Network Sampling to Estimate Distribution of Pavement Condition and Costs

J. TEMPLETON, R. L. LYTTON, and A. GARCIA-DIAZ

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 presented the results of a study of data from a 1982 survey of the total mileage of three district networks in Texas that was undertaken to answer several questions concerning the accuracy that can be achieved with different sample sizes. It is found that the sample size 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 among classes of highway: Interstate, U.S., state, and farm-to-market. Several convenient relations were found among the mean pavement score and the variance of pavement scores, percentage of pavements needing no repairs, average costs per square yard, and percentage error in the estimated average costs. It is not surprising to find that the average costs are reduced and the percentage 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.

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. Results of the analysis of data collected on 100 percent of the 2-mi-long pavement sections in three districts of the Texas State Department of Highways and Public Transportation, each of which is responsible for between 2,500 and 3,000 mi of pavement of all functional classes, are reported. The objective of the study was to determine the answers to several questions about the minimum sample size and the 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. It is 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 distribution 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 sample size is required to get an accurate estimate of the distribution of pavement scores?
5. What sample size is needed 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 that could be supported by the data and 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 2 mi 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 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. 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.

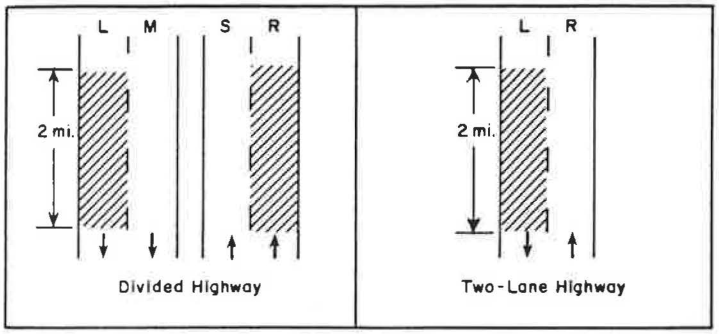

FIGURE 1 Observation from a 2-mi segment of a divided highway (a) and observation from a 2-mi segment of a two-lane highway (b).

FIGURE 2 Location of 100 percent sampled districts.

For each observation, in both the 5 and the 100 percent samples, the serviceability index was determined with the May's ridemeter and a visual defects rating was performed. In the visual rating, the following 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 to 10, 11 to 25, or >25 percent).

In Texas, the pavement evaluation score is a number between 0 and 100 that represents a weighted condition of the riding quality rating and visual distress rating of the pavement. A detailed description of the scoring procedure is contained elsewhere (1). The visual distress rating method is described by Epps et al. (2).

ESTIMATING COSTS OF MAJOR MAINTENANCE AND REHABILITATION

In evaluating the results of the statewide condition survey, each pavement section with a pavement score of 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/yd²);
2. One-inch asphaltic concrete pavement (ACP) overlay or seal plus level-up ($1.58/yd²);
3. Two-and-one-half-inch ACP overlay ($3.41/yd²);
4. Four-inch ACP overlay ($6.05/yd²); or
5. Seven-and-one-half-inch ACP overlay ($11.93/yd²) or its equivalent in reconstruction.

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

For each of the strategies, the estimated rehabilitated pavement score was computed and deterioration calculations were made to determine life expectancy. This expected life was compared to a minimum allowable expected life of 3 to 5 years depending on the class of highway in order 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-mi 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 1. It is noted that District 11 has no Interstate highway mileage.

When these results are plotted against the mean pavement score, clear trends emerge. For example, in Figure 3 the relations between the variance of pavement score and the mean pavement score for the different highway systems are shown. Knowing this relation and the probability density function for pavement scores, it is possible to construct an accurate pavement score distribution when a good estimate of the mean score has been determined. In Figure 4 the relations between the mean pavement score and the percentage of pavements with zero costs are shown. The trends are again quite clear, and it is significant that the farm-to-market system curve is above the state curve that, in turn, is above the U.S. and the Interstate highway curve. In Figure 5 the average costs per square yard are plotted against the mean pavement score to show that even these costs can be estimated when 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 3-5 to make accurate estimates of the condition and the costs of rehabilitating pavement networks.

