SOME USES OF THE STUDENT t AND CHI-SQUARE TESTS IN MAINTENANCE COST CALCULATIONS

Mathew J. Betz, Department of Civil Engineering, Arizona State University

•THIS PAPER discusses some of the uses of two statistical techniques on particular problems in highway maintenance cost evaluations. This is not intended as a thorough statistical presentation of the theory of the Student t and chi-square tests. This information can be obtained from any good statistical textbook. On the other hand, it is not meant to be an extensive compilation of example problems. It is hoped that it falls somewhere between the two, thus giving the engineer involved in highway maintenance cost calculations a general feeling for the statistical tests and a few examples that would allow him to use the statistical tests in cases that are common in highway engineering. The author has attempted to select problems that are common in this field and does not suggest that this paper covers all of the possible uses of these techniques.

Moreover, this paper should be taken as one of a series concerned with the application of statistics to highway maintenance costs. Thus, the paper will assume a knowledge of some elementary statistics. This includes a knowledge of statistical distributions, more specifically the normal distribution; a knowledge of the mean (the first moment about the origin) and other measures of central tendency; a knowledge of the variance (the second moment of the distribution about the mean) and the standard deviation; some basic concepts of probability; and the use of simple regression including th concept of the coefficient of correlation.

CHARACTERISTICS OF HIGHWAY MAINTENANCE COSTS

Although expenditures for the maintenance and renewal of highways represent a substantial proportion of total highway expenditures, until quite recently relatively little attention had been given to the components of these costs and the analysis of their variability. It has only been in the past few years that scientific, comprehensive studies of highway maintenance have been undertaken, and the basic independent variables that affect highway maintenance have yet to be fully evaluated.

In general, highway maintenance operations and costs are difficult to analyze because of the large number of different physical operations involved, which range all the way from the patching and resurfacing of pavement surfaces to the cutting of grass and maintenance of rest areas. In addition, these operations are highly dispersed in time and place. This means that the majority of operations are relatively small in size and may be highly seasonal. Compounding the problem further is the fact that there are a large number of variables that may directly affect the various maintenance needs on a specific roadway. This in turn affects the cost. At least three major categories may be defined. The first category is the highway design characteristics, which include number of lanes, lane width, pavement design criteria, topography, etc. The second is the traffic characteristics, which include the average daily traffic volume, the percentage of trucks, the maximum wheel loads, etc. Last, there is the question of environmental characteristics. These could include temperature, precipitation (and a combination of these two), depth of frost, etc. All of these factors may affect the overall maintenance costs of the roadway. When the individual maintenance operations are evaluated separately, each of the independent variables has a different relative effect. In addition, many are closely interrelated.

Sponsored by Committee on Maintenance Costs and presented at the 48th Annual Meeting.

Although substantial highway maintenance data are now being accumulated, it is still very difficult to obtain much data where all of the parameters except the one of interest can be held constant. Thus, for most analyses the data sample size will be small. There are many cases where data that will measure the effect of only one parameter are not available.

The application of the Student t and chi-square tests is suitable in those cases where the sample size is relatively small. Therefore the use of these tests in determining statistical significance is a powerful tool in the analysis of specific maintenance parameters. They can be applied to maintenance costs of roads of different surface type, from different environmental areas, from different maintenance districts (this might indicate inefficient operation or necessary corrective measures in a particular district), or with different traffic characteristics.

USEFUL CONCEPTS

Before discussing the tests and numerical examples, it is useful to establish a few basic concepts and definitions that will be used later. The first of these is the difference between a population and a sample. A population is all of the possible items in the set under consideration. Thus, theoretically the population mean and the population variance could only be evaluated if a 100 percent count were made of all of the individuals within that population. In all highway and traffic engineering work this is a virtual impossibility because of economic restraints. Thus, in the real case the analyst must be satisfied with a sample. A sample is a selection of individual items out of the population. The reliability of the sample depends on the size of the sample. All engineers are at least intuitively aware of this in that they will put more reliance on figures based on several samples rather than on a figure based on only one observation. Thus, the parameters of the sample (the mean, variance, and other statistical values) tend to approach the corresponding parameters of the population as a sample size increases. In many cases where the sample is greater than 30 to 50 items the sample means are, for all practical purposes, equal to the population means. However, as previously indicated, very small samples may have to be used in the evaluation of highway maintenance costs.

