# Required Number of Specimens for Moisture Susceptibility Testing

KWANG WOO KIM, JAMES L. BURATI, JR., AND SERJI N. AMIRKHANIAN

**The moisture susceptibility of asphaltic concrete mixtures can be evaluated using a one-tail lower *t*-test for the mean difference between dry and wet tensile strengths. The practical number of specimens to be used for moisture susceptibility testing in the laboratory should also be valid statistically. On the basis of the data from 960 laboratory specimens, the reliabilities for using various numbers of samples for the moisture susceptibility test were evaluated by several statistical analyses. The power of the *t*-tests, a measure of the reliability of the moisture susceptibility tests, was also evaluated on the basis of sample size, difference between dry and wet means, and variances among samples. The analyses showed that the decision derived from the *t*-test based on small sample sizes had low reliability. Factors that could possibly influence the power of the test were examined. It was found that the most likely method for improving reliability was to increase the number of samples used. Therefore, it is recommended that the minimum number of samples be determined based on sample variance, difference between dry and wet means, and the expected importance of the result.**

The number of samples to be tested for evaluating moisture susceptibility must be valid statistically. The design of any laboratory experiment must therefore be planned carefully to determine the appropriate number of samples to obtain an acceptable reliability from the experiment. As a general rule, increasing the number of samples results in increased accuracy, increased testing costs, and possibly increased time. However, personnel in most central laboratories or field laboratories wish to prepare and test only the minimum number of samples. This paper presents a procedure to examine and evaluate the required number of samples for moisture susceptibility testing of asphaltic concrete mixtures in the laboratory. The methods described here are statistical approaches based on analysis of data produced from previous moisture susceptibility testing (*1*).

The data used are from moisture susceptibility tests conducted on 3 aggregate sources (designated as A, B and C) from South Carolina, 4 AC-20 asphalt cements (designated as I, II, III, and IV), and 5 treatments (one control designated as 0 and 4 antistripping additives designated as 1, 2, 3, and 4), resulting in a total of 60 mixture combinations. Eight replicates were prepared for dry testing and 8 for wet testing for each combination (16 specimens for each combination), resulting in a total of 960 specimens (specimens were 4 in. in diameter and 2.5 in. high).

The indirect tensile strength (ITS) test was used to measure the tensile strength of wet- or dry-conditioned specimens for each mixture combination (*2, 3*). Then, *t*-tests for mean strength difference of the dry- and wet-conditioned mixtures were conducted to evaluate the moisture susceptibility of the mixtures. Mixture combinations and tensile strength values for each combination for each moisture condition are shown in Table 1.

Department of Civil Engineering, Clemson University, Clemson, S.C. 29634-0911.

## APPLICATION OF *t*-TEST FOR MOISTURE SUSCEPTIBILITY TESTING

The average tensile strengths of wet-conditioned specimens were compared with those of dry-conditioned specimens by use of the *t*-test, a statistical technique suitable for use with small numbers of data. A 5 percent significance level ($\alpha = 0.05$) was used in the *t*-test on the difference of two means. A one-tail lower *t*-test was used to test the null hypothesis that tensile strength of wet-conditioned mixtures is equal to the strength of the dry-conditioned mixtures ($H_o$: $\mu_{wet} = \mu_{dry}$) against the alternate hypothesis that tensile strength of the wet-conditioned mixtures is lower than that of the dry-conditioned mixtures ($H_a$: $\mu_{wet} < \mu_{dry}$).

Significant difference in the *t*-test at the $\alpha = 0.05$ level meant that the average tensile strength of wet specimens was significantly lower, with a 95 percent likelihood, than that of dry specimens. This interpretation can be applied differently for the control mixture and the additive-treated mixtures. For the control mixture, a significant difference meant that a reduction in the strength of wet-conditioned samples compared with that of dry-conditioned samples had occurred, and that an antistrip additive might be needed to increase the strength of the wet-conditioned mixture. For mixtures treated with an antistrip additive, a significant difference meant that the wet strength of the mixture is still lower than the dry strength after an additive treatment (i.e., that mixture may still be moisture susceptible with the additive) (*4*).

If $N1$ and $N2$ are the numbers of samples for the dry and wet conditions, respectively, in the *t*-distribution, the degrees of freedom (*df*) for the two-mean comparison are ($N1 + N2 - 2$). The degrees of freedom are a minimum value of 2 when $N1$ and $N2$ are both 2 (*5*).

