Effect of Number of Test Specimens on Test Results on Slag-Lime-Flyash Mixtures

GEORGE W. HOLLON, Assistant Professor of Civil Engineering, and ELLIS DANNER, Assistant Professor of Civil Engineering and Professor of Highway Engineering, respectively, University of Illinois, Urbana, Illinois

This paper sets forth a scientific approach which may be used in evaluating new highway materials. This approach has been used in the evaluation of slag-lime-flyash mixtures at the University of Illinois in cooperation with the Illinois Division of Highways and the Bureau of Public Roads. Wet-bottom boiler slag and flyash are two by-products of the combustion of powdered bituminous coal. These materials have been combined with lime and water to produce a mixture that may be used as a base course for bituminous surfaces.

Numerous laboratory techniques and testing methods have been developed that will be of value to those who may be concerned with the analysis of this or similar mixtures in the future. This paper is concerned with the importance of relating the number of test specimens prepared to the variability of the laboratory techniques employed, in order that the significance of test results may be determined. If this approach is followed the laboratory test results may be interpreted with a known degree of reliability. Any conclusions drawn from the test data will therefore be of scientific value to all concerned.

NATIONAL INTEREST

● THE MAGNITUDE of the present highway program makes it imperative that adequate scientific methods for evaluation, design, and construction of highway pavements be developed and put into use. The scientific approach must be used in the evaluation of new materials which are not currently being used in the highway field in order that these new products may be effectively incorporated in the production of strong, durable, and economical highways.

Highway engineers are continually confronted with the problem of improving the load carrying capacity of natural soil. Numerous methods of soil stabilization have been used in the past and new methods are continually being developed. The basic methods of stabilization may be classified as follows: (a) mechanical-improving the aggregate gradation; (b) compaction-increasing density; (c) bituminous-waterproofing and giving additional cohesion; (d) cementation; and (e) chemical. Methods of stabilization which satisfy more than one of the above functions are highly desirable. In order to determine whether a material will serve any or all of the above functions it is necessary to develop a laboratory test program which will give test results that are significant and that are truly measures of the desired properties. After the laboratory tests have been obtained it will no doubt be desirable to test the best materials under actual field conditions in order that information may be obtained regarding the workability of the material as well as the performance under field conditions.

One method of stabilization, lime-pozzolan stabilization of natural soils, has received considerable attention during the past decade. A pozzolanic material is defined by ASTM as a silicious or alumino-silicious material which in itself possesses little or no cementitious value but which in finely divided form and in the presence of moisture will chemically react with alkali and alkaline earth hydroxides at ordinary temperatures to form or to assist in forming compounds possessing cementitious properties.

Campaigns to clean up smoky cities have resulted in large quantities of flyash being precipitated from the smoke. Some of the flyashes have been found to possess physical and chemical properties that meet the above requirements for a pozzolanic material. Since large quantities of flyash are presently available at a nominal cost, it is logical that uses for this material be investigated.

The problem of developing a laboratory program which will give test results that are significant and that truly measure the desired properties of the material is a very challenging and interesting task.

SLAG-LIME-FLYASH MIXTURE STUDY

In the state of Illinois there are numerous plants producing power by burning pulverized bituminous coal. Two by-products of this burning process are wet-bottom boiler slag and flyash. Wet-bottom boiler slag is a dense, hard, angular material and is not to be confused with blast furnace slag. Wet-bottom boiler slag is known also as black slag. These waste products are presenting large disposal problems in the state of Illinois, since over 1½ million tons of the two materials are produced annually.

In some Illinois counties the supply of aggregate for base courses is becoming limited, and suitable substitute materials that may be economically used are highly desirable. Since slag and flyash are primarily waste products, the economic consideration in localized areas where they are available and where the usual base-course aggregates are scarce will favor their use—if mixtures containing these materials have the required physical properties.

The original purpose of the slag-lime-flyash mixtures study was to evaluate the merits of the slag-lime-flyash mixture in regard to its possible use as a base course material with a bituminous surface. The general requirements were adequate workability, strength, durability, and ability to be produced at a reasonable cost. This paper is not a report on the research project in its entirety, but describes only a scientific approach that was used to obtain reliable and meaningful laboratory test results.