Another useful fact that has application later in this paper is that the distribution of sample means taken from a population approaches a normal distribution no matter what the original population distribution function. For example, if a population were composed of 100 cards an equal number of which each have the numbers 0 to 9 printed on them, then the distribution of the population would be rectangular, i.e., each number would have $\frac{1}{10}$ of the total. If samples of two or more are selected from this population, it is clear that the mean of the samples would be the same as the mean of the population (in this case $4\frac{1}{2}$); however, sample means such as 0 or 9 would be quite unusual and values near the mean would be more common (Fig. 1). It can be shown that the distribution of these means is in fact a normal distribution. Furthermore, if the population function is skewed to one side or the other, the distribution of the means will also be skewed, but decidedly less so. Thus, a normal distribution can be used to compare sample means taken from a population.

The Student t and chi-square tests are used to determine whether there is a statistically significant difference between samples, between samples and populations, and between other statistical indices and some fixed value. This is accomplished by establishing a hypothesis and accepting or rejecting it. As explained in the following paragraphs, this is normally accomplished through consideration of the null hypothesis. This procedure assumes that the values are equal or from the same population and then attempts to reject this assumption. This paper will emphasize the rejection of a null hypothesis as the strongest statistical statement that can be made. The reason for this requires a knowledge of the types of errors that are inherent in the acceptance or rejection of a hypothesis.

There are two basic errors that are possible in rejecting or accepting a hypothesis. The first of these is called a type I error, which is the rejection of a hypothesis when it is true. This is associated with the alpha risk, where alpha is the probability that the error could occur. As with all statistical techniques, it is impossible to reject a

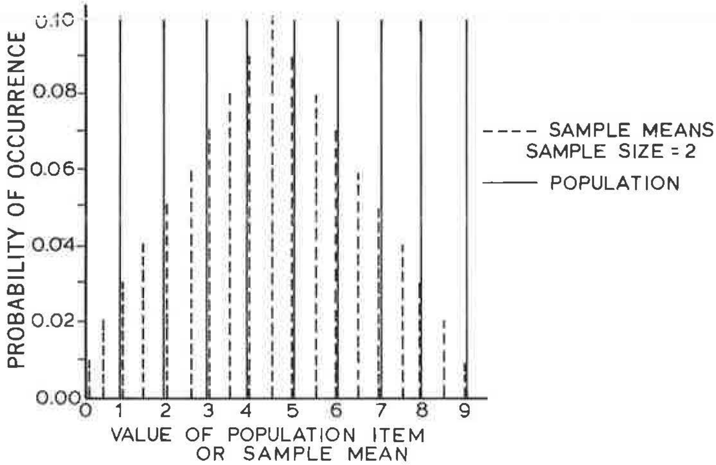

Figure 1. Distribution of a population and a mean therefrom.

hypothesis with 100 percent certainty. For example, if the acceptable alpha is equal to 0.05, the analyst is stating that he will reject the hypothesis realizing that 5 times out of 100 he will reject something that is in fact true.

The second error is the type II error, which is the acceptance of a hypothesis when it is not true. This is associated with the beta risk, where beta is the probability of this occurrence.