## COMPARISON OF PERCENTAGE OF REJECTING THE NULL HYPOTHESIS

The *t*-test on two means (wet and dry tensile strength means) was conducted using all eight samples for each moisture condition. If the *t*-test based on eight samples showed a significant difference between the average dry and wet tensile strengths,

TABLE 1 DATA USED FOR STATISTICAL ANALYSIS

| Mixture Combination | | | No. of samples | Dry ITS Mean (psi) | Dry ITS St Dev (psi) | Wet ITS Mean (psi) | Wet ITS St Dev (psi) |
|---|---|---|---|---|---|---|---|
| A | I | 0 | 8 | 67.7 | 8.37 | 44.1 | 4.76 |
| A | I | 1 | 8 | 65.2 | 7.51 | 57.3 | 10.34 |
| A | I | 2 | 8 | 62.3 | 8.65 | 55.8 | 11.03 |
| A | I | 3 | 8 | 65.0 | 11.01 | 58.3 | 14.80 |
| A | I | 4 | 8 | 65.7 | 9.29 | 52.6 | 11.44 |
| A | II | 0 | 8 | 69.3 | 15.79 | 46.2 | 9.77 |
| A | II | 1 | 8 | 73.3 | 9.36 | 60.8 | 10.19 |
| A | II | 2 | 8 | 71.4 | 12.71 | 61.4 | 9.13 |
| A | II | 3 | 8 | 71.6 | 12.66 | 64.8 | 9.22 |
| A | II | 4 | 8 | 63.1 | 8.39 | 59.3 | 10.27 |
| A | III | 0 | 8 | 81.9 | 5.30 | 61.2 | 7.75 |
| A | III | 1 | 8 | 86.3 | 12.85 | 75.2 | 9.89 |
| A | III | 2 | 8 | 82.0 | 14.62 | 74.3 | 9.38 |
| A | III | 3 | 8 | 85.6 | 11.63 | 72.0 | 10.72 |
| A | III | 4 | 8 | 73.7 | 12.43 | 68.0 | 9.31 |
| A | IV | 0 | 8 | 76.6 | 14.37 | 61.6 | 12.79 |
| A | IV | 1 | 8 | 78.0 | 11.61 | 71.2 | 6.83 |
| A | IV | 2 | 8 | 78.9 | 15.19 | 73.4 | 8.10 |
| A | IV | 3 | 8 | 76.2 | 14.65 | 78.0 | 7.50 |
| A | IV | 4 | 8 | 79.8 | 16.98 | 66.4 | 12.90 |
| B | I | 0 | 8 | 50.6 | 6.77 | 45.1 | 4.19 |
| B | I | 1 | 8 | 51.7 | 9.68 | 46.3 | 4.72 |
| B | I | 2 | 8 | 48.8 | 9.58 | 46.2 | 8.16 |
| B | I | 3 | 8 | 52.4 | 6.94 | 47.6 | 8.01 |
| B | I | 4 | 8 | 45.0 | 7.11 | 44.8 | 5.84 |
| B | II | 0 | 8 | 54.7 | 8.72 | 47.1 | 5.63 |
| B | II | 1 | 8 | 56.0 | 8.09 | 53.7 | 7.79 |
| B | II | 2 | 8 | 55.3 | 6.30 | 46.9 | 9.85 |
| B | II | 3 | 8 | 54.9 | 8.16 | 51.3 | 5.50 |
| B | II | 4 | 8 | 52.2 | 7.14 | 45.8 | 9.26 |
| B | III | 0 | 8 | 58.4 | 5.94 | 57.8 | 7.05 |
| B | III | 1 | 8 | 63.6 | 10.23 | 66.3 | 8.53 |
| B | III | 2 | 8 | 64.9 | 9.96 | 63.3 | 5.22 |
| B | III | 3 | 8 | 64.6 | 9.84 | 58.2 | 11.27 |
| B | III | 4 | 8 | 60.3 | 8.51 | 56.6 | 5.92 |
| B | IV | 0 | 8 | 61.2 | 12.56 | 58.0 | 9.98 |
| B | IV | 1 | 8 | 61.2 | 12.75 | 63.9 | 7.81 |
| B | IV | 2 | 8 | 66.5 | 14.92 | 55.4 | 6.34 |
| B | IV | 3 | 8 | 56.0 | 10.94 | 58.1 | 11.93 |
| B | IV | 4 | 8 | 57.8 | 11.89 | 57.0 | 8.15 |
| C | I | 0 | 8 | 58.9 | 10.23 | 29.4 | 8.17 |
| C | I | 1 | 8 | 58.0 | 11.37 | 51.0 | 8.70 |
| C | I | 2 | 8 | 62.6 | 6.80 | 65.4 | 5.66 |
| C | I | 3 | 8 | 57.6 | 6.79 | 48.5 | 8.13 |
| C | I | 4 | 8 | 57.4 | 9.52 | 42.1 | 8.40 |
| C | II | 0 | 8 | 66.3 | 8.20 | 26.2 | 3.67 |
| C | II | 1 | 8 | 67.8 | 10.32 | 56.8 | 8.11 |
| C | II | 2 | 8 | 64.3 | 11.92 | 61.9 | 9.81 |
| C | II | 3 | 8 | 65.3 | 8.67 | 58.6 | 12.64 |
| C | II | 4 | 8 | 58.6 | 12.64 | 53.5 | 7.00 |

