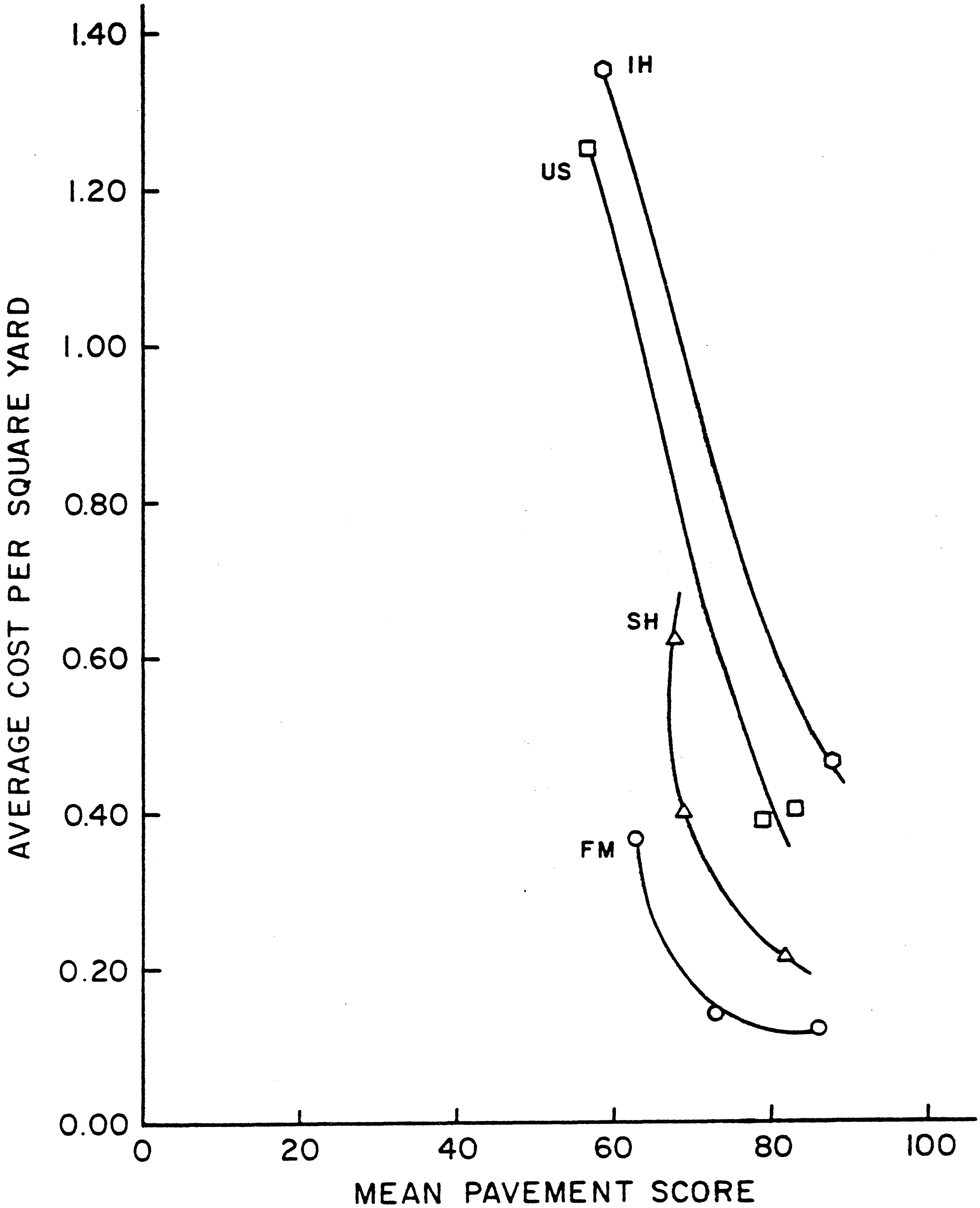


Figure 6. Relations Between Mean Pavement Score and the Average Cost per Square Yard