TABLE 1 Pavement Score Costs for the 100 Percent Sample Districts

Data Group	District	No. of Pavement Sections	Average Pavement Score	Variance of Pavement Score	Percentage of Pavements with Zero Costs	Mean Cost per Square Yard ($)
Interstate	8	154	59	1,392	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	1,343	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	1,101	86	386	96.2	0.12

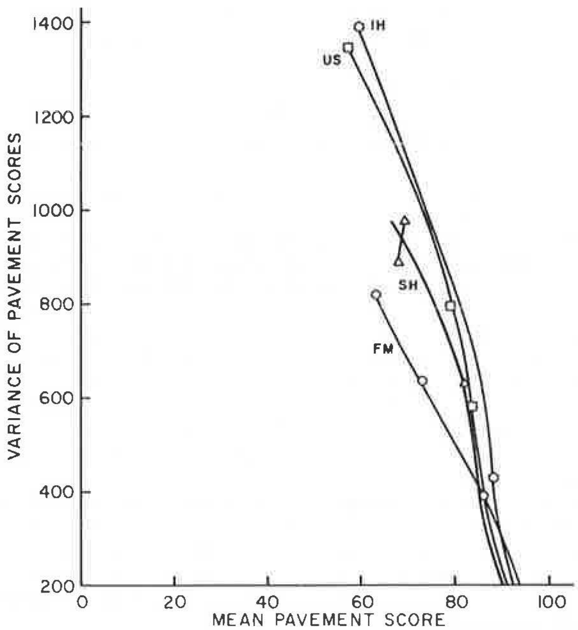

FIGURE 3 Relation of the variance of pavement score to the mean pavement score.

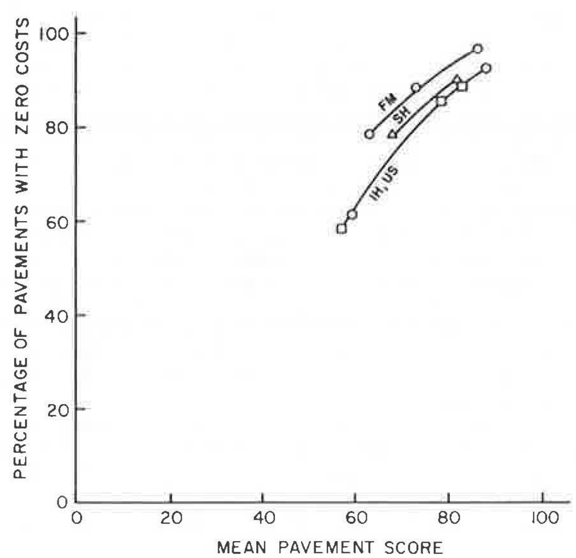

FIGURE 4 Relations between mean pavement score and the percentage of pavements with zero costs.