TABLE 1 *(continued on next page)*

TABLE 1 (*continued*)

| Mixture Combination | No. of samples | Dry ITS Mean (psi) | Dry ITS St Dev (psi) | Wet ITS Mean (psi) | Wet ITS St Dev (psi) |
|---|---|---|---|---|---|
| C III 0 | 8 | 72.2 | 13.99 | 45.9 | 5.21 |
| C III 1 | 8 | 71.9 | 10.37 | 63.9 | 8.94 |
| C III 2 | 8 | 72.6 | 12.91 | 61.6 | 9.01 |
| C III 3 | 8 | 71.5 | 9.94 | 64.5 | 6.04 |
| C III 4 | 8 | 68.0 | 11.55 | 62.0 | 6.86 |
| C IV 0 | 8 | 73.5 | 17.24 | 35.7 | 3.80 |
| C IV 1 | 8 | 74.6 | 17.79 | 63.8 | 8.63 |
| C IV 2 | 8 | 77.7 | 13.35 | 73.1 | 9.86 |
| C IV 3 | 8 | 71.9 | 17.29 | 68.9 | 10.07 |
| C IV 4 | 8 | 68.5 | 11.76 | 59.7 | 6.30 |

Note: A, B, C = aggregate
I, II, III, IV = asphalt cement
0, 1, 2, 3, 4 = anti-stripping treatment

TABLE 2 *t*-TEST RESULTS FOR SAMPLE SIZE OF 8

| t-Test Result | Aggregate A | | Aggregate B | | Aggregate C | | Total | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | C* | T* | C | T | C | T | C | T | |
| Significant | 4 | 6 | 2 | 2 | 4 | 5 | 10 | 13 | 23 |
| Not Significant | 0 | 10 | 2 | 14 | 0 | 11 | 2 | 35 | 37 |
| Total | 4 | 16 | 4 | 16 | 4 | 16 | 12 | 48 | 60 |

* C = Control mixtures, T = additive treated mixtures.

then the two means of the populations were considered significantly different. When there was no significant difference in the *t*-test of eight samples, the means of the two populations were considered not significantly different.

The *t*-test results (significant or not significant) for all combinations were recorded and are presented, for comparison, in Table 2. Out of 60 mixture combinations, 23 wet versus dry comparisons were found to be significantly different by the *t*-test based on eight samples. Population means of dry and wet strengths for those 23 mixture combinations were therefore considered significantly different. In this case, the $H_o$ that "the mean tensile strength of the dry-conditioned mixture is the same as the wet-conditioned" is actually false. Therefore, for any number of samples, rejecting $H_o$ is a correct decision in the *t*-tests for these 23 mixture combinations.

Using those 23 mixtures, a number of *t*-tests were conducted for sample sizes (number of samples) of 2 to 7 for each mixture combination. The number of *t*-tests for each sample size was determined by the mathematical combination for selecting the number of samples out of 8, as specified in Table 3 (*5*). For example, the numbers of combinations (the number of *t*-tests) for sample sizes 2 and 3 are $8!/(2! \times 6!) = 28$ and $8!/(3! \times 5!) = 56$, respectively. For each sample size, a different combination of samples selected from the 8 was used for each *t*-test. This procedure was performed the appropriate number of times for each sample size for each mixture combination.