Thus, if one were to test whether a sample mean came from a particular population, the strongest position that could be taken would be to establish the null hypothesis (that they did in fact come from the same population) and then statistically reject this hypothesis. This involves only the selection and evaluation of the alpha risk because rejection of the hypothesis that they are the same involves only a type I error. If, on the basis of an acceptable alpha risk, one cannot indicate that the items under consideration are significantly different, then one is left with the possibility of accepting the hypothesis that they are the same. A consideration of the errors described previously would indicate that this is not as simple as the rejection of the hypothesis. The value of the beta risk must be considered. In other words, the probability of saying that they are the same when in fact they are not the same must be considered. Generally when one is attempting to reject the null hypothesis, one will select a relatively small alpha risk to ensure that his rejection will be true in the vast majority of cases. For example, an alpha of 0.05 to 0.10 is not uncommon. However, as the acceptable alpha risk is decreased in size, the beta risk increases for a sample of a given size. Although a null hypothesis may not be rejected at the 5 or 10 percent level, acceptance of the hypothesis may involve the acceptance of a substantial beta risk (often greater than 0.50) because the value of the beta risk can approach $1-\alpha$. Thus, acceptance of a hypothesis may not be desirable because of the danger of accepting something that is not fact.

THE STANDARDIZED NORMAL DISTRIBUTION

Any normal distribution can be transformed into what is referred to as the standardized normal distribution by using the following transformation:

$$Z_i = \frac{X_i - \mu_x}{\sigma}$$

where

Z_i = a transformed value for the standardized normal curve,
X_i = a value for the normal curve,

μ_x = population mean for normal curve, and
σ = standard deviation for normal curve.

Such a distribution will have a mean value of zero and a standard deviation around that mean equal to 1.0. All normal distributions can be transformed in this way. By using standardized tables for this transformed function, comparisons can be made that require the use of the area under part of the distribution function. Although not covered in this report, this procedure can be used to establish confidence limits about the mean of a population.

The standardized normal distribution is presented here because it may be used to solve a type of problem that occurs in the comparison of maintenance costs, which is similar in nature ot the problems that can be solved using the Student t test. The particular problem is the comparison of a sample mean to a population mean when the population standard deviation (σ) is known. The test statistic from the standardized normal distribution is formed by the calculation of the following quantity:

$$\frac{\overline{X} - \mu}{\sigma} (\sqrt{N})$$

where

$\overline{X}$ = sample mean,
μ = population mean,
σ = population standard deviation, and
N = sample size (number of observations).

Normally, in maintenance calculations one is concerned with both abnormally high and low figures, and thus the areas under both of the extreme tails of the distribution curve need to be considered. One can use a table of standard values and obtain the probability that the mean does in fact come from the population. Thus, if the value obtained from the table is 0.28, then 28 times out of 100 one could expect to get this value of sample mean or larger from the given population. If the probability is 0.05, then only 5 times out of 100 would this be expected. Such a low probability might lead the analyst to conclude that it did not in fact come from the population. As indicated previously, the normal technique would be to establish the null hypothesis and an acceptable alpha risk and to attempt to reject the null hypothesis. The following example indicates the use of this technique.

Suppose that data collected from a large number of locations indicate that the annual cost to maintain a 2-lane rural road is $1,200 per mile and that the standard deviation is 100. A sample of five maintenance sections is taken in an adjoining state that has a mean of $1,308 per mile per year and a standard deviation of 233. Is this sample mean significantly different from the mean of the first state (could they represent the same maintenance cost function)? An alpha risk of 0.05 or less can be accepted. Thus,

$$\frac{\overline{X} - \mu}{\sigma} (\sqrt{N}) = \frac{1{,}308 - 1{,}200}{100} \sqrt{5} = 2.40$$

From a standard table for a two-tail test, it is found that the probability that a sample mean of this value or larger could be obtained from this population is 0.016. Because this is less than the acceptable alpha risk, the null hypothesis can be rejected and we can state that the sample mean is significantly different from what would be expected from the population.

