This paper is published as:

McCabe, BY, AbouRizk, SM, Gavin, J, 1999, "Sample Size Analysis for Asphalt Pavement Quality Control", ASCE Journal of Infrastructure Systems, 5(4): 118-123

SAMPLE SIZE ANALYSIS FOR ASPHALT PAVEMENT QUALITY CONTROL

Brenda McCabe¹, Member, ASCE, Simaan AbouRizk², Member, ASCE, Jim Gavin³

ABSTRACT

A review of the quality assurance practices of Alberta Transportation and Utilities was undertaken for each of three quality test measures: asphalt content; percent compaction; and, aggregate gradation. The primary objective of the study was to determine how reducing the number of tests per lot affect the lot mean values and the degree of confidence in the quality results. The secondary objective was to determine the effect a change in the sample size would have on the quality-related price adjustments to the contractor.

Two methods were used to compare the number of samples and their resulting mean values. First, the minimum number of samples required to stay within the variance allowed in the specifications, given the variance of the samples, was determined. Second, the mean values derived from sample sizes of five, four and three were compared using one-way ANOVA and independent T-test. The study found that the minimum number of tests for asphalt content, percent compaction and aggregate gradation could be reduced. The effect on the price adjustments to the contractor given the reduced number of test samples was negligible.

Keywords: quality assurance, asphalt pavement, sample size

BACKGROUND

Due to imposed budgetary constraints, many government agencies are working to reduce their operating costs by reevaluating their procedures and efficiency. Alberta Transportation & Utilities (AT&U), in the Province of Alberta, Canada, worked with University of Alberta to review some of their policies related to quality assurance testing of newly constructed asphalt pavements. Quality Assurance (QA) measures are based upon test results of samples collected on a lot by lot basis. In AT&U specifications, a lot is generally defined as representing a full day of hot mix production and placement.

The objective of reviewing the sampling procedure was twofold. First, AT&U wanted to determine how changes in the lot sample sizes would affect confidence in the test results. QA tests were undertaken to estimate asphalt content, percent compaction and aggregate gradation. The second objective was to determine the effect any change

¹ Assistant Professor, Department of Civil Engineering, University of Toronto, Toronto, ON, M5S 1A4

² Professor, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 2G7

³ Roadway Construction Standards Specialist, Alberta Transportation & Utilities, 4999 - 98 Ave., Edmonton, AB, T6B 2X3

to the sampling procedure might have on the contractors.

The objectives have both a theoretical and a practical side. Statistically, the analysis can determine whether there is a difference in the means of different sample sizes. The increase in risk of determining the incorrect mean lot value, measured by confidence in the standard deviation of the mean, can also be calculated. The practical issues relate to the effect on the payment adjustments in the field of changing the number of samples. The monetary effect can be calculated, however, the effect on other aspects of the project, such as contractor and owner confidence, cannot be calculated.

NUMBER OF SAMPLES PER LOT

The current quality assurance (QA) program involves sampling and testing five cores per lot for each of the previously mentioned QA measures. AT&U has used five samples as a standard number since the inception of their End Product Specifications (EPS) in the mid 1980s [Palsat and McMillan 1994]. In EPS, penalty/bonus payment adjustments are provided to the contractor based upon the quality of the product received as determined using specified sampling and testing protocols. To illustrate, if the average test value of the five samples (e.g. asphalt content) is outside the specified tolerance limits, the contractor is penalized. Conversely if the average test value indicates a high quality product is produced, (e.g. high pavement compaction), the contractor receives a bonus.

The standard number “five” was originally chosen because five was generally felt to provide accurate mean estimates at a reasonable cost to the highway agency [Rilett 1998]. In more recent years AT&U has been examining ways to reduce their operating costs, including means of reducing QA sampling and testing costs, without compromising the quality of the final product. One proposed alternative was to reduce the standard number of QA core samples from five to four. However, concerns were raised on part of the contracting industry that such a change would increase their risk in receiving greater penalties based upon greater variance in the mean test values.