EARLY TESTING PROBLEMS

Since this was a new material there were no standard ASTM tests which could be performed to determine whether the mixture was acceptable. Therefore, methods of evaluating the mixture had to be developed. This necessitated the development of a laboratory procedure that would reveal the true nature of the mixture. A preliminary study was required to answer some of the early testing problems in order that the variability of the mixture could be determined.

Compressive strength was chosen as the basic test to evaluate the variability of the mixture because this test was easy to run and it measured the combined effect of cohesion and internal friction of the mixture. Since the load carrying capacity of a material consists primarily of cohesion and internal friction, a comparison of the compressive strengths should indicate the most promising mixture. The remainder of this report presents the procedure that was used in limiting the variables involved in compressive strength tests within reasonable limits and in determining the number of test specimens required to provide the results within an acceptable degree of accuracy.

One of the early questions was the determination of the size of specimen to be used in testing. After evaluating research procedures used by others on similar materials a 2- by 4-in. cylindrical specimen was selected for use in the investigation. Until the effect of mixing time was investigated it was necessary to arbitrarily fix the mixing time. Densities of the 2- by 4-in. specimens were equated to maximum density by determining a compactive effort that would produce the same density as that produced by the standard compaction test (AASHO Designation T99-57, Method A). The initial number of test specimens was determined by the capacity of the mixing equipment. Additional questions were raised in regard to capping, removing, curing and testing specimens. Initially obtained results, although indicating high compressive strength, were not consistent enough to enable the data to be of scientific use for evaluation purposes.

It was on the basis of this preliminary study that the research project was developed in order that some of the above questions could either be resolved or eliminated by extreme care in the laboratory procedure. This large variation simply necessitated an analysis of all known variables and the determination of the effect of both the inherent variables of the materials and the variables produced by the testing procedures. This approach will enhance the development of standard procedures and will also enable the development of significant test results for a given degree of reliability.

LABORATORY TECHNIQUES AND EFFECT ON TEST RESULTS

Sampling and Preparation of Materials

In this test program numerous samples of the raw materials were obtained in accordance with the ASTM D75-48 Specification. Whenever possible, the size of sample was large enough to complete one phase of the testing. As the testing progressed, it was evident that different materials obtained from the same plant at different times had different characteristics.

In determining the size of the test specimen it was anticipated that several specimens would be required to give an average compressive strength value that would be significant. It was desirable to use as small a specimen as possible to prevent the handling of extremely large quantities of the materials. Since more than 95 percent of all the raw materials passed the No. 4 sieve, the 2- by 4-in. cylindrical mold was used. This size of specimen also eliminated any L/D corrections since the ratio was two. In the laboratory testing, all material larger than ½ in. was removed from the sample and discarded. In order to minimize physical variation in the flyash and slag, the total sample of each material was mixed separately in a pug mill mixer for 45 minutes after which it was placed directly in its storage container.

The lime was obtained in sealed metal drums and transferred to 1 gal paint cans which had an air-tight seal. Type No. 3 Permatex was used to coat the lids as an added precaution to assure a completely air-tight seal. It was found that lime must be handled with care in the laboratory to prevent it from carbonating. Carbonation is the formation of calcium carbonate due to the reaction between lime and carbon dioxide from the air. The calcium carbonate is actually inert and does not combine to form calcium silicates. In the preliminary study

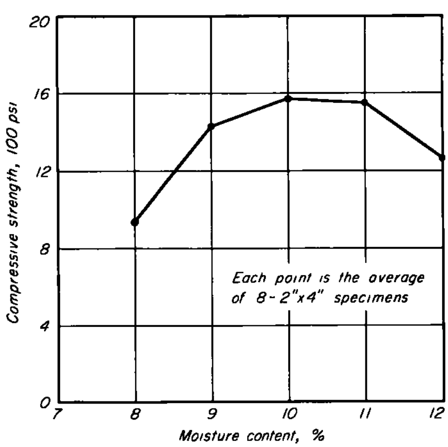

Figure 1. Relationship of compressive strength to moisture content for 72.1 percent slag, 24.0 percent flyash, and 3.9 percent lime mixture.