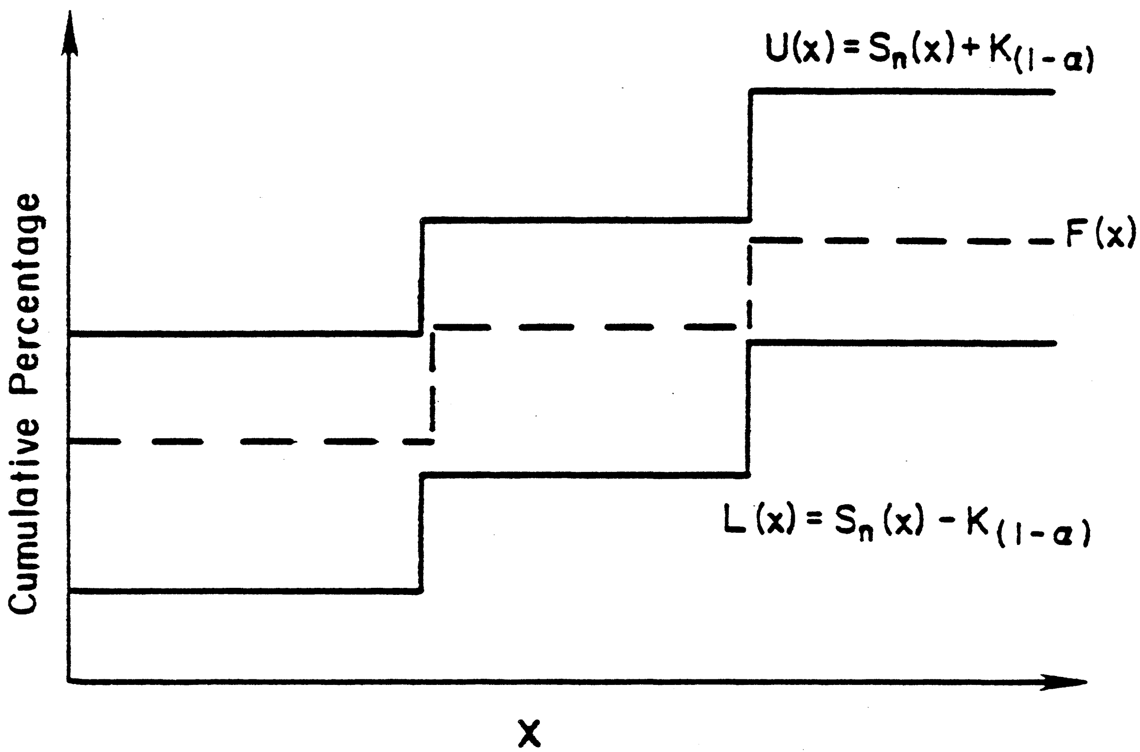


Figure 7. Confidence Bands for F(X)