TABLE 3 PERCENTAGE OF $t$-TESTS REJECTING $H_o$ AT THE $\alpha = 0.05$ LEVEL

| Mixture | Percentage of Significant Differences | | | | | | |
|---|---|---|---|---|---|---|---|
| | Sample Size | | | | | | |
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | Number of t-tests | | | | | | |
| | 28 | 56 | 70 | 56 | 28 | 8 | 1 |
| A I 0 | 21 | 98 | 100 | 100 | 100 | 100 | 100 |
| A I 5 | 14 | 20 | 30 | 38 | 79 | 100 | 100 |
| A II 0 | 18 | 50 | 81 | 100 | 100 | 100 | 100 |
| A II 1 | 11 | 27 | 39 | 63 | 86 | 100 | 100 |
| A II 2 | 4 | 18 | 16 | 20 | 22 | 38 | 100 |
| A III 0 | 71 | 98 | 100 | 100 | 100 | 100 | 100 |
| A III 1 | 18 | 18 | 14 | 29 | 39 | 63 | 100 |
| A III 3 | 25 | 30 | 41 | 64 | 75 | 100 | 100 |
| A IV 0 | 11 | 9 | 21 | 34 | 50 | 75 | 100 |
| A IV 4 | 11 | 21 | 16 | 23 | 32 | 25 | 100 |
| Average | 25.9 | 38.9 | 45.8 | 57.1 | 68.3 | 80.1 | 100 |
| B I 0 | 11 | 14 | 26 | 30 | 50 | 50 | 100 |
| B II 0 | 11 | 11 | 11 | 18 | 32 | 63 | 100 |
| B II 2 | 7 | 11 | 16 | 30 | 36 | 50 | 100 |
| B IV 2 | 29 | 39 | 51 | 46 | 32 | 38 | 100 |
| Average | 14.5 | 19.3 | 26.0 | 31.0 | 37.5 | 50.3 | 100 |
| C I 0 | 68 | 93 | 100 | 100 | 100 | 100 | 100 |
| C I 3 | 11 | 29 | 29 | 45 | 93 | 100 | 100 |
| C I 4 | 32 | 61 | 76 | 89 | 100 | 100 | 100 |
| C II 0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| C II 1 | 11 | 27 | 39 | 61 | 71 | 100 | 100 |
| C II 3 | 18 | 34 | 66 | 98 | 100 | 100 | 100 |
| C III 0 | 57 | 82 | 100 | 100 | 100 | 100 | 100 |
| C IV 0 | 75 | 98 | 100 | 100 | 100 | 100 | 100 |
| C IV 4 | 7 | 20 | 23 | 29 | 29 | 25 | 100 |
| Average | 42.1 | 60.4 | 70.3 | 80.2 | 88.1 | 91.7 | 100 |
| Total Avg. | 30.3 | 43.9 | 51.9 | 57.6 | 70.7 | 79.5 | 100 |

It was assumed that the samples were homogeneous and therefore that any number of samples selected from the 8 still represented the population statistically.

The number of significant differences in $t$-tests was converted to a percentage of the total number of $t$-tests conducted for each sample size for each mixture combination. The percentages obtained for each sample size for all mixture combinations are presented in Table 3. The percentage of significant difference was low for small sample sizes. For example, using the mixture combination of Aggregate A, Asphalt I, and Antistripping Additive 5, or AI5 in Table 3, the percentage of significant difference was 14 percent when the sample size was 2. On the average for Aggregate A, 25.9 percent of the $t$-tests rejected the $H_o$ when the sample size was 2, and 38.9 percent rejected the $H_o$ when the sample size was 3. For sample sizes of 2 to 4 in Table 3, the chance of detecting a significant difference was less than approximately half that for a sample size of 8. This meant that the chance of obtaining a correct decision would be 50 percent, or less, when 4 or fewer samples were used.

## EVALUATION OF SAMPLE SIZE BASED ON SIMULATED DATA

Because the analysis in the previous section was based on a limited number of samples and $t$-tests, it is not easy to see the consistent change of reliability of the $t$-test with changes in sample size. Monte Carlo simulation was therefore used to evaluate the reliability based on a larger number of simulated sample values and $t$-tests. The combinations of A-control and C-control were used for simulation because those are the groups that showed all significant differences in the $t$-tests for eight samples (Table 2). The probability distributions for ten-

TABLE 4 PROBABILITY DISTRIBUTIONS FOR TENSILE STRENGTH

| Agg. | Moisture Condition | Treatment | No. | Probability Distribution | Mean (psi) | St Dev (psi) |
|---|---|---|---|---|---|---|
| A | Dry | Control | 32 | Normal | 73.9 | 12.6 |
| | Wet | Control | 32 | Normal | 53.3 | 12.1 |
| C | Dry | Control | 32 | Normal | 67.6 | 13.6 |
| | Wet | Control | 32 | Normal | 34.3 | 9.3 |

sile strength values were determined separately for each aggregate and moisture condition using goodness-of-fit tests. The results are shown in Table 4.

All distributions for tensile strength were found to follow normal distributions at the $\alpha = 0.05$ level by both chi-square and Kolmogorov-Smirnov (K-S) goodness-of-fit tests. Using these probability distributions, a given number of sample values for wet and dry conditions was simulated and used for one-tail lower *t*-tests at the $\alpha = 0.05$ level. The simulated *t*-tests were conducted 5,000 times for each sample size for each combination. This process was conducted for sample sizes of 2 to 8. The probabilities of rejecting a false $H_o$ (reliability) were recorded for the different sample sizes, and the results are illustrated in Figure 1.

Both A-control and C-control show steep rates of change in the probability of rejecting $H_o$ as the sample size increases from 2 to 5. The reliability of using 2 samples for A-control is approximately one-third that of using 8 samples, whereas using 3 samples has approximately half the reliability of using 8 samples. The reliability of using small sample sizes for C-control is relatively high because of the large difference between dry and wet strengths (Table 4). The rate of change was almost zero after a sample size of 6 for C-control because reliabilities were close to 1.0. The reliability for A-control increased consistently as the sample size increased beyond 5.