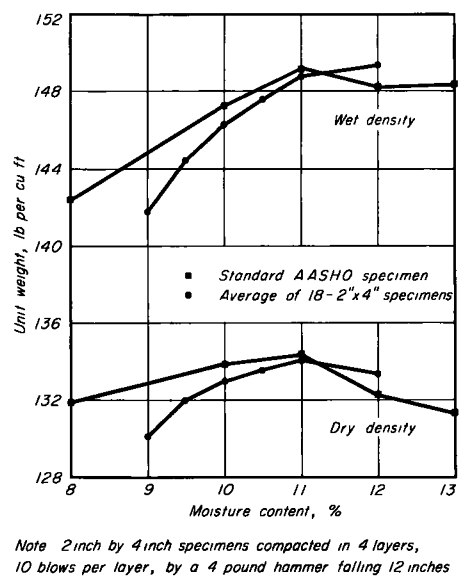

Figure 2. Relationship of standard AASHO density to 2- by 4-in. specimen density for 72.1 percent slag mixture.

mentioned previously, the lime was left exposed to air in an open bag, and it was determined by subsequent tests that carbonation had occurred. The degree of carbonation depended on the interval of time that the lime was exposed. The carbonation does not extend more than a fraction of an inch into the lime, but this is sufficient to cause the laboratory results to be inconsistent since a very small quantity of lime is used in making the cylinders.

Proportioning of Materials

The major problem under investigation in this research was the determination of the effect of various proportions of slag, lime, and flyash on compressive strength. In order to evaluate the various combinations of slag, lime, and flyash, the moisture content had to be standardized. The effect of moisture content on compressive strength is shown in Figure 1. The relationship between unit weight and moisture content for both the AASHO compaction specimens (4-in. diameter) and 2- by 4-in. specimens is shown in Figure 2. Since the optimum moisture content from the standard AASHO compaction test corresponded so closely with the moisture content which gave the best compressive strengths, it was decided that the moisture content would be determined for each combination of slag, lime, and flyash by running a standard compaction test to obtain the optimum moisture content for each mixture.

Mixing

The mixer used throughout this study was a Lancaster Mixer equipped with a muller. It was found that an increase in mixing time resulted in a higher compressive strength for a given mixture. This relationship is shown in Fig. 3. Since most present-day standards simply state that the material should be thoroughly mixed, it was necessary to analyse this data and determine an appropriate mixing time which would be comparable with average field conditions. Two minutes was determined to be a reasonable time, based on observations of field mixes and of laboratory-prepared mixtures, and the results of the laboratory mixing time study. These two minutes were divided into one minute of dry mixing, followed by the addition of the water, and one minute of wet mixing.

At one time it was thought that the additional mixing was actually changing the particle size and thus producing a higher compressive strength. A sieve analysis, before and after four minutes of mixing, revealed that the particle size was not changed. It was thus assumed that the increase in compressive strength was due to a better distribution of all ingredients.

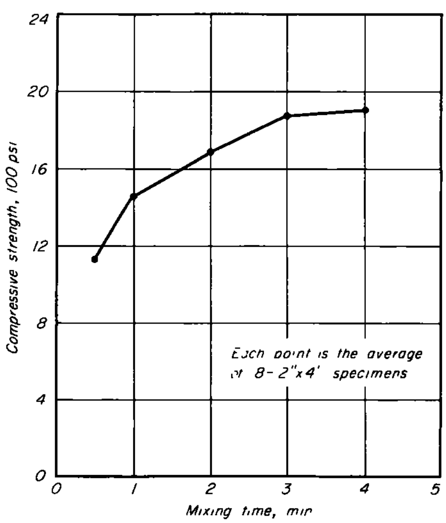

Figure 3. Relationship of compressive strength to mixing time for 72.1 percent slag, 24.0 percent flyash, and 3.9 percent lime mixture.

Molding

The test specimens were molded immediately after each batch of material was mixed. The mixture was compacted in the molds in four equal layers with each layer receiving ten blows of a 4-lb hammer dropping 12 in. As shown in Fig. 2 this is equivalent to standard AASHO maximum density.