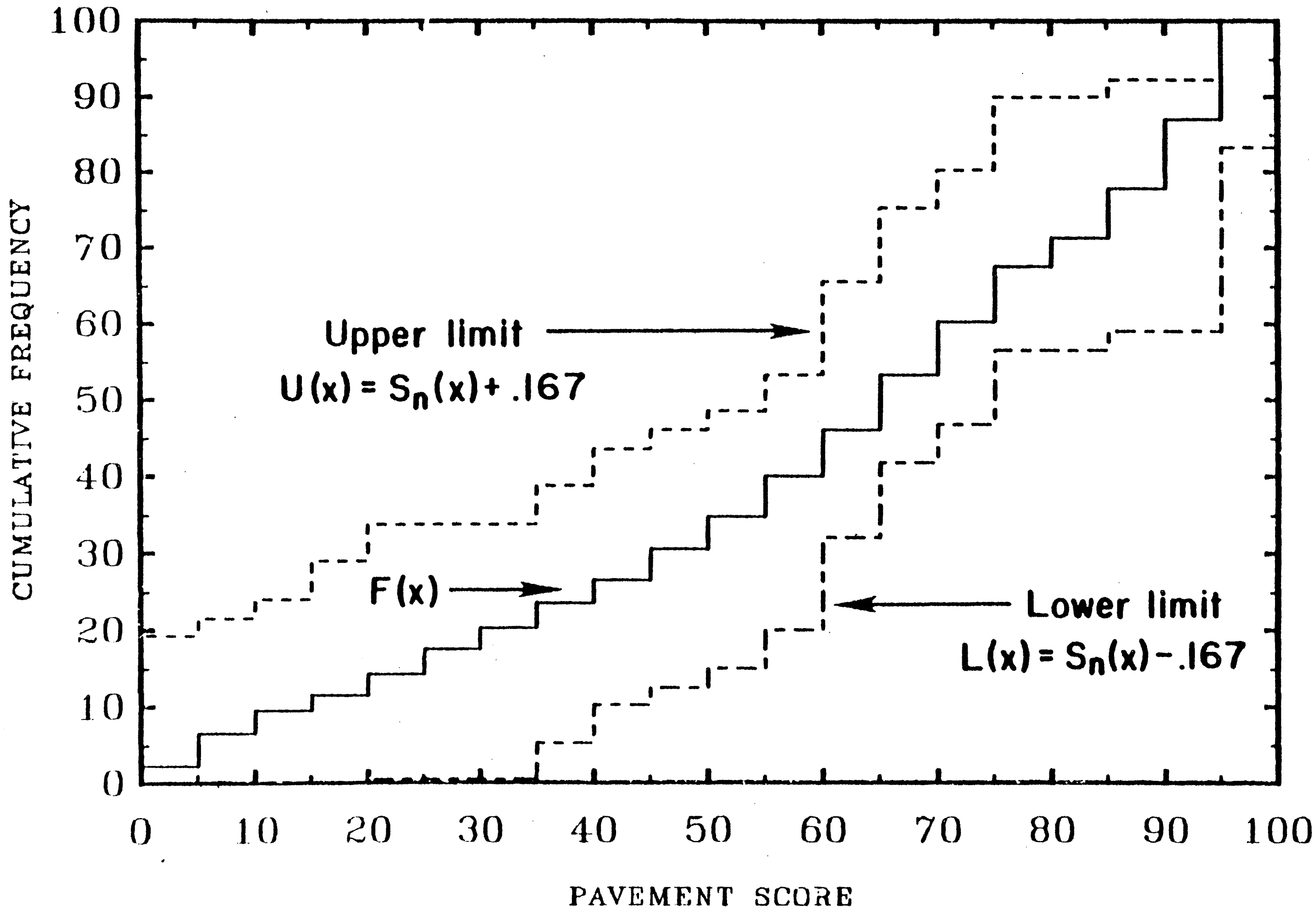


Figure 8. Confidence Band for Cumulative Distribution Function of Pavement Scores F(X), District 11, System FM

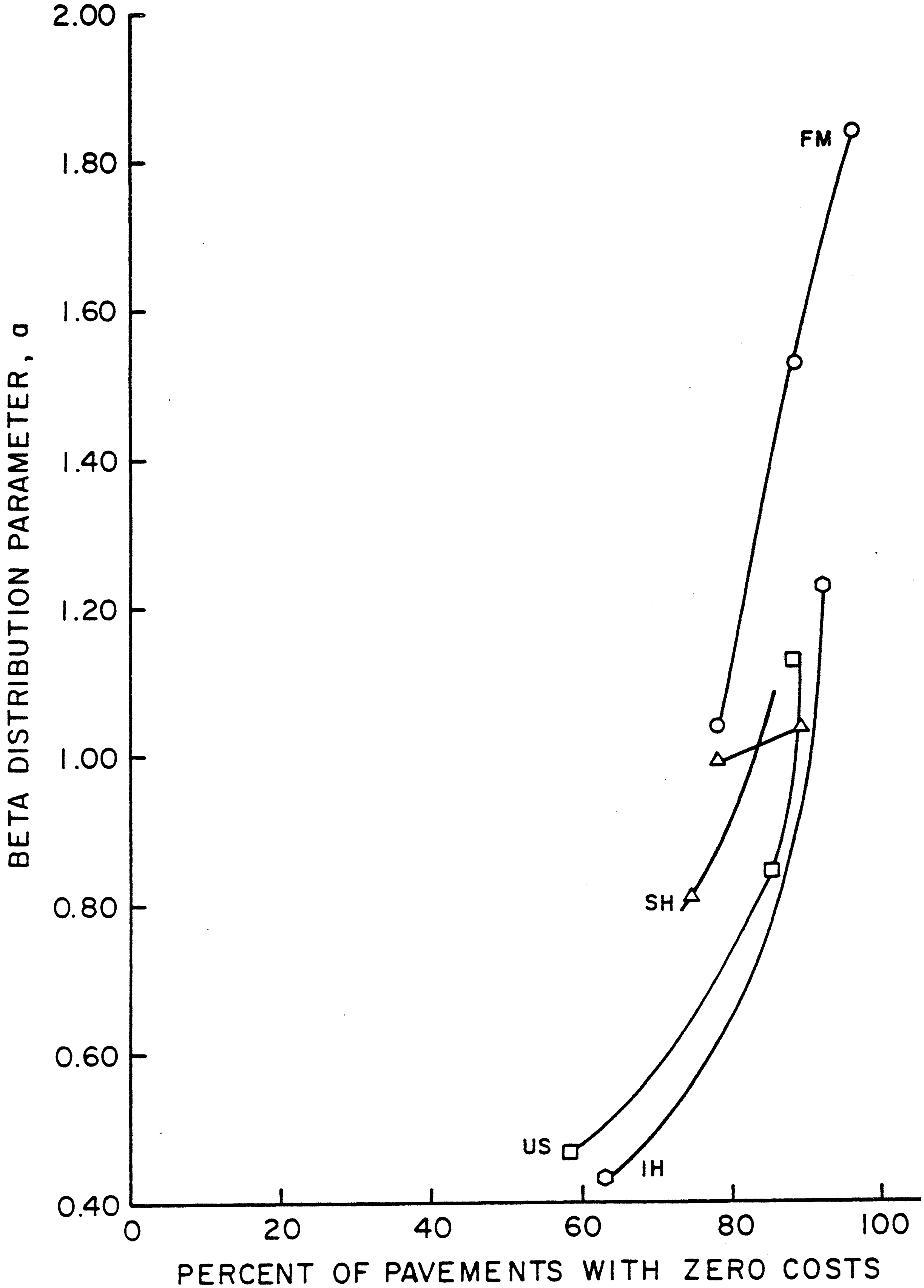


Figure 9. Relation Between the Percent of Pavements with Zero Costs and the Beta Distribution Parameter, a