METHOD OF ANALYSIS

The initial step in the analysis was to determine the minimum number of samples required per lot to meet the variance in the mean allowed in the specifications before the contractor is penalized. Because the calculation is directly proportional to the standard deviation of the sample, confidence in the standard deviation was evaluated. Therefore, for various sample sizes, the confidence interval in the standard deviation was determined. The minimum number of samples required, then, is dependent on the amount of variance in the sample that is acceptable to the owner.

The second part of the evaluation required statistical comparisons of the lot means of three, four, and five samples per lot using analysis of variance (ANOVA) and independent T-tests [Green and Salkind 1997]. ANOVA is a method of testing the null hypothesis that several group means are equal in the population, by comparing the sample variance estimated from the group means to that estimated within the groups. The T-test procedure compares the means of two variables for a single group by

computing the differences between values of the two variables for each case and tests whether the average differs from zero. In order for the lot means to be considered equal in general, ninety percent of the individual lot comparisons made must be found to be equal. Ninety- percent compliance was justified through analysis of the theoretical probability of achieving that compliance.

Hypotheses

For each of the quality assurance tests under investigation, the null, or accepted hypothesis, H_0 , states that the mean of five samples is equal to the mean of four, or three samples. The alternative hypothesis, H_1 , states that the means are not equal.

$$H_0: \mu_5 = \mu_4$$

$$H_1: \mu_5 \neq \mu_4$$

and

$$H_0: \mu_5 = \mu_3$$

$$H_1: \mu_5 \neq \mu_3$$

where μ_5 , μ_4 and μ_3 are the means of five, four, and three sample tests per lot, respectively.

Analysis Approach

Actual QA test data were collected from the 1994 construction season. From these data, groups of three numerically *adjacent lots* within a project were selected for comparison of the means. Adjacent lots refer to lots that are consecutively numbered within a highway project e.g. Lots 4, 5, and 6. The assumption is that numerically adjacent lots represent physically adjacent lots or road sections constructed on consecutive days. The numerically lowest lot was evaluated using the mean of 5 samples, the second lot was evaluated using the mean of 4 samples, and the numerically highest lot was evaluated using the mean of 3 samples. Because the analysis compared the means of different lots with different sample sizes, special consideration was taken to ensure the lots were comparable. For example, suppose lots 4, 7, 15 were adjacent data in the historical records for one project i.e. data for lots 5, 6 and 8 through 14 were missing. These records were not used as a group of three lots in the analysis because the data records are missing for reasons unknown to the investigator. The jump between non-consecutive lots may represent a break in the work, and therefore, could indicate a non-steady state. A non-steady state of construction could affect the comparison, and was therefore avoided. To further ensure the project is in steady state, means from lots within three lots from the start or the finish of the project were not included in the analysis.

Level of Confidence, α , was 0.05 for this analysis. Therefore, the results may be viewed with a $1-\alpha$, or 95% confidence. In all cases, the data has been confirmed as approximately normal through standard tests from normality. Further, where comparisons were made between lots using ANOVA analysis, it was confirmed that the variances were within an acceptable range of each other.

Minimum Number of Tests Required

Two calculations to determine the minimum number of tests required to meet the statistical variance allowed by the specification were performed. Equation 1 is commonly

used to evaluate the number of tests required to ensure, with a certain confidence or probability, that the variance of the mean is equal to or less than the variance allowed in the specifications before adjustments to the payment schedule are made.