After each layer was compacted its surface was scarified in order to develop adequate bond between layers. In order that the final layer would receive the same amount of compaction as the preceding layers, a collar was made to fit over the mold so that it could be filled with loose material to a height greater than the top of the mold. To assure the same amount of material in each layer, a scoop was constructed with a volume which approximated the volume of loose material required for a 1-in. compacted layer.

Capping and Removal

Various removal times and capping materials were investigated. In order to determine the amount of variation in compressive strength of the test specimens which may be attributed to the capping and the removal procedures, the coefficient of variation for compressive strength was determined for various methods. The coefficent of variation is a convenient way to express the degree of variation as a percentage of the mean value. A certain amount of variation may be attributed to the testing technique but the increment between various coefficients of variation may be attributed to the capping and removal techniques. This test data is summarized as follows:

1. Extruded immediately and capped with Hydrocal, coefficient of variation, 10.5 percent.
2. Cured in the molds and capped with Hydrocal, coefficient of variation, 7.9 percent.
3. Cured in the molds overnight and capped with neat cement, coefficient of variation, 7.1 percent.

On the basis of the foregoing data, method 3 was used as the standard. Neat cement released the molds for continuous use following the initial curing period and the specimens were prepared for testing in one continuous operation.

The molded specimens were weighed to determine the wet unit weight, after which a thin layer of material approximately ⅛ of an in. thick was removed from each end of the specimen. A mixture of cement and water was placed on each end of the specimen and metal plates were used to form a smooth surface. The specimens were then left undisturbed for 20 ± 4 hr in the laboratory air. Following this initial curing period each specimen was removed from the mold by the use of a pedestal. Care was taken

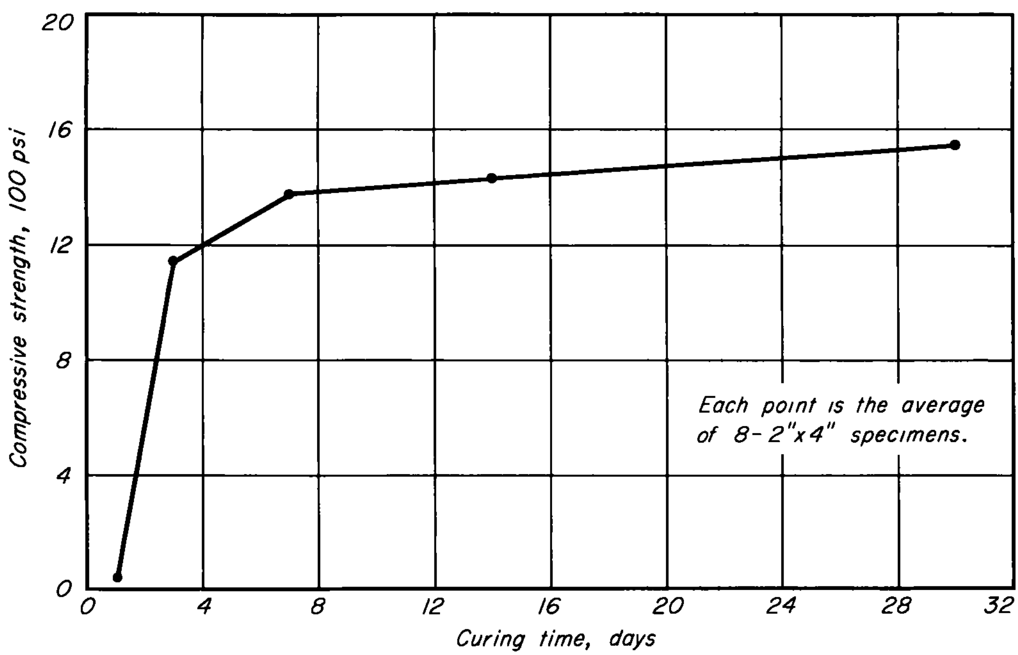

Figure 4. Relationship of compressive strength to curing time for 72.1 percent slag, 24.0 percent flyash, and 3.9 percent lime mixture.

to keep the specimens in compression since they are very weak in tension at this early age. The specimens were then given an appropriate indentification number. Each series of specimens was gently placed and sealed in a 1 gal paint can.