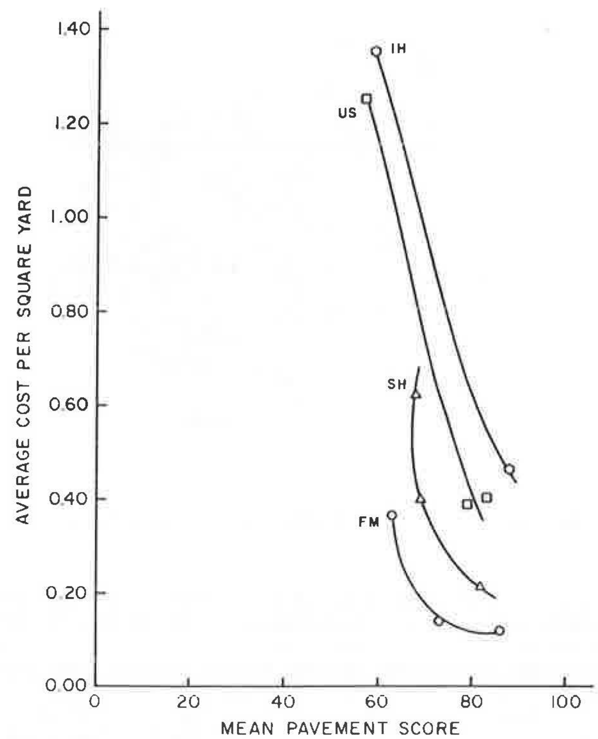

FIGURE 5 Relations between mean pavement score and average cost per square yard.

How large a sample is necessary to make accurate estimates of these quantities? Is a 5 percent sample large enough? What is the best way to organize a sample survey 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 3-5. The answers are determined by simulating different sizes of surveys using the data from the 100 percent sampled districts.

ACCURACY OF THE 5 PERCENT SAMPLE IN ESTIMATING NETWORK PAVEMENT SCORE DISTRIBUTIONS

To determine the accuracy resulting from a 5 percent sample in predicting pavement condition, the data from the 100 percent sampled districts were first divided into 15 groups, and then a 5 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 20 class intervals, each with a width of five. A comparison of the 100 percent sample histogram and the corresponding 5 percent histogram shows that the 5 percent histogram more closely resembles the 100 percent histogram in those data grouped with a larger number of observations. Table 2 gives the maximum absolute difference between the two histograms for each data group along with the number of observations in the 5 percent sample. The maximum difference ranges between 0.0601 and 0.1769.

TABLE 2 Maximum Absolute Difference Between 100 Percent and 5 Percent Sample Histograms

Data Group	District	Maximum Absolute Difference Between 100% and 5% Histograms	No. 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-rarket roads	8	0.0601	48
	11	0.1129	41
	15	0.0703	55

CONFIDENCE BANDS ON THE TRUE CUMULATIVE DISTRIBUTION

To statistically compare the pavement score distribution based on the 5 percent sample with the pavement score distribution based on the 100 percent sample, a percentile of the Kolmogorov-Smirnov test statistic (3) is used. This percentile along with an empirical cumulative distribution $[S_n(X)]$ can be used to form a $100(1 - \alpha)$ percent confidence band for a true cumulative distribution $[F(X)]$. As the percentage of confidence is increased, the band becomes wider. Figure 6 shows one of the better comparisons of the true 100 percent histograms $[F(X)]$ and the confidence bands generated using a 5 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 (3), 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 $100(1 - \alpha)$ 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 of 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

$$F(X) = [1/B(a,b)]\, X^{a-1} (1 - X)^{b-1} \quad \text{for } a,b > 0;\ 0 < X < 1 \tag{1}$$

In Equation 1, $B(a,b)$ is defined as

$$B(a,b) = \int_0^1 X^{a-1} (1 - X)^{b-1}\, dx \tag{2}$$

Because the random variable (X) must be in the interval between zero and one, all of the pavement scores are divided by 100 to satisfy this condition.

The parameters a and b are estimated by the following procedure. The mean of the 100 percent sample distribution is set equal to $\hat{\mu}$ and the variance to $\hat{\sigma}^2$. The estimated values of a and b are then given by

$$\hat{a} = \{[\hat{\mu}^2 (1 - \hat{\mu})] / \hat{\sigma}^2\} - \hat{\mu} \tag{3}$$

and

$$\hat{b} = (\hat{a} - \hat{a}\hat{\mu})/\hat{\mu} \tag{4}$$

Table 3 gives 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 (4), $\hat{a}$ and $\hat{b}$ are consistent estimators of a and b.