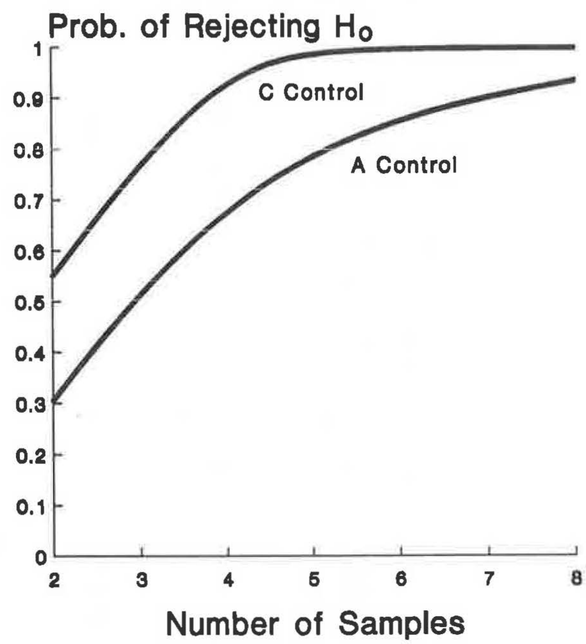


**FIGURE 1 Probability of rejecting $H_o$ versus sample size.**

## EVALUATION OF SAMPLE SIZE USING POWER OF TEST

The reliability of the *t*-test can be evaluated by the power of the test. The power of test is the probability of rejecting $H_o$ when it is not true. Therefore, the larger the value for power of the test, the higher the reliability of the test result (*5*). The basic concept of power of test when testing the difference between two means is shown in Figure 2. The power of test, $1 - \beta$ (shaded area in the figure), can be explained with the following:

$$\beta = Pr\,[\,X_w > X_\alpha |\, X_d > X_w\,] \quad (1)$$

$$\text{Power} = 1 - \beta = 1 - Pr\,[\,X_w > X_\alpha |\, X_d > X_w\,] \quad (2)$$

where

$X_d$ = mean strength of dry-conditioned samples,
$X_w$ = mean strength of wet-conditioned samples,
$X_\alpha$ = critical strength at given level of $\alpha$, and
$\beta$ = probability of type II error.

If the *t*-distribution is used for the test of the difference between two means, the mathematical expression is as follows:

$$\beta = Pr\,[\,t_o > t_\alpha |\, t_o \neq 0\,] \quad (3)$$

$$\text{Power} = 1 - \beta = 1 - Pr\,[\,t_o > t_\alpha |\, t_o \neq 0] = Pr\,[\,t_o < t_\alpha |\, t_o \neq 0\,] = Pr\,[t_o < t_\alpha\,] \quad (4)$$

where

$t_o$ = calculated $t$ and
$t_\alpha$ = critical $t$ at given level of $\alpha$.

The value for the power can be obtained using the SAS (Statistical Analysis System) function PROBT(*X,df*), where PROBT denotes the probability for the *t*-distribution function. The PROBT function computes the probability that a random variable with a *t*-distribution with *df* degrees of freedom falls below the *X* value given (*6*). The *X* value is given by:

$$X = |\text{Calculated } t| - \text{Critical } t \quad (5)$$

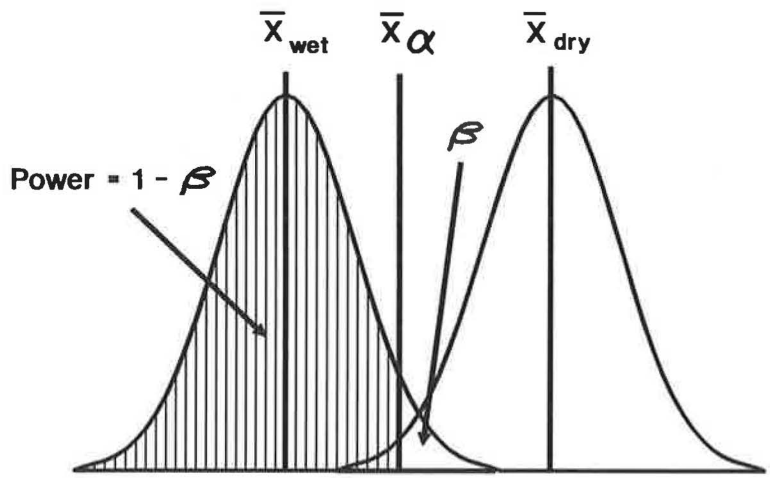


**FIGURE 2 Power of test for lower one-tail *t*-test on difference between dry and wet means.**

The data were combined for 6 groups (2 groups per aggregate—one for all control mixtures and one for all additive-treated mixtures, resulting in 6 groups for 3 aggregates as shown in Table 2) for power of test determination. The powers of test for lower one-tail *t*-tests were calculated at the $\alpha$ = .05 level. The power of test is plotted in Figure 3 against the number of samples for each combination. The power in Figure 3 represents the probability of detecting a difference when the wet and dry means are actually different.