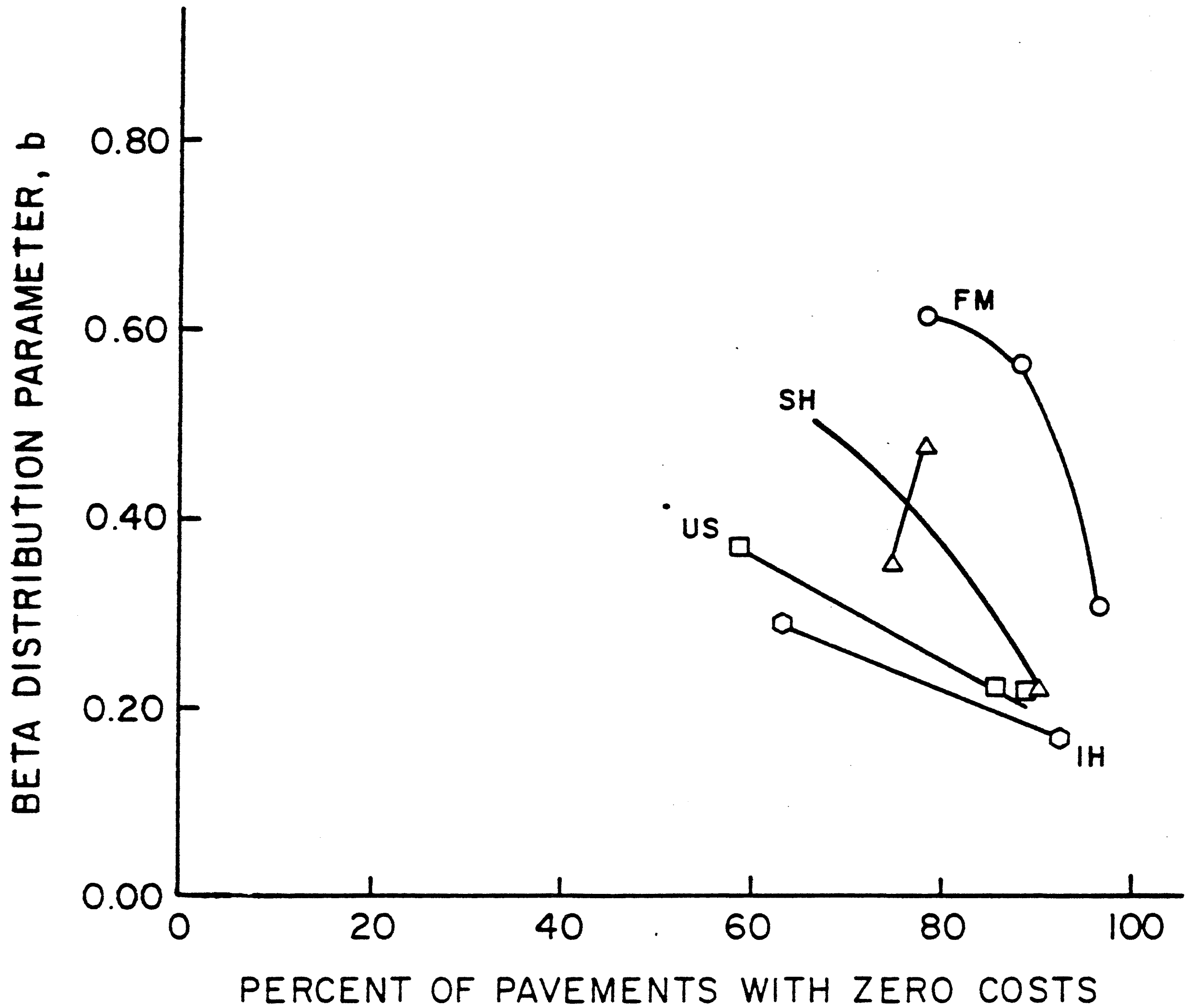


Figure 10. Relation Between the Percent of Pavements with Zero Costs and the Beta Distribution Parameter, b