A Kolmogorov-Smirnov goodness of fit test (3) was used to determine if the 100 percent samples do actually come from a beta distribution with the a and

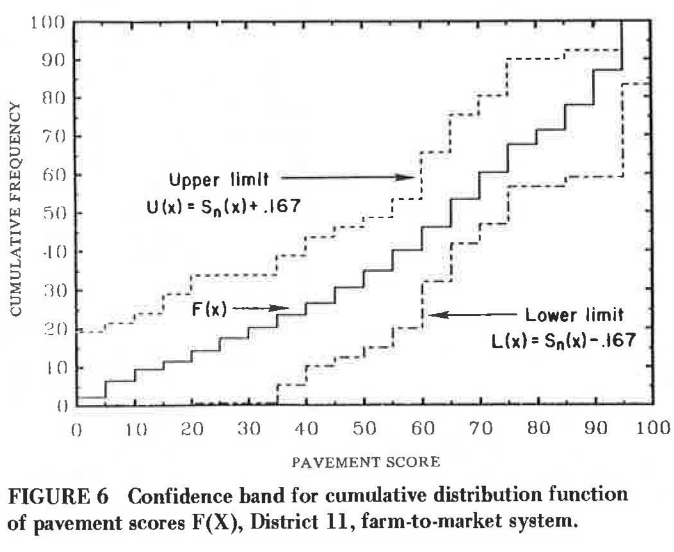

FIGURE 6 Confidence band for cumulative distribution function of pavement scores F(X), District 11, farm-to-market system.

TABLE 3 Estimated Values of the Parameters a and b of the Beta Distribution

Data Group	$\hat{a}$	$\hat{b}$
I-08	0.434	0.291
FM-08	1.522	0.564
SH-08	0.810	0.351
US-08	0.843	0.220
FM-11	1.048	0.612
SH-11	0.990	0.476
US-11	0.467	0.369
I-15	1.226	0.161
FM-15	1.844	0.304
SH-15	1.038	0.217
US-15	1.125	0.218

b parameters given in Table 3. The tests showed that the samples did come from a beta distribution.

ACCURACY OF ESTIMATING THE PERCENTAGE 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 randomly selected 5 percent samples. The second method made use of the cumulative beta distribution. The values of $\hat{\mu}$ and $\hat{\sigma}^2$ were estimated from the sample. Making use of Equations 3 and 4, $\hat{a}$ and $\hat{b}$ were estimated from selected 5 percent samples. Using the International Mathematical and Statistical Libraries, Inc., (IMSL) subroutine, MDBETA, the percentage of observations falling at or below 40 was then estimated.

Table 4 gives 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.

TABLE 4 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

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

How large a sample size is needed to obtain consistent accuracy in estimating the pavement score cumulative distribution? This question was studied using a simulation program. 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 7 shows the results of the study for farm-to-market roads. The least accuracy is found for the Interstate sections, and the greatest accuracy is for the farm-to-market roads.

Figure 7 shows that there is a point of diminishing returns that is reached by attempting to improve the accuracy by increasing the sample size. Although there is no prescribed way of choosing the optimal sample size from these data alone (5), 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 of around 3 to 10 percent.

USE OF SAMPLING TO ESTIMATE AVERAGE COSTS

Simulation runs were made for sample sizes ranging from 5 to 50 percent to estimate average costs. This required the selection of a sampling method and a means of calculating the average cost. Simulation studies were used to determine both of these and the detailed results are reported elsewhere (6). The most accurate sampling method proved to be a random sampling in each county of each district. The most accurate method of calculating the average cost is to take an average of the costs per square yard on a sample-by-sample basis. Table 5 gives the results of the simulation study, in which the percentage error in the average costs per square yard was calculated for different sample sizes.

It is worth noting that the better condition a 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 8, which shows that the mean percentage error in average costs per square yard for a 25 percent sample increases with the mean pavement score.