$$w = 2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \text{ or } n = \left(\frac{2z_{\alpha/2}\sigma}{w} \right)^2 \quad [1]$$

where:

z is the standardized normal distribution value for the required confidence

σ is the standard deviation of the population

w is the width of the confidence interval covering both positive and negative deviation from the mean allowed in the specifications

n is the minimum number of samples required given the above constraints

To address concerns about confidence in the means from smaller sample sizes, a more conservative approach is used. Shown in Equation 2, $\sigma = \sigma'$, where σ' is a confidence interval of the standard deviation that accounts for variance between lots being evaluated. This is attractive because the population standard deviation has been estimated from sample sizes of 5 for each lot, and by using the mean standard deviation only, the probability that the standard deviation of the population is less than or equal to the standard deviation of the sample will be $P(S \leq \sigma) = 50\%$. A higher level of confidence may be desired. Therefore, σ' is:

$$\sigma' = \bar{\sigma} \pm z_{\alpha/2} \sigma$$

$$\text{where } \sigma = \frac{\bar{\sigma}}{2} c \quad [2]$$

$$c = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\frac{n}{2}}$$

The value of c [Cowden, 1957] may be determined using the Gamma function, with $\Gamma(j)$ as $\Gamma(j) = \pi^{1/2}$ for $j=1/2$ and for $\Gamma(j) = (j-1)!$ when j is an integer. Because n is required to evaluate c , which is then used to evaluate n , an initial value of n is required. For the calculations in this section, the value of $n=3$, and $c=1.382$ were used. The value of $z_{\alpha/2}$ is the standardized normal distribution value for the level of confidence α desired and will be varied to show the effect of different confidence levels in the value of the standard deviation on the number of samples required. The mean standard deviation, $\bar{\sigma}$, and the resulting number of required tests, n , will be used as input.

Analysis of the Means

Analysis of the means provided indication of whether the three-, four-, and five-sample means are equal. For practical reasons, the effect on the payment adjustments to the contractor was evaluated through the difference between the means of the sample sizes. This may be looked upon as the error. If the mean error is approximately zero, then

there is no tendency to consistently over or under estimate the sample mean compared to the current practice of five samples per lot. Where the error is zero, the total effect on the payment adjustments is also zero.

One-way ANOVA tests and independent T-tests were performed to compare the means and variances of the three sample sizes. The tests were performed on groups of three numerically adjacent lots as outlined in the previous discussion. Within each group, there were three lots, say A, B, and C, with a total of twelve data records, five from lot A, four from lot B and three from lot C, where lots A, B, and C are numerically adjacent lots within one project.

Analysis of variance, or ANOVA, is useful in all fields of research when data are measured quantitatively. It can be used to test the null hypothesis that several group means are equal by comparing the sample variance estimated from the group means to that estimated within the groups. It is assumed that the data are a random sample from a normal population. The ANOVA analysis was based on simultaneous comparisons of the three sample size means. If the null hypothesis was not rejected, (there was insufficient evidence to conclude the three means are not equivalent), then no further analysis was required. However, if the analysis showed that the means were not equal, two-sample student's t-test was used to compare the means on a pair-wise fashion to determine which means were not equal.

After the analysis, the number of successful comparisons was determined. Successful comparisons are those that indicate that there is insufficient evidence to reject the null hypothesis, and therefore hypothesis that $\mu_5 = \mu_4$, or $\mu_5 = \mu_3$ may be accepted. A target of 90% was set for the number of successful comparisons relative to the total number of comparisons made in order for the means of the different sample sizes to be considered equal. The value of 90% was evaluated as follows [Cowden 1957]:

$$p = P \left[\frac{|\bar{X}_5 - \bar{X}_4|}{\sigma \sqrt{\frac{2}{5}}} \leq z_{\alpha/2} \right] = 1 - \alpha \quad [3]$$

where $\bar{X}_5$ is the mean of five samples, $\bar{X}_4$ is the mean of four samples, σ is the standard deviation of the samples, z is the normalized statistic for normal distributions, α is the level of confidence, and $P[\]$ is the probability that the difference of the means would be equal to or less than z . Therefore, with a probability of 95% ($\alpha=0.05$), and consequently $z=1.96$, the difference between the means should be less than or equal to: 1.240σ as demonstrated in equation 4.

$$z_{\alpha/2} * \sigma \sqrt{\frac{2}{5}} = 1.240\sigma \quad [4]$$

If the number of groups tested is n , and 90% of the tests must be successful in order for the means to be considered equal, then the number of successful ANOVA analysis required, k , is $0.9 n$. The likelihood of m or more successful comparisons occurring is:

$$P(\text{Success}) = \sum_{k=m}^n \binom{n}{k} * 0.95^k * 0.05^{n-k} \quad [5]$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ evaluates number of combinations of n items, taken k at a time.