Curing

The sealed paint cans containing compressive strength specimens were placed in a constant temperature oven at 140 F and cured for an additional 13 days, thus giving a total curing period of 14 days. The relationship between the curing time in the oven and compressive strength is shown in Figure 4.

To compare field curing with oven curing, some specimens were placed in the field and companion specimens were placed in the oven. The field-cured specimens were buried in the soil in such a manner that the tops of the specimens were covered with 2 in. of soil. These specimens were cured and tested after various periods of time. The temperatures and relative humidities prevailing during the field curing are shown in Fig. 5. The compressive strength test results are shown in Fig. 6. It may be observed that for this mixture the 7-day oven-cured strength approximates the 1-yr field strength.

Preparation for Testing

The average diameter of each cylinder was determined to the nearest 0.02 of an in. in order to accurately determine the area. After the specimens were measured they were allowed to reach room temperature, and then weighed to determine their net gain or loss in weight during the curing.

An investigation of the effect of saturation on the compressive strength of test specimens showed that saturated specimens had an average compressive strength of 100 psi less than comparable specimens tested in the dry state. This difference was not considered great enough to warrant the additional time and extra handling required for saturation, and all subsequent specimens were tested in the dry state.

Testing

The rate of deformation used in all tests was 0.05 in. per min. All compressive

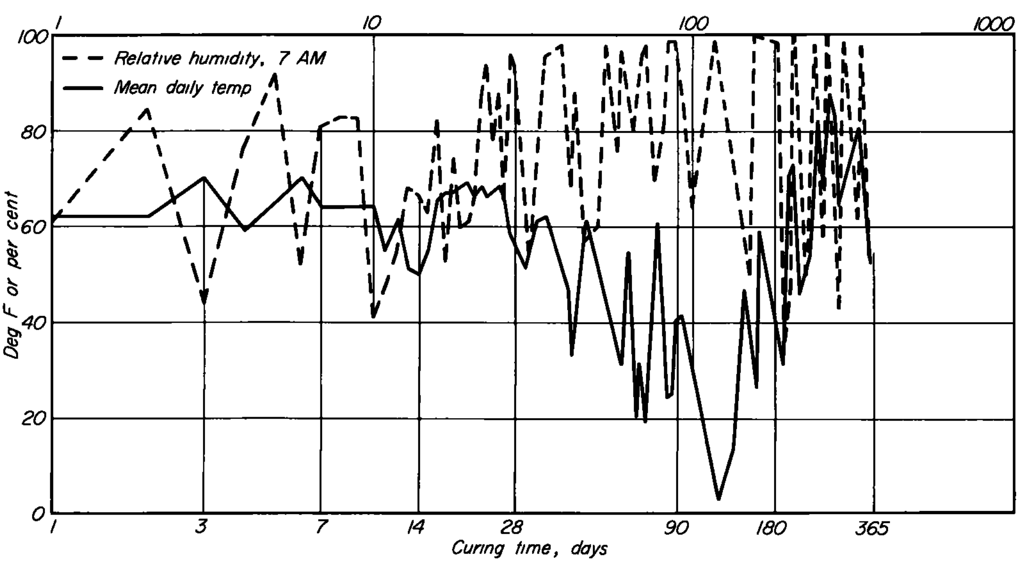

Figure 5. Temperature and relative humidity prevailing during field curing.

strength values were determined to the nearest 10 psi. There was no L/D correction applied to the standard 2- by 4- in. cylinder.

DETERMINATION OF THE NUMBER OF TEST SPECIMENS IN RELATION TO THE TECHNIQUES USED AND THE RELIABILITY OF TEST RESULTS DESIRED

Pilot Study to Determine Variations

After the laboratory techniques and known variables were standardized, a series of test specimens was made by the standardized techniques and tested to determine the effect of the remaining unknown variables on test results. The pilot study consisted of making 24 2- by 4-in. specimens with all of the known variables fixed. The compressive strength of the 24 specimens was analyzed statistically. The relationship between the variation in compressive strength and the number of test specimens used in the pilot study is shown by the dashed lines for two different probabilities in Fig. 7.