CONCLUSIONS

Through the use of simulation procedures, the following basic results were obtained:

- It was found that the beta distribution fits the pavement score data in each of the 11 data groups.
- The percentage of roads with a pavement score of 40 or less can best be estimated through the use of the cumulative beta distribution. This procedure leads to a better estimate of the percentage of pavements with a score below 40 than does a direct estimate from the sample.
- The more variable the pavement scores are within a district, the greater is the mean maximum difference between the sample cumulative distribution and the 100 percent sample cumulative distribution.
- Pavement condition can be estimated accurately with a smaller sample size than is required to estimate the costs of maintenance and rehabilitation to the same degree of accuracy.
- Simple random sampling on a county-by-county basis usually leads to the smallest relative errors in predicting costs.
- The better the condition of roads within a district, the more difficult it is to accurately estimate a mean cost per square yard.

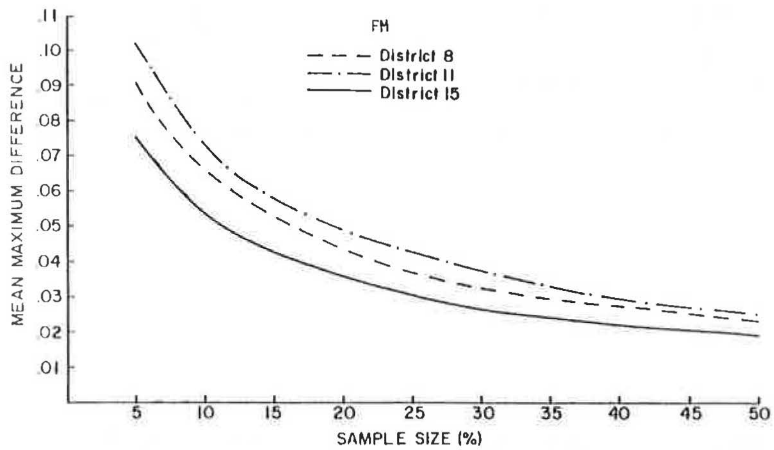

FIGURE 7 Comparison of district results of the mean maximum difference using county stratification Method A, farm-to-market system.

TABLE 5 Percentage Error in Average Cost per Square Yard with County Stratification

System Sample Size (%)	I	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

- When relations have been established between mean pavement score and the variance of pavement scores, the percentage of pavements requiring no costs and average costs per square yard, as illustrated in Figures 3-5, can be estimated if an accurate estimate of the mean score is known. A 5 percent sample can give such an estimate of the mean score.
- The use of a 5 percent sample to estimate the parameters of a beta distribution can produce reasonably accurate estimates of the percentage of pavements falling below a minimum acceptable score.
- Although the details of the distress and riding quality of individual pavement sections can only be found by observation of the sections themselves, estimates of average costs and pavement network condition distributions, which are used in network-level pavement management, may be achieved with a 5 percent sample if the relations referred to previously have been established.
- Although the data analyzed here are for Texas conditions, costs, and pavement rating methods, the same general approach may be used by any other state, with the result that more efficient use can be made of state highway agency manpower resources in surveying the condition of the state highway network.
- The level of accurate detail that is desired as a result of a statewide condition survey will dictate the percentage of the pavement sections that should be sampled. From the observations made in this study, the percentage can range from 5 percent up to 40 or 50 percent, depending on the type of information desired and the level of accuracy that is

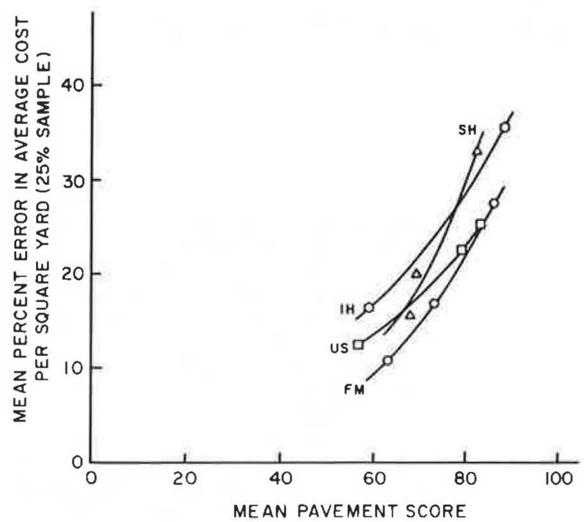

FIGURE 8 Relation between mean pavement score and mean percentage error in estimated average cost per square yard.

required. This paper illustrates typical results that can be expected.