THE STUDENT t TEST

In many cases the standard deviation of the population is not known. The sample and its parameters are available. The hypothesis to be evaluated is whether this sample mean is significantly different from a population mean where the population standard deviation is not known. To accomplish this, an estimate of the value of the population standard deviation must be constructed from the standard deviation of the sample. The accuracy of this estimate is, naturally, a function of sample size. Therefore, a comparison using the standardized normal distribution as described previously might be misleading. The distribution used in this case is the Student t distribution. This is a distribution that is symmetrical around a mean, which has a value of 0 similar to the standardized normal. However, the distribution is a function of the degrees of freedom, which in turn is related to the sample size. The function approaches the standardized normal distribution as the degrees of freedom increase. At approximately 30 degrees of freedom the two functions are the same. At smaller values the Student t distribution is wider at the base than the standardized normal. The Student t distribution has several applications in the analysis of maintenance cost data.

As already indicated, the first of these is the comparison of a sample and population mean when the population standard deviation is not known. The statistic in this case is as follows:

$$t_{(N-1)} = \frac{(\bar{X} - \mu)\sqrt{N}}{S}$$

where

$$S = \text{standard deviation of the sample} = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N-1}}$$

$t_{(N-1)}$ = Student t statistic with N - 1 degrees of freedom.

Again, the normal procedure is to establish the hypothesis that the two means are equal, to select an alpha risk, and then to attempt a rejection of the hypothesis. Standard tables of the value of the t statistic are available that present the statistic as a function of the level of significance and the number of degrees of freedom.

An example of the application of this technique may be illustrated using essentially the same data as in the previous example. In this case the population mean will still be considered as $1,200 per mile per year, but it will be assumed that the population standard deviation is unknown. The sample values will also remain the same. Because the sample size is 5, the t distribution has four degrees of freedom:

$$t_{(N-1)} = \frac{(\bar{X} - \mu)\sqrt{N}}{S} = \frac{(1{,}308 - 1{,}200)\sqrt{5}}{233} = 1.04$$

From the standard table of t values it is established that the probability of the sample mean coming from the population is greater than 0.35. Thus, the null hypothesis cannot be rejected, and the sample mean is not significantly different (at $\alpha = 0.05$) from what would be expected from the population. The statistic would have to exceed a value of 2.776 for there to be a significant difference at $\alpha = 0.05$.

Another application of the Student t distribution that is probably even more common to highway maintenance cost analysis is the determination of a significant difference between two samples means where their means and standard deviations are known but nothing is known about the population. In this case the t statistic to be used is the difference between the means divided by an estimate of the standard error of the difference between the two sample means. This may be written as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

where the hypothesis is that the samples come from populations with equal means and variances X_1, X_2 = sample means and $\sigma_{\bar{X}_1 - \bar{X}_2}$ = standard error of the difference between the two sample means.

Because small samples are involved the population means are unknown, and therefore an estimated standard error of the difference of the means must be calculated based on the sample deviations and the sample sizes. Thus, the statistic may be rewritten as follows:

$$t_{(N_1 - 1 + N_2 - 1)} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 - 1 + N_2 - 1}} \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

where

$t_{(N_1 - 1 + N_2 - 1)}$ = the t statistic with $N_1 - 1 + N_2 - 1$ degrees of freedom,
N_1, N_2 = sample sizes, and
S_1, S_2 = sample standard deviations.

In addition, the use of the foregoing formulation of the statistic assumes that the samples are independent. This would normally be the case for highway maintenance cost samples. The t distribution can also be used for nonindependent samples, and any statistics text will give the formulation of the t statistic in this case.

For example, assume that a highway maintenance cost sample was taken from another state with roads having characteristics similar to the previsously used sample. This second sample has a mean of $1,230 per mile per year and a standard deviation of 162. The sample size is five. (The sample sizes need not be equal.) The hypothesis is that the two sample means come from the same population. The α value is set at 0.05. Thus,

$$t_{(5 - 1 + 5 - 1)} = \frac{1{,}230 - 1{,}308}{\sqrt{\frac{5(162)^2 + 5(233)^2}{5 - 1 + 5 - 1}} \sqrt{\frac{1}{5} + \frac{1}{5}}}$$

$$= \frac{78}{\sqrt{50{,}188}\ \sqrt{0.40}} = \frac{78}{224(0.633)} = 0.55$$

Using a table of t values for 8 degrees of freedom, it is established that these could be expected to come from the same population more than 50 percent of the time. Thus, the null hypothesis cannot be rejected at $\alpha = 0.05$. The value of the statistic would need to exceed 2.262 in order to reject the null hypothesis at that alpha level.