In order to understand the curves shown in Fig. 7, the reader is referred to ASTM Designation E122-56, which presents the recommended practice for "Choice of Sample Size to Estimate the Average Quality of a Lot or Process."

A definition of standard error and limit of accuracy as shown in Fig. 7 follows: The standard error implies that for a given number of specimens, n, the mean value of n specimens will be within plus or minus the value shown on the ordinate (Fig. 7) of the population mean 68.3 times out of 100. The population mean is obtained by testing all the samples in the universe. Similarly, limit of accuracy means that for a given number of specimens, n, the mean values of n specimens will be within plus or minus the value shown on the ordinate of the population mean 95.5 times out of 100.

It was believed that for the data to be meaningful the maximum allowable difference between the estimate of the true mean and the true mean, which would be the result of testing (by the same methods) all of the units in the universe, should be set at plus or minus 100 psi for a probability of 95.5 percent. This limiting condition was based solely on judgment.

From the pilot study curve it is seen that six 2- by 4-in. specimens fulfill the above requirement. The mixer used had a capacity such that eight specimens could be made in one batch. Based on a consideration of the above factors it was decided that eight

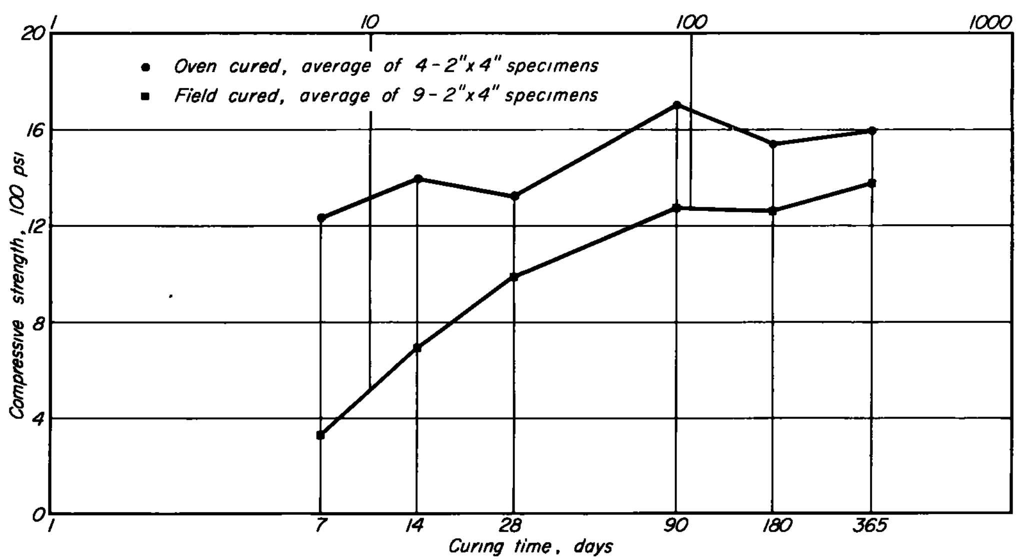

Figure 6. Relationship of compressive strength to curing time for 72.1 percent slag, 24.0 percent flyash, and 3.9 percent lime mixture.