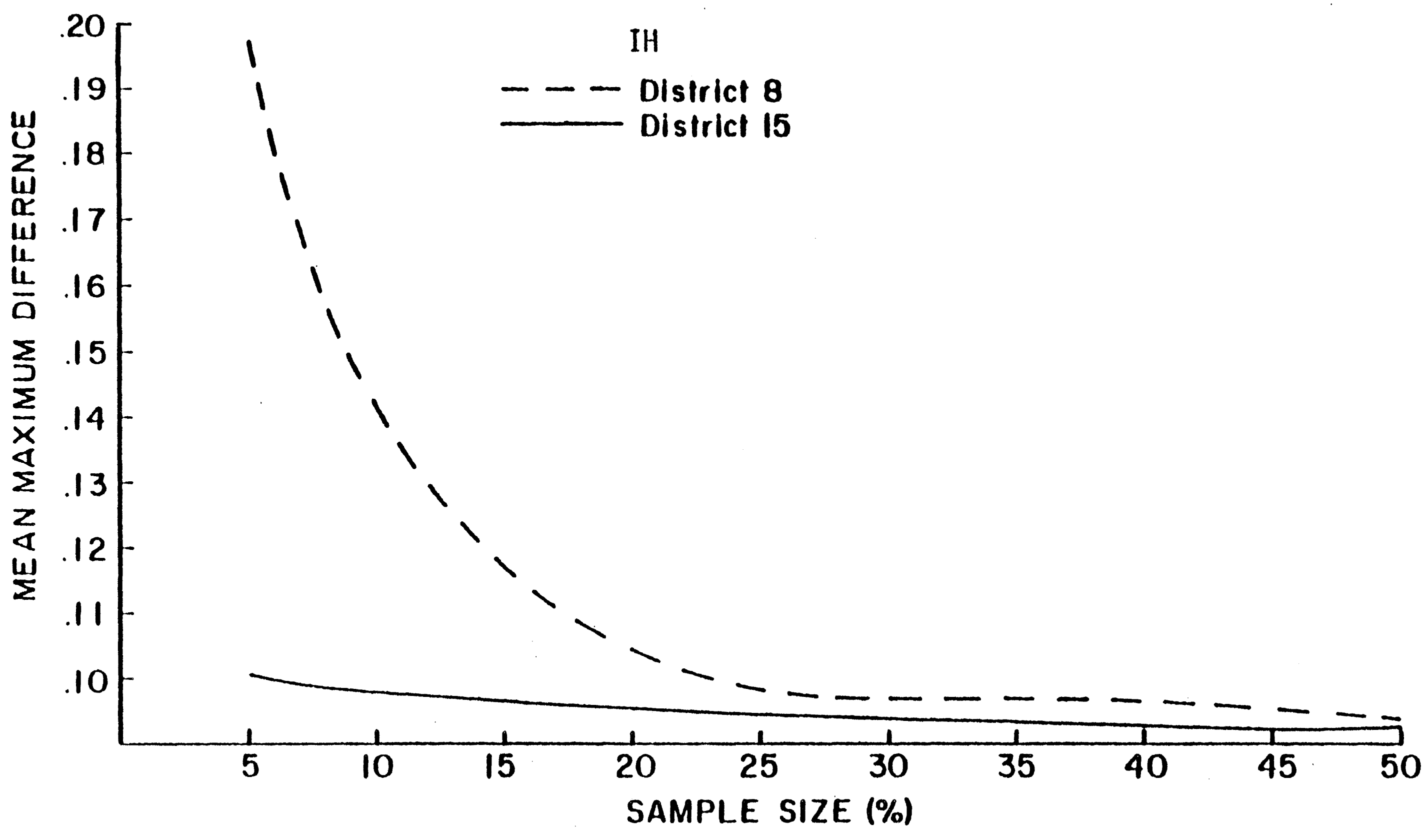


Figure 11. Comparison of District Results of the Mean Maximum Difference Using County Stratification Method A, System IH