REFERENCES

1. T. Scullion and R.L. Lytton. The Development of the RAMS-State Cost Estimating Program. Research Report 239-6F. Texas Transportation Institute, Texas A&M University System, College Station, Nov. 1984.
2. J.A. Epps, A.H. Meyer, I.E. Larrimore, and H.L. Jones. Roadway Maintenance Evaluation User's Manual. Research Report 151-2. Texas Transportation Institute, Texas A&M University System, College Station, Sept. 1974.
3. W.J. Conover. Practical Nonparametric Statistics. John Wiley and Sons, Inc., New York, 1971.
4. R.V. Hogg and A.T. Craig. Introduction to Mathematical Statistics. MacMillan Publishing Co., Inc., New York, 1978.
5. 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 System, College Station, 1978.
6. C.J. Templeton and R.L. Lytton. Estimating Pavement Condition and Rehabilitation Costs Using Statistical Sampling Techniques. Research Report 239-5. Texas Transportation Institute, Texas A&M University System, College Station, June 1984.

Publication of this paper sponsored by Committee on Pavement Management Systems.

Optimal Pavement Management Approach Using Roughness Measurements

BENJAMIN COLUCCI-RIOS and KUMARES C. SINHA

ABSTRACT

Many state highway departments are placing major emphasis on the development of cost-effective procedures for maintaining their existing pavement network. The state of Indiana is in need of a systematic procedure for allocating Interstate resurfacing funds to its pavement network. An optimization procedure for establishing resurfacing priorities at the network level, which can be incorporated in a pavement management system for the state, is described. Roughness measurements, increase in roughness over time, and traffic are the primary factors considered in the optimization scheme. Different types of resurfacing activities are considered in the model. A performance function model was developed to relate resurfacing strategies to the overall reduction in pavement roughness present in the pavement section just before resurfacing. Regression equations based on roughness measurements were also developed in this study for predicting future roughness levels. The optimization model has the capability of considering deficient pavement sections at any point within the specified analysis period. In addition, it has the capability of analyzing the impact of different budget scenarios. The model, in its present format, can predict what pavement section and resurfacing strategy combination should be adopted in order to achieve an optimal resurfacing program in Indiana during the next 5-year period. The application of the optimization model to the Indiana Interstate highway network is discussed in the paper.

The Indiana Department of Highways (IDOH) through the Research and Training Center (R&TC) has been collecting pavement roughness measurements on a continuing basis for the entire highway system since 1979. These data are summarized annually along with other information including average daily traffic (ADT) in one direction, surface type and texture, contract number, length, and last time a major rehabilitation was performed. This information currently forms the basis for most of the decisions about major rehabilitation, primarily for the Interstate system. Although this information is useful in identifying pavement sections that exceed the minimum acceptable values established by the state in any given year, it is not useful in the process of selecting those miles that have the greatest need given a constraint on the amount of money available for major rehabilitation. For an effective management approach, it is necessary to have a mathematical model that can answer questions such as (1,2)

1. Which specific pavement contract sections as well as how many miles of roads should be rehabilitated during a given year or during the time frame specified with available budget?
2. What type of resurfacing strategy should be applied to the pavement contract section selected in order to use the total available budget in the most cost-effective manner?
3. How many additional lane-miles can be improved if the budget is increased by a certain percentage?
4. How much additional budget is required to upgrade the pavement condition of the entire network (or a part of it) to a minimum acceptable level?

In the following paragraphs a systematic procedure, which uses a mathematical model for allocating maintenance and rehabilitation funds to existing pavements within the state of Indiana, is described. The results of applying the mathematical model to