For example, if 30 lots are tested ($n=30$), then the number of successful analysis is $k=0.9*30=27$. Using the average standard deviation over all of the aggregate gradation lots and sieve sizes of 0.994, the theoretical difference between the means is 1.233 using Equation 4. The probability of observing 27 of 30 successful tests, then, is 93.9% as evaluated using Equation 5. The resulting parameters for each of the quality assurance tests are shown in Table 1. Column 5 shows the allowable difference of the means. The allowable difference is $w/2$, where w is the width of the allowable range of the mean test results above and below the value in the specifications before a penalty is applied to the contractor. One may expect the theoretical and the allowable differences to be reasonably similar, and in all cases, the allowable is within 20% of the theoretical. Using a minimum limit of 90%, then appears to fit well with the theoretical probability.

Table 1: Evaluation of Minimum Level of Successful Tests

QA Measure	σ	n	Theoretical $ \bar{X}_5 - \bar{X}_4 $	Allowable $ \bar{X}_5 - \bar{X}_4 $	Theoretical P(Success)
[1]	[2]	[3]	[4]	[5]	[6]
Asphalt Content	0.225	83	0.279	0.3	97.7%
Percent Compaction	1.207	42	1.500	1.75	94.2%
Aggregate Gradation	0.994	30	1.233	1.0	93.9%

Analysis Results for Asphalt Content

Minimum Number of Tests Required

The following values were used as input to the equation $n = \left(\frac{2z_{\alpha/2}\sigma}{w} \right)^2$.

$z=1.96$ for 95% confidence

$\sigma=0.225$ is the average standard deviation of the samples included in the analysis
 $w=0.6$ was determined using the allowable variance of 0.3% from the specified asphalt content in each direction before penalties are applied to the payment schedule to the contractor

The number of samples required to meet the specified variance evaluated using the empirical formula is 2.2, or if rounded up to the nearest integer, 3 samples. Equation 2 is used below to determine the number of samples required using a more conservative value of the standard deviation.

The confidence levels of the standard deviation are shown in column 1 of Table 2,

along with the resulting minimum number of samples required per lot based on the standard deviation. The value of n in Equation 1 is highly dependent on S , the estimate of the population standard deviation, σ . Note that Equation 1 was based on a 50% probability that the population standard deviation exceeded the value of the sample set standard deviation, and resulted in a minimum number of samples per lot of just over 2. As the number of samples per lot is reduced, then, the sample standard deviation becomes less certain or more variable. If a greater certainty is required in the value of the sample standard deviation or variance, then more samples are required. Note that as the number of samples per lot increases, the corresponding increase in the confidence in the value in the standard deviation decreases. For example, an 11% increase in the confidence in the standard deviation is gained by increasing the number of samples from 3 to 4. However, the increase in confidence is only 8% when the sample size is further increased to five. Conversely, decreasing the sample size has a greater effect on the confidence with each reduction.

Table 2: Minimum Number of Samples for Asphalt Content

P($S \leq \sigma$) [1]	$Z_{\alpha/2}$ [2]	σ' [3]	n [4]
50%	0.000	0.225	2.2
61%	0.279	0.264	3.0
72%	0.583	0.307	4.0
80%	0.842	0.344	5.0
86%	1.080	0.377	6.0
90%	1.282	0.406	7.0

The current practice of using five samples implies an 80% confidence in that value. If AT&U and the contractors are comfortable using a lower confidence, then reducing the number of samples may be warranted.