The last application of the Student t test to be discussed here is somewhat different from the preceding. The t distribution can be used to evaluate if a correlation coefficient, r, in simple correlation is significantly different from zero. The question is often raised in correlation analysis as to the satisfactory value of the correlation coefficient. This is a most difficult question to answer and probably varies depending on the type of problem and sources of data. In many cases the correlation coefficient that would not be acceptable in measuring physical properties might be highly acceptable in measuring socioeconomic relationships. This is one case where the rejection of the null hypothesis may not be as meaningful to the analyst as the inability to reject the hypothesis. Again, the analyst must be cautioned that the inability to reject the hypothesis does not necessarily mean the acceptance of it. More specifically, then, the test can be used to reject the null hypothesis that the coefficient or correlation is equal to zero. If this is rejected, it still does not answer the question as to whether the correlation coefficient is acceptable in a given analysis. However, it would seem that if it cannot be rejected, i.e., if it cannot be said that the coefficient or correlation is different from zero, then this would cast considerable doubt on the meaningfulness of the relationship between the dependent and independent variables.

The Student t test can be used to test the difference from zero in simple correlation for linear or nonlinear relationships. In multiple correlation, a similar hypothesis test can be conducted using the F test, which is not discussed in this paper. The testing of partial correlation coefficients can be accomplished by the t test. In each case, the statistic is somewhat different. The only example included here is where simple linear regression is considered. In this case, the t statistic is as follows:

$$t_{(N-2)} = r\sqrt{\frac{(N-2)}{1-r^2}}$$

where

$t_{(N-2)}$ = the t statistic with N - 2 degrees of freedom,
r = simple correlation coefficient, and
N = number of points (observations) in correlation analysis.

Figure 2 shows a plot of maintenance effort as a function of average annual daily traffic. Also included is a plot of the trend line established by conventional correlation analysis. The value of the correlation coefficient for these data is 0.721. An observation of the trend line and data would indicate that no strong trend is evident. It is possible that a number of analysts, because of the r value, might accept this as a significant indication of relationship. Part of this may be due to a built-in bias on the part of the analyst in that he feels there should be a relationship. Thus, an evaluation of the significance of the correlation is of interest. The statistic is

$$t_{(7-2)} = 0.721\sqrt{\frac{(7-2)}{1-0.52}} = 2.34$$

To reject the hypothesis with an alpha value of 5 percent or less, the statistic would need to equal or exceed 2.571. Therefore, in this case, the analyst cannot state that this correlation is significantly different from zero. If a type I error were allowed 10 times out of 100, then it could be stated that the correlation coefficient is significantly different from zero for the standard tables to show that the statistic need only exceed 2.015.