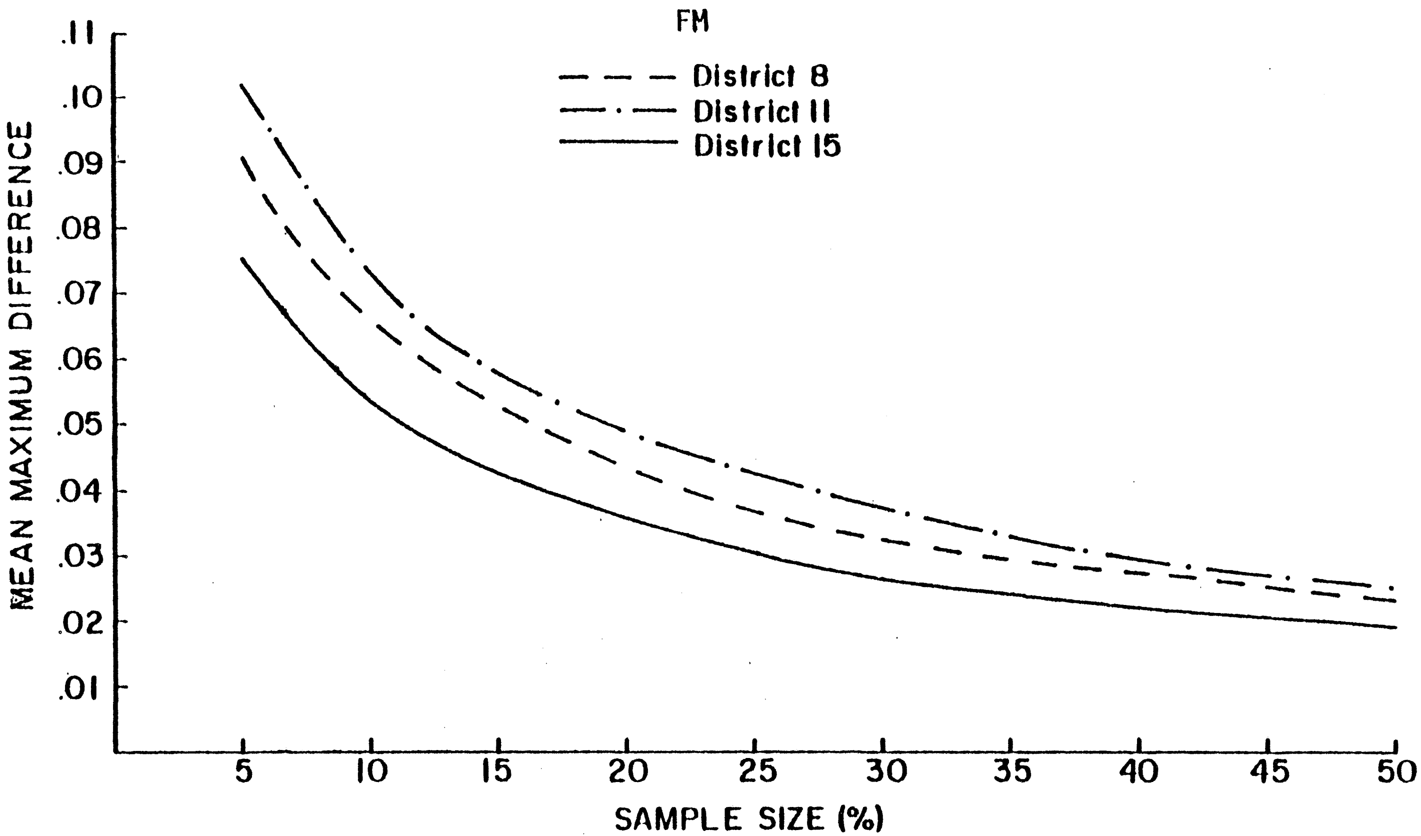


Figure 12. Comparison of District Results of the Mean Maximum Difference Using County Stratification Method A, System FM