Analysis of the Means

Of the eighty-three comparisons made between the five-, four-, and three-sample groups, seventy-six, or 92% of the cases indicated that the means are statistically equal. This result exceeds the target of 90% set at the start of the study.

Although the means are statistically equal, they are not identical. Therefore, as shown in Table 3, the means were compared to determine the extent of any differences that existed. The mean of each five-sample test was subtracted from the corresponding four- and three-sample means. The mean and standard deviation of the differences, plus the maximum absolute difference are provided.

Table 3: Analysis of the Means for Asphalt Content

Comparison [1]	$\bar{X}_4 - \bar{X}_5$ [2]	$\bar{X}_3 - \bar{X}_5$ [3]
Mean Difference	0.00861	0.00223
Std. Deviation	0.0403	0.0965
Maximum Difference	0.124	0.235
Number of Comparisons = 83		

The mean difference between the sample sets is approximately zero, indicating that the mean of four or three samples was neither consistently higher nor lower than the standard five samples. For the contractor, this result implies that, on average, they will be neither penalized nor profited by a reduction in the number of samples used. The maximum difference is also small relative to the values of asphalt content, which ranged from 4.5% to 6%. The maximum values were also smaller than $w/2$, the variance from the specified asphalt content allowed before penalties were assigned to the contractor.

Summary for Analysis of Asphalt Content

Reduction of the number of samples will decrease confidence in the mean value, increasing risk to the contractor and to the owner. The result of this analysis provides support that the number of samples can be reduced to four or three per lot without significantly affecting price adjustments to the contractor. The mean difference between the sample size means is approximately equal to zero, indicating the average price adjustment to the contractor will not be significantly affected. The effect of reducing the number of samples will, however, reduce the cost of quality assurance testing by AT&U.

Analysis Results for Percent Compaction

Minimum Number of Tests Required

The following values were used as input to the equation $n = \left(\frac{2z\sigma}{w} \right)^2$

$z=1.96$ for 95% confidence in the evaluation

$\sigma=1.207$ is the average standard deviation of the samples

$w=3.5$ was determined from Specification 3.50, where the lift is greater than 35mm.

At 96.5% compaction, the penalty to the contractor is -\$0.50 per tonne, the balance to the maximum bonus of +\$0.50 for compaction greater than 98.0%. The difference between 100% and 96.5% was then used as the allowable width.

Table 4: Minimum Number of Samples for Percent Compaction

P($S \leq \sigma$) [1]	$Z_{\alpha/2}$ [2]	σ' [3]	n [4]
0.50	0.000	1.207	1.8
0.53	0.075	1.264	2.0
0.67	0.440	1.540	3.0
0.78	0.772	1.791	4.0
0.85	1.036	1.991	5.0
0.90	1.282	2.176	6.0

The number of samples, n in Column 4, required to meet the specified variance evaluated using the empirical formula is 1.8, or if rounded up to the nearest integer, 2 samples. This is shown in Column 1 of Table 4 as $P(S \leq \sigma)=50\%$, along with the more conservative calculations utilizing Equation 2. Note that the current practice of using five samples relates to an 85% confidence that the population standard deviation is equal to or less than the one used in the calculation, $\sigma'=1.991$. Similar to the results for asphalt content, as the number of samples per lot increases, the rate of increase in confidence in the value of the standard deviation decreases. For example, the increase in confidence

is 11% when the sample size is increased from 3 to 4. However, the increase in confidence is only 7% when the sample size is further increased from 4 to 5 samples.

Analysis of the Means

Eighty-seven percent, or 34, of the thirty-eight comparisons made between the five- and four-sample test indicated that the means were equal. Ninety-seven percent of the thirty-eight comparisons between five- and three-sample tests indicated the means were equal. With the targeted minimum of 90%, the evaluation has shown that four sample tests do not have the same mean as the five sample tests. However, the three sample tests *do* have the same mean as five sample tests. This obviously does not make sense.