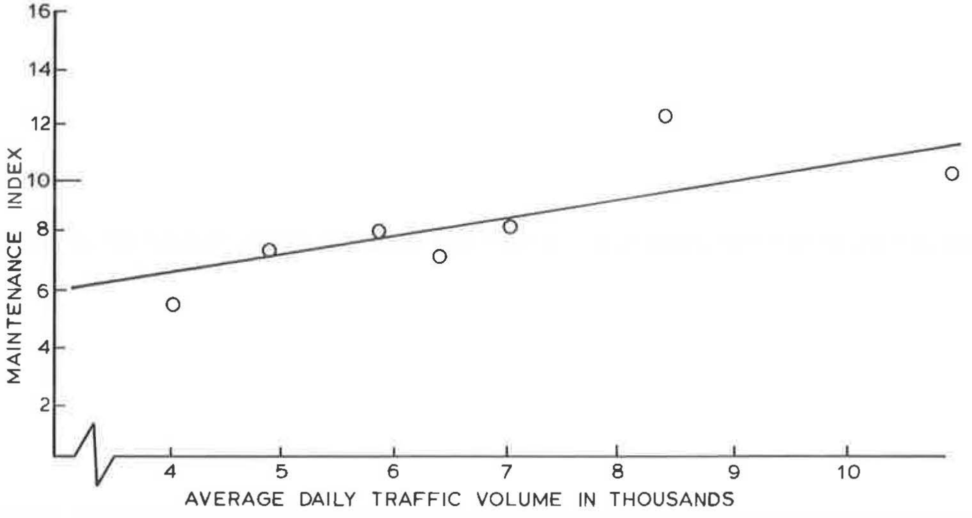

Figure 2. Maintenance effort versus traffic volume.

THE CHI-SQUARE DISTRIBUTION

The chi-square distribution is the function that would be obtained when values of sample variances are repeatedly drawn from a normal distribution. Naturally, this distribution function is related to sample size as was the Student t distribution. The chi-square statistic is normally proportional to the square of the difference from the sample value divided by the expected value. The distribution itself is skewed to the right at lower degree of freedom values. As the value of the degree of freedom increases, the function approaches symmetry around a mean that has a value of the number of degrees of freedom. Therefore, a chi-square distribution with 15 degrees of freedom is fairly symmetrical around the value 15.

Because the chi-square distribution represents the expected variation of sample variances taken from a population, it can be used in a manner similar to the standardized normal curve for evaluating the significant difference between a sample variance and a population variance when the population variance is known. The chi-square statistic in this case is as follows:

$$\chi^2 = \frac{(N-1)S^2}{\sigma^2}$$

This type of problem is not common in the statistical analysis now generally used for maintenance cost data. It is more useful for the study of the range of errors that may be expected rather than the direct evaluation of significant differences between samples and populations. Therefore, no example is included.

However, the chi-square test can also be used where the expected frequency of occurrences may be developed through probability. This permits its use in an area that is much more meaningful to highway maintenance costs and to a much broader range of problems, including the evaluation of the performance of maintenance crews and equipment. The procedure is to establish the theoretical frequency of occurrences and then to establish the null hypothesis that the actual observed occurrences come from the same distribution. Again, an acceptable alpha risk will need to be established. The application of the technique and the details of calculation are probably best explained through the use of an example.

One may wish to compare several brands of equipment, each of which is supposedly designed for the same job and the same capacity. The index of equipment maintenance could be the number of hours of maintenance over a given time span. Model A needed 9 hours of maintenance whereas model B needed 23, model C needed 20, and model D needed 16. The null hypothesis is that this distribution could have come from a population where each had equal maintenance. The χ^2 statistic is

$$\chi^2_{(N-1)} = \sum \frac{(f - f_e)^2}{f_e}$$

where

$\chi^2_{(N-1)}$ = chi-square statistic for N - 1 degrees of freedom,
N = number of items (rows),
f = actual value, and
f_e = expected value (note: value should exceed five).

Table 1 gives the data for calculation of the statistic.

If the alpha risk were established at 0.05 or smaller, the χ^2 statistic would need to be 7.815 or larger to reject the null hypothesis. Therefore, the analyst cannot say that this distribution is significantly different from one where all units have equal maintenance requirements.

If one would accept an alpha of 0.10, standard tables would indicate that χ^2 need only exceed 6.251, for three degrees of freedom, to reject the null. Therefore, in one were willing to be wrong 1 in 10 times, one could say this distribution is significantly different

TABLE 1

STATISTIC DATA

Model	Actual Value f	Expected Value* f_e	$f - f_e$	$(f - f_e)^2$	$\frac{(f - f_e)^2}{f_e}$
A	9	17	-8	64	3.76
B	23	17	6	36	2.12
C	20	17	3	9	0.53
D	16	17	-1	1	0.06
Σ	68	68			6.47

*The expected number of hours of maintenance if all models required an equal amount.