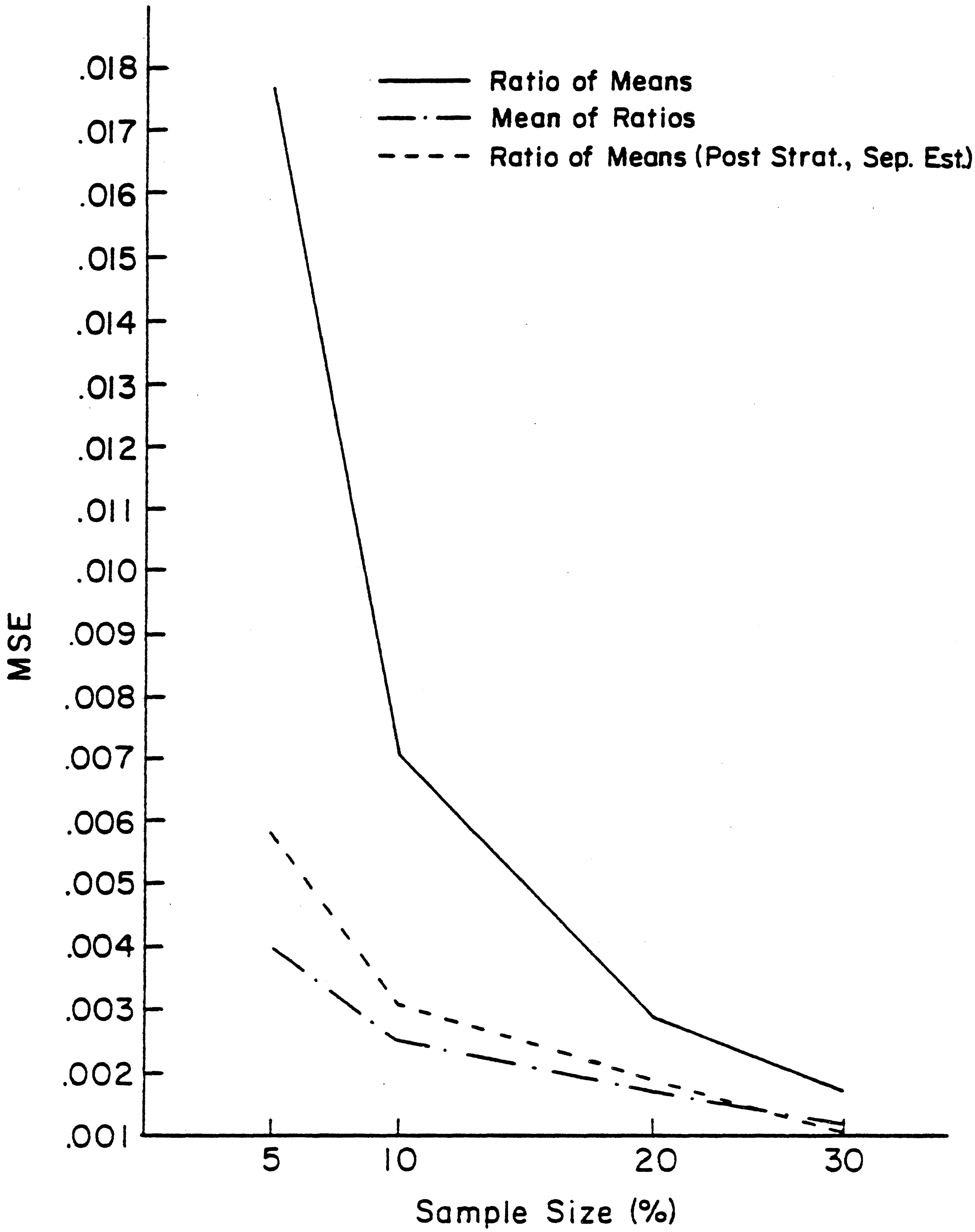


Figure 13. Comparison of Estimations, District 15, System FM

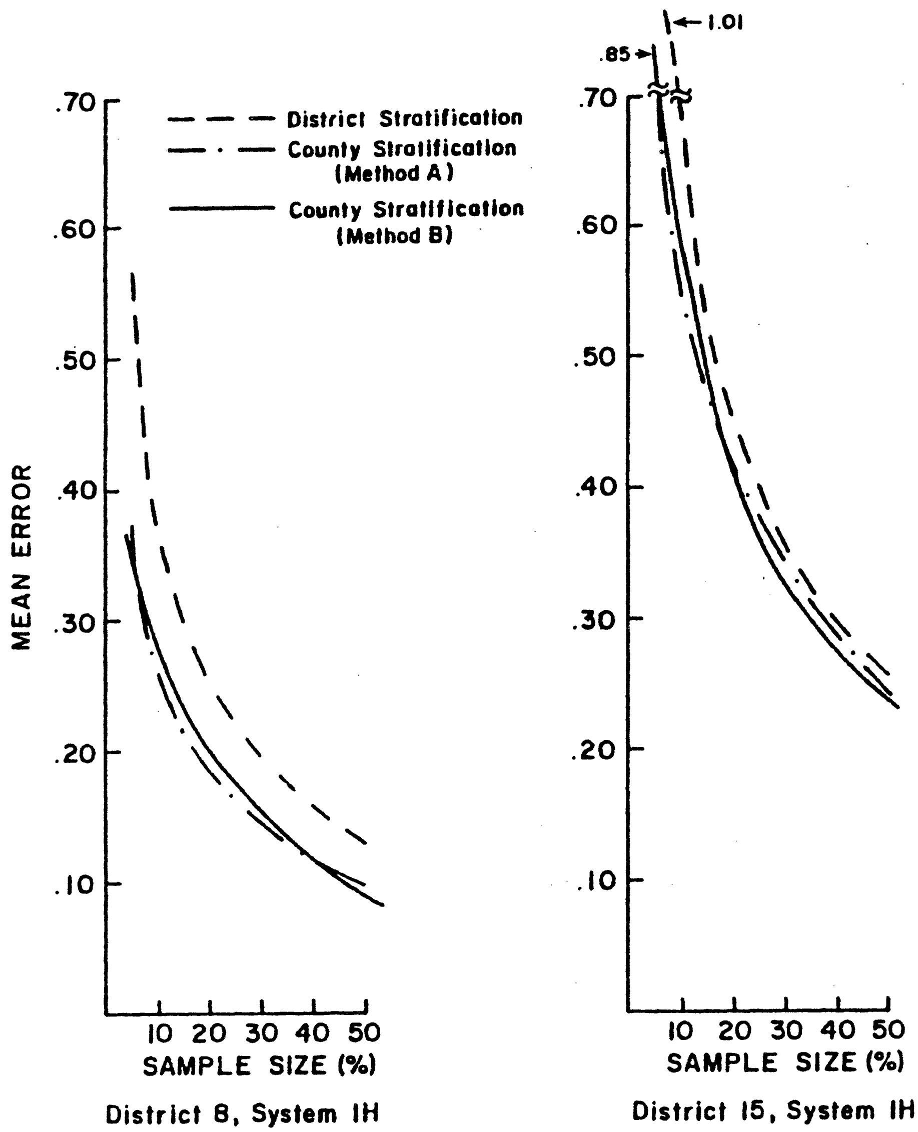


Figure 14. Comparison of Sample Designs