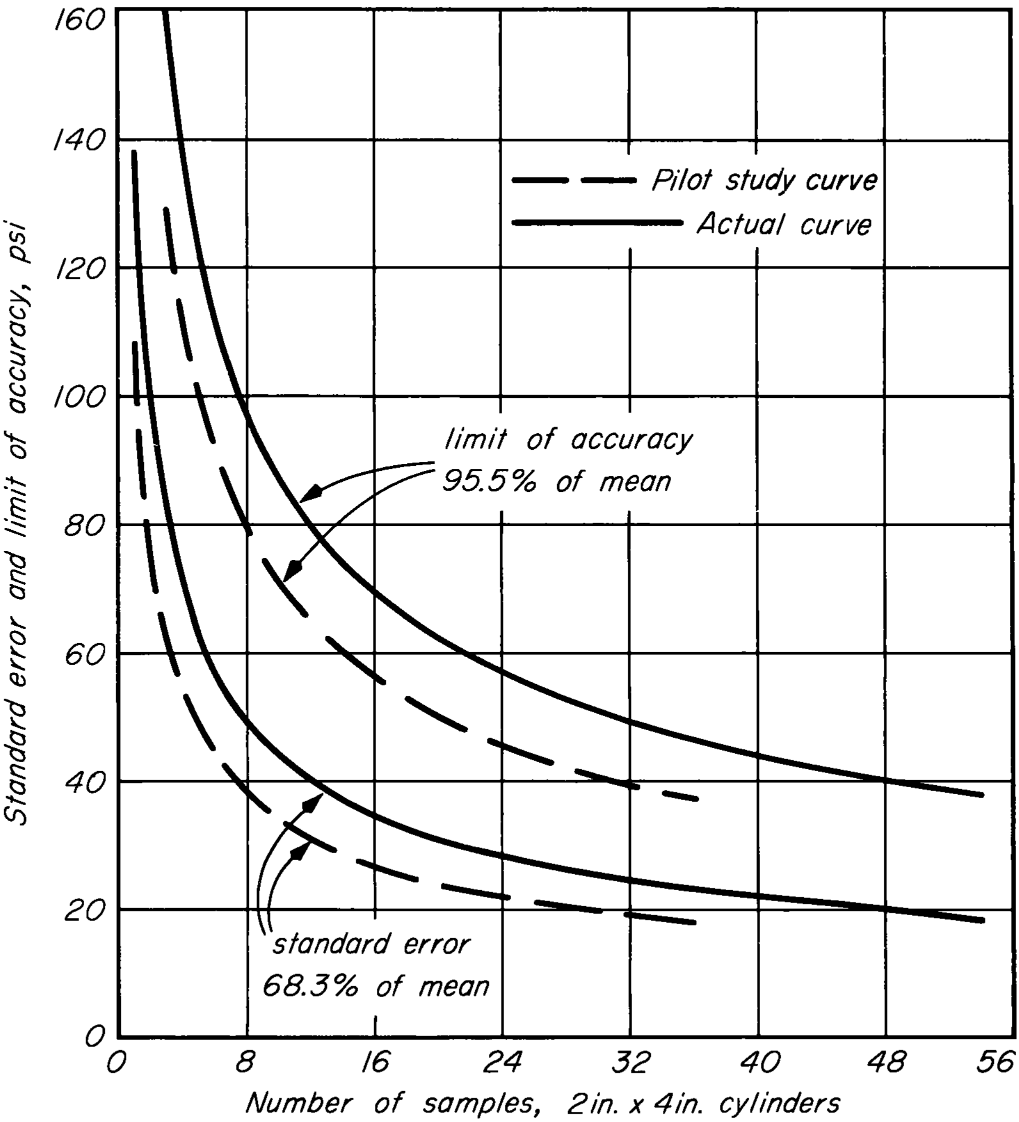

Figure 7. Standard error and limit of accuracy curves for compressive strength of 2- by 4-in. specimens.

2- by 4- in. specimens would be made to evaluate each condition.

Correlation of the Reliability, Probability, and Number of Test Specimens for the Actual Test Program

After completing the test program, the reliability of the pilot study was checked against the reliability of a curve which was developed from an analysis of all the routine test specimens. Figure 7 shows both the original pilot study curve and the actual curve. It is interesting to note that the actual curve based on an analysis of 54 specimens has a slightly larger variation than the computed curve. For eight specimens the pilot study curve indicated that the mean compressive strength would be within plus or minus 80 psi of the population mean, whereas, the actual curve indicates that the average of eight specimens would be within plus or minus 100 psi of the population mean. Somewhat less care used in the routine testing may be responsible for this variation. It was fortunate that the average of eight 2- by 4-in. specimens satisfied the initial criteria.

SIGNIFICANCE OF TEST RESULTS

A statistical analysis of the actual data based on ASTM Designation E122-56T showed that the average compressive strength of eight 2- by 4-in. specimens was with ± 100 psi of the population mean 95.5 times out of 100. It follows that a significant difference for a probability of 95.5 percent of the time requires that the average compressive strength of eight specimens differ by more than 200 psi from the average of another set of eight 2- by 4-in. specimens. The reason for the significant difference being 200 psi is that it is possible for the average of eight 2- by 4-in. specimens to be 100 psi above the population mean and the average of another set of eight 2- by 4-in. specimens to be 100 psi below the population mean. This information makes the task of analyzing the test data a relatively simple matter.