TABLE 2

CALCULATION OF χ^2

Operation	f	f_e	$f - f_e$	$(f - f_e)^2$	$\frac{(f - f_e)^2}{f_e}$
Tire	70	54	16	256	4.74
Other minor	60	72	-12	144	2.00
Motor	25	36	-11	121	3.36
Other major	25	18	7	49	2.72
Σ	180	180			12.82

from a case where all units could be expected to have equal maintenance. Further analysis could indicate that the difference is due to model A. This would indicate that A is better than the others at some level of alpha.

In the preceding example, each item was assumed to be equivalent to the other and therefore to have an expected frequency equal to the others. This assumption is not necessary for the application of the chi-square test to this type of problem. The following example indicates the application of a case where this is not true.

Assume the following: When a given type of equipment needs maintenance, previous experience has shown that 30 percent of the time it is for tire or track repair, 40 percent for other light maintenance, 20 percent for motor removal and overhaul, and 10 percent for other major repair. In a particular job location, the frequencies have been 70, 60, 25, and 25, respectively. The null hypothesis is that these frequencies come from a population with the historical distribution. The calculation of χ^2 is given in Table 2.

The standard χ^2 table indicates a rejection of the null hypothesis with $\alpha = 0.05$ and 3 degrees of freedom when the statistic exceeds 7.815. Therefore, the null hypothesis can be rejected in this case. In fact, it can be rejected with an alpha of less than 0.01 (i.e., with a chance of error of less than 1 in 100). Thus, there is something unique about the job or location that may need investigation.

There are innumerable maintenance and maintenance cost problems that can be structured in the same way. For the analysis of maintenance costs, the use of chi-square is probably the most important of all the tests discussed in this paper.

CONCLUSIONS

This paper has tried to indicate briefly some applications of the chi-square and Student t tests to highway maintenance problems. It has discussed the establishment of hypotheses and the types of errors involved. Several examples of different types of highway maintenance problems have been presented. It is hoped that this paper will interest highway maintenance analysts in the use of these statistical techniques in the evaluation of their data.

ACKNOWLEDGMENTS

The author would like to thank Mr. Robert L. Lynch for his help during the preparation of this report and Dr. O. B. Moan for his review of the draft document.

REFERENCES

1. Betz, Mathew J. Highway Maintenance Costs—A Consideration for Developing Areas. Highway Research Record 94, pp. 1-27.
2. Bowker, Albert H., and Lieberman, Gerald J. Engineering Statistics, 7th Printing. Prentice Hall, Englewood Cliffs, N. J., 1965.
3. Byrd, L. G. Interstate Highway Maintenance Requirements. Meeting Preprint 594, ASCE National Meeting, San Diego, Calif., Feb. 1968, 9 pp.
4. Croxton, Frederick E., and Cowden, Dudley J. Applied General Statistics. Prentice Hall, New York, 1955.
5. Davis, D. W. Highway Maintenance. Proc. 2nd North Carolina Highway Conference, 1960, pp. 40-43.
6. Dixon, Wilfrid J., and Massey, Frank J. Jr. Introduction to Statistical Analysis. McGraw-Hill, 1957.
7. Greenshields, Bruce D., and Wyeda, Frank M. Statistics With Applications to Highway Traffic Analysis. Eno Foundation, Saugatuck, Conn., 1952.
8. Jorgensen, Roy, and Associates. Highway Maintenance Operations in Virginia. Virginia Maintenance Study, Jan. 1966.
9. Leonhard, William E. A Study of the Effect of Income on Trip Characteristics for Tucson, Arizona. Unpublished thesis, Arizona State University, June 1968.
10. Schwar, Johannes F., and Puy-Huarte, Jose. Statistical Methods in Traffic Engineering. Ohio State University, Engineering Experiment Station, Spec. Rept. 26, 1962.
11. Tallamy, Bertram D., Associates. Interstate Highway Maintenance Requirements and Unit Maintenance Expenditure Index. NCHRP Rept. 42, 1967, 144 pp.