Used to Estimate Costs, System IH

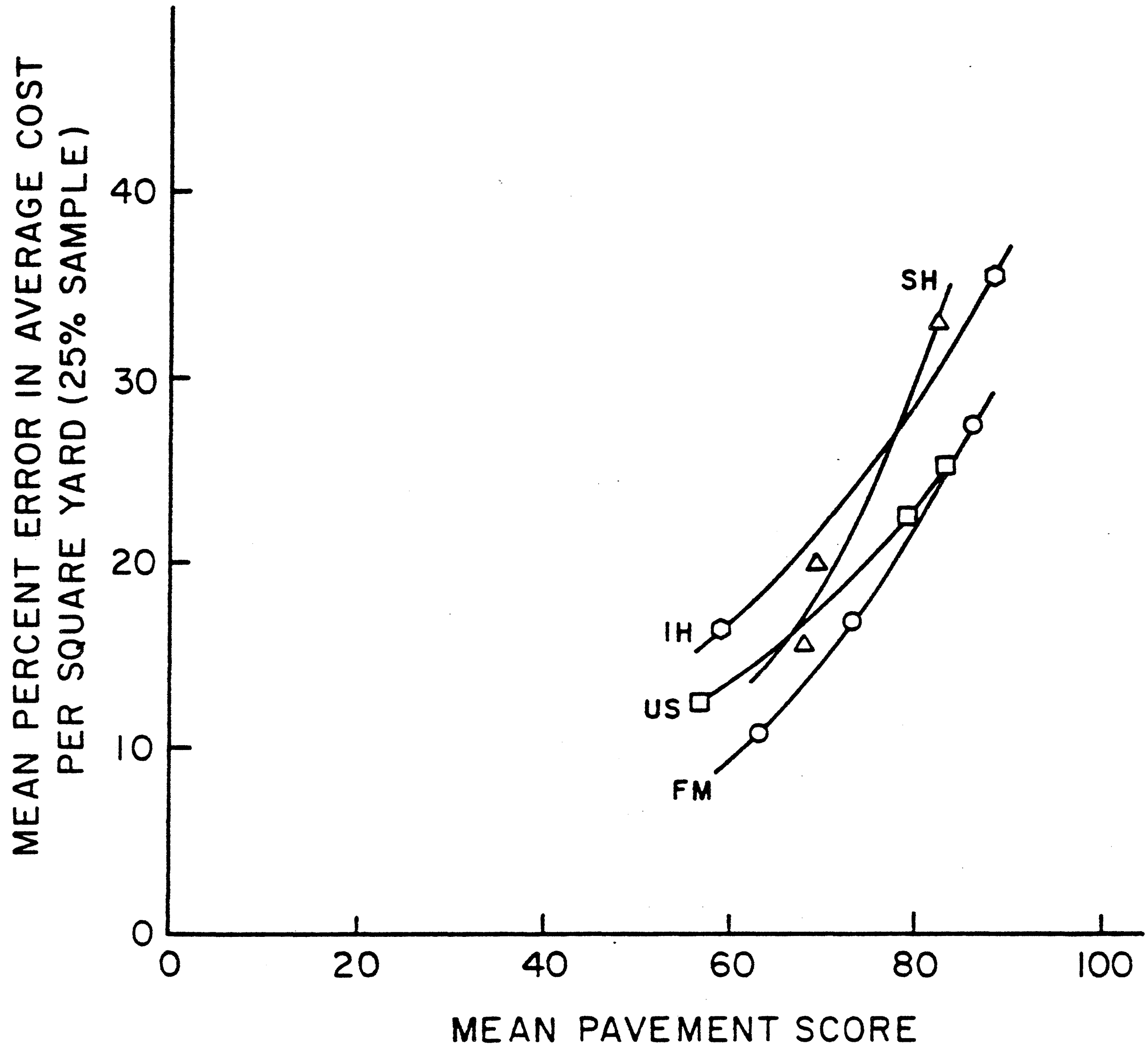


Figure 15. Relation Between Mean Pavement Score and Mean Percent Error in Estimated Average Cost per Square Yard