In the figure, the power of test increases as the sample size increases for all the groups. The rate of increase is high for A-control and C-control. This result is identical to that in the previous section. However, the rates for the other 4, 3 treated groups and B-control, were very low. The reason for this is that those samples have small differences in their dry and wet means. For those mixture groups, almost no difference was detected between wet and dry tensile strength means by the *t*-test for 8 samples. Therefore, $H_o$ is considered correct in this case. When there is a small difference in two means, the calculated *t* value will be affected mainly by variations among samples. For those types of mixtures, therefore, it is beneficial to make the sample size as large as possible so that variances of the sample means will be reduced as low as possible.

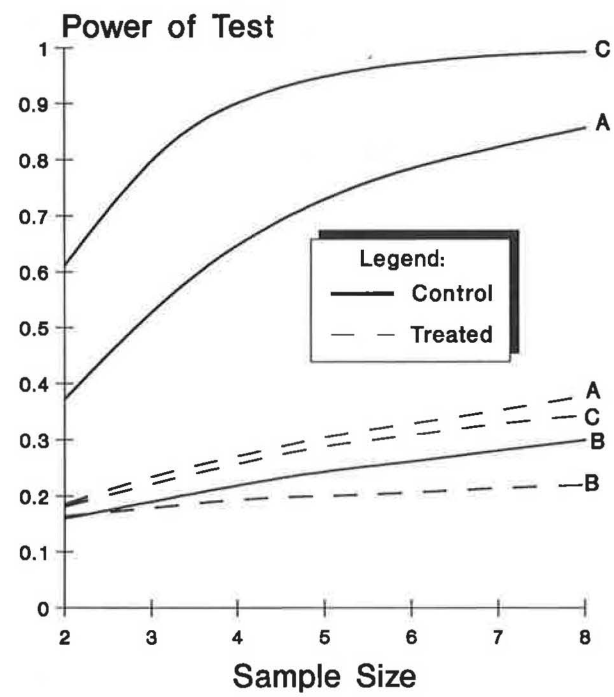


**FIGURE 3 Power of test at $\alpha$ = 0.05.**

## CONTROL OF THE POWER OF TEST

There are several factors affecting the power of tests, including sample size, sample variance, and the difference between the two means. As shown in Figure 4, the power is low for small sample sizes, small differences between two means, or large sample variances. If the variance can be reduced in the laboratory, the power can be improved. If sample size increases, the power will also be improved. However, differences between the two means, which are based on the intrinsic strengths of a specific mixture, are not a controllable factor and are, in fact, what the experimenter intends to find. Therefore, variance and sample size are the factors that can be controlled for improving the power of tests.

The variance among samples consists of uncertainties associated with many factors, such as material variations, experimental errors, and construction or sampling errors. Because construction or sampling errors are related to field samples, they are not considered in this study.

Variations of materials may include variabilities in quality, quantity, and source of materials used for the specimens. The materials include coarse aggregates, fine aggregates, asphalt cements, filler materials, and, possibly, additives. Variability of size distribution (gradation), quality and quantity of aggregates, and variability of asphalt quality and asphalt content will cause variation of mixture strength. Variability of air voids of specimens resulting from variability of aggregate gradation, asphalt content, compaction, and the like is also related to the variation of mixture strength.