SUMMARY

This paper sets forth an approach which may be used in determining whether new materials possess qualities that will enable them to be used satisfactorily in the ever expanding highway field. First, it is necessary to evaluate all of the known variables that may affect the specific material under consideration. After this has been done and all of the variables have been standardized there will still be some variation in the test results due to uncontrollable variations or unknown variables. These variations can be reduced to a known value by means of a pilot study and a statistical analysis. The exact amount of the variation can be either raised or lowered by decreasing or increasing the number of test specimens used in the evaluation of each item. Thus, to obtain significant data within a given range for a given probability it will be necessary to test a given number of specimens.

By use of this technique researchers may avoid to a large extent the pitfall of drawing conclusions from data based on too few test results. This will also help eliminate the embarassing situation of subsequently finding that the conclusions are erroneous because all of the variables were not considered.

Discussion

B. B. CHAMBLIN, JR., Highway Research Engineer, Virginia Council of Highway Investigation and Research, Charlottesville, Virginia—This application of statistical techniques to a practical investigation into the compressive strength of slag, lime, flyash mixtures illustrates the advantages gained in the interpretation of test data when these techniques are applied to experimental design.

The same principles are useful in many other fields of testing, some of which are: concrete beam and cylinder work; California Bearing Ratio soil comparisons; field density measurements; and routine control of research test specimens. Most laboratories are faced with a variety of possible applications of "statistics."

When a number of supposedly identical specimens are tested for some characteristic, the test results are seldom identical; they are distributed about a mean value, some values larger, some smaller. If this distribution can be assumed "normal" (bell shaped), and it usually can, two statistics can be computed. They are: (a) Mean of test values and (b) standard deviation of values. These two statistics, and the number of samples tested, completely determine the characteristics of the distribution, and should be reported with the data. From these, other useful measures can be computed. Some are: (a) Coefficient of variation and (b) confidence limits for true mean. The "standard error" and "limit of accuracy" used by Hollon and Danner are forms of the confidence limit.

The writer's use of confidence limits is best illustrated by an example. Assume this hypothetical data, normally distributed, of strength values from cylindrical specimens.

1. Number of samples tested N = 24
2. Mean of results $\bar{X}$ = 1,600 psi
3. Standard deviation s = 50 psi

It can be seen that this sample data can be utilized in making assumptions about the population (or true parameters). The relationship between sample mean, $\bar{X}$, and true mean, μ, is given by this equation.

$$\mu = \bar{X} \pm \frac{ts}{\sqrt{N}} \qquad (1)$$

where "t" is a standard score computed for various degrees of certainty. It is found in tabular form in most statistical references. At the 95 percent confidence level, for 24 samples equals 2.07. By Equation 1,

$$\mu = 1600 \pm \frac{(2.07)\,(50)}{4.90} = 1600 \pm 21.1$$

This suggests that the true mean value of cylinder strength lies between 1,580 and 1,620 psi. If this statement is made, the probability of its being correct is 95 in 100. This probability was controlled when t was selected from the tabled values.

In this same manner, by working backwards, the number of samples necessary to insure a certain width of confidence limits may be calculated. Solving Equation 1 for sample number,

$$N = \left(\frac{ts}{\mu \pm \bar{X}}\right)^2$$

($\mu \pm \bar{X}$) is selected, t is determined from tables, and s is found by testing a number of nominally identical samples.

It is hoped that this discussion will amplify the presentation by Hollon and Danner, and that papers such as this will lead to the wider use of statistical techniques by testing engineers throughout the country. The assignment of definite numerical levels of confidence to conclusions drawn from testing program results is certainly a positive step in the advancement of experimental projects.

CLOSURE, George W. Hollon and Ellis Danner—We wish to thank Mr. Chamblin for his discussion of our paper and we appreciate his support of the concepts which we presented.