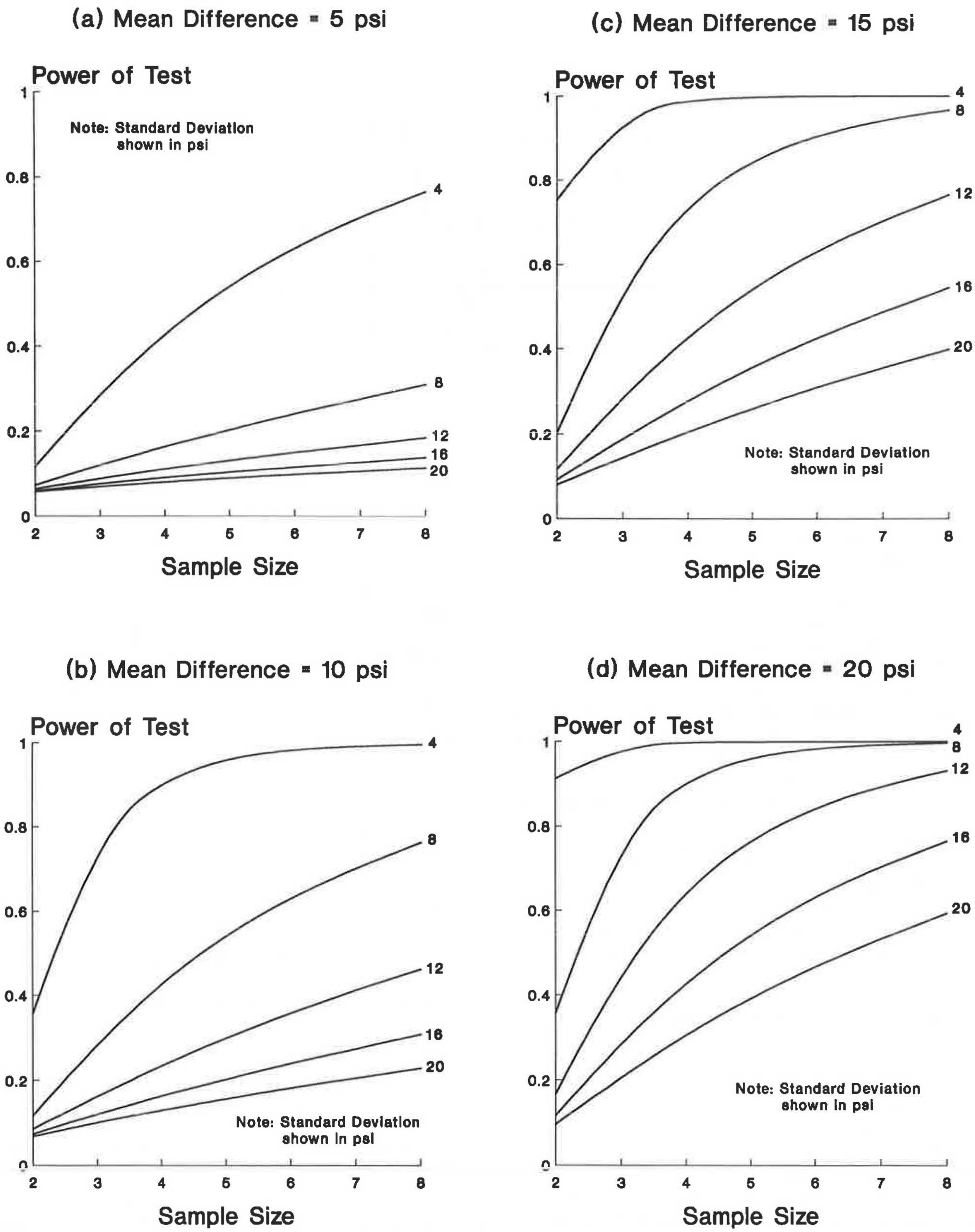


**FIGURE 4** Comparison of power of test for various mean differences and standard deviations.

Experimental error may be induced from mixing and compaction procedures, incorrect calibration of machines, testing and reading errors, and variation of test conditions, such as temperatures for mixing, compacting, cooling, and breaking of the specimens. Those experimental variations result in variabilities in specimen size, percent compaction, air voids, percent saturation of wet-conditioned specimens, and strength of mixtures. If the experimenter is changed from time to time, personal differences reflected in the results cannot be ignored. Testing errors may constitute up to one-half of total variability (*3*).

Estimating every variation or error separately is not easy; however, these variations or errors will accumulate and will contribute to increased variance of sample strengths. Assuming that there are three major independent variations to be considered, and that each can be evaluated as a coefficient of variation, total variation can be estimated by Equation 6 (*7, 8*):

$$V_t = (V_1^2 + V_2^2 + V_3^2 + V_1^2V_2^2 + V_1^2V_3^2 + V_2^2V_3^2 + V_1^2V_2^2V_3^2)^{1/2} \qquad (6)$$

where

$V_t$ = coefficient of variation of total value,
$V_1$ = coefficient of variation of material quality,
$V_2$ = coefficient of variation of experimental error, and
$V_3$ = coefficient of variation of construction error.

If these variations are assumed to be independent of each other, then Equation 6 can be rewritten as

$$V_t = (V_1^2 + V_2^2 + V_3^2)^{1/2} \qquad (7)$$

Assume that there are two major uncertainties, material variations and experimental errors, in asphaltic concrete mixture strengths measured in the laboratory. If, for example, the native variation of asphaltic concrete materials is assumed at approximately 12 percent or 0.12, and if minimum estimated experimental error is assumed at 7 percent or 0.07, by Equation 7, the total variation ($V_t$) involved with the strength of the sample for laboratory specimens will be approximately 14 percent or 0.14. Applying this value to the average tensile strength for A-control (74 psi) used in this study, the standard deviation will be $74 \times 0.14 = 10.4$ psi. This standard deviation, 10.4 psi in this case, should be considered an originally existing variation that is almost uncontrollable by laboratory efforts. In other words, in this case, on the average the standard deviation cannot be reduced to less than 10.4 psi by laboratory endeavor. Therefore, there is a limit to the control of variation as long as there are inevitable variations in materials and tests.

Increasing the sample size is therefore the most likely method for improving the power of test and reliability of experimental results. Increasing the sample size to an unlimited number, however, is unrealistic because of the time and cost involved with the testing. Therefore, the sample size should be selected on the basis of the importance of the expected experimental result. In Figure 4(d), for example, if the experimenter requires a 0.9 power of test with a pooled standard deviation of 12 psi and a mean difference of 20 psi, an acceptable sample size will be between 7 and 8. If the required power is 0.8 with the same conditions, the required sample size will be between 5 and 6. The data for variation (standard deviation or coefficient of variation), difference of two means, and required power of test can be obtained from previous data if available, from the literature, or from the engineer's judgment.

## SUMMARY AND CONCLUSIONS

On the basis of the analysis of data from 960 laboratory specimens, the number of samples that should be used in acceptance testing for moisture susceptibility was examined. The *t*-test was used for testing, at the $\alpha = 0.05$ level, the null hypothesis that the dry and wet tensile strengths were equal against the alternate hypothesis that the wet strength was less than the dry strength. Using several statistical analyses, the reliabilities of decisions derived from the *t*-tests for evaluating moisture susceptibility were evaluated and compared for different sample sizes.

Twenty-three mixture combinations out of a total of 60 were found to be significantly different in *t*-tests based on all 8 samples for each mixture combination. For each of the 23 combinations, percentages for rejecting $H_o$ were calculated by repeating *t*-tests for various numbers of samples. Comparing those percentages revealed that the chances of obtaining a correct decision (rejecting $H_o$ in this case) in the *t*-test were low for small numbers of samples when compared with larger numbers of samples. The *t*-tests based on simulated data showed a consistent increase of reliability by increasing sample size from 2 to 8. Powers of test were also evaluated for different numbers of samples. The power of test increased as the sample size increased for all mixtures. From those analyses, it was shown that the reliability for testing small sample sizes (4 or less) was, in general, half or less than half the reliability for testing 8 samples.

The reliability of test results can be improved in the laboratory by reducing material variation and experimental error or by increasing sample size. Because there are inevitable uncertainties in the materials and testing procedures used, however, there is a limit to the reduction that can be obtained in the variations or experimental errors. Therefore, increasing the number of samples is the most likely method for improving the power of test and reliability of the test result.

However, increasing the number of samples more than the minimum number required is unrealistic because of limited time and budget. The minimum number of samples can be determined on the basis of sample variance values, difference of dry and wet means, and the expected importance of the result. The variance, difference between two means, and expected level of the power of test can be obtained from previous data, from the literature, or through the experimenter's judgment. Once these factors are determined for a certain mixture, Figure 4 can be used for determining the approximate number of samples to be tested.

## ACKNOWLEDGMENTS

Partial funding for the work presented in this paper was provided by the National Science Foundation, Presidential Young Investigator Award. Matching support was provided by the South Carolina Asphalt Pavement Association.

## REFERENCES


1. H. W. Busching, J. L. Burati, and S. N. Amirkhanian. *An Investigation of Stripping in Asphalt Concrete in South Carolina.* Final Report FHWA-SC-86-02. FHWA, U.S. Department of Transportation, July 1986.
2. T. W. Kennedy. *Tensile Characterization of Highway Pavement Materials.* Research Report 183-15F. Center for Transportation Research, University of Texas at Austin, July 1983.
3. E. J. Yoder and M. W. Witczak. *Principles of Pavement Design*, 2nd ed. John Wiley and Sons, Inc., New York, 1975.
4. D. G. Tunnicliff and R. E. Root. *NCHRP Report 274: Use of Antistripping Additives in Asphaltic Concrete Mixtures—Laboratory Phase.* TRB, National Research Council, Washington, D.C., Dec. 1984.
5. R. G. D. Steel and J. H. Torrie. *Principles and Procedures of Statistics: A Biometrical Approach*, 2nd ed. McGraw-Hill Book Company, New York, 1980.
6. *SAS User's Guide: Basics, 1984 ed.* SAS Institute Inc., Cary, N.C.
7. A. H. S. Ang and C. A. Cornell. Reliability Bases of Structural Safety and Design. *Journal of Structural Division ASCE*, Vol. 100, No. St9, Sept. 1974, pp. 1755–1769.
8. C. A. Cornell. A Probability Based Structural Code. *ACI Journal*, Dec. 1969, pp. 974–985.


*Publication of this paper sponsored by Committee on Quality Assurance and Acceptance Procedures.*