The number of successful tests required to meet the target minimum is $90\% \times 38$, or 34.2 tests; however, only 34 tests were successful. Considering the failure of the test is balanced on a fraction of a test and the fact that the three-sample comparisons overwhelmingly passed, the failure may be a result of random error. As there was no other apparent explanation for this anomaly in the data nor in the method, the test was determined inconclusive.

The mean and standard deviation of the difference of the means of the four and three sample tests for percent compaction, compared to the standard five sample tests, are presented in Table 5. The maximum absolute difference is the largest value that was observed above or below the mean. The mean error in both cases is approximately zero, indicating that the difference was neither consistently high nor low.

Table 5: Analysis of the Means for Percent Compaction

Comparison [1]	$\bar{X}_4 - \bar{X}_5$ [2]	$\bar{X}_3 - \bar{X}_5$ [3]
Mean Difference	0.134	-0.049
Std. Deviation	1.269	1.186
Maximum Absolute Difference	4.615	3.473
Number of Comparisons = 38		

Summary for Analysis of Percent Compaction

The analysis of the four-sample mean was inconclusive. Reluctance to accept the failure is based upon very strong results (97% successes) from the three-sample analysis. A reduction of samples to four per lot will decrease confidence in the standard deviation by 7%, with a further loss of confidence of 11% to 67% confidence if the number of samples per lot is decreased to three. However, the mean difference between the lot means of five samples and three- and four-samples was approximately zero, indicating that the effect of reducing the number of samples to the contractor price adjustment will be negligible.

Analysis Results for Aggregate Gradation

Minimum Number of Tests Required

The following values were used as input to the above equation $n = \left(\frac{2z\sigma}{w} \right)^2$

$z=1.96$ for 95% confidence

$\sigma=0.994$ is the average standard deviation of the samples over all sieve sizes

$w=2.0$ was determined from Table 3.50F of AT&U Specification 3.50 Asphalt Concrete Pavement - EPS, which states that penalties begin with each 1% deviation from the specified gradation

The number of samples required to meet the specified variance is 3.8, or if rounded up to the nearest integer, 4 samples. The results of Equation 2 including the result of Equation 1 at $P(S \leq \sigma)=50\%$ are shown in Column 1 of Table 6. The current practice of using five samples relates to a probability of 59% that the population standard deviation is equal to or less than the standard deviation used in the calculation. This analysis differs from the previous two in that the standard deviation is rather large, likely caused by the percent passing values documented as the nearest whole percent. However, as seen in the next section, this does not appear to have any effect on the analysis of the means.

Table 6: Minimum Number of Samples for Aggregate Gradation

P(S≤σ) [1]	z [2]	σ' [3]	n [4]
50%	0.000	0.994	3.8
52%	0.050	1.025	4.0
59%	0.228	1.136	5.0
66%	0.412	1.251	6.0
72%	0.553	1.339	7.0
77%	0.739	1.454	8.0
81%	0.842	1.518	9.0
84%	0.994	1.613	10.0

Analysis of the Means

Gradation results were divided into two designations of aggregate top size: 10 mm and 16 mm. For each, thirty lots were randomly selected from the data provided. Seven sieve sizes were analyzed for the 10mm designation, nine for the 16mm, totaling over two hundred tests for each comparison of sample size. At least ninety-nine percent of the tests for 16mm and 10mm designation for both four and three-sample size analysis, over all of the sieve sizes, indicated that the means of the three- and four-sample tests, compared to the five-sample tests, are equal. As this is significantly higher than the minimum percent of successes (90%) previously set, the means of each of the sample sizes may be accepted as equal.

The mean and standard deviation of the difference of the means of the four and three sample tests, compared to the standard five sample tests, over all of the sieve sizes,

are provided in Table 7. The maximum absolute difference is the largest difference that was observed above or below the mean. The mean differences are approximately zero, indicating that the difference was neither consistently high nor low. Further, the maximum difference is reasonably low.

Table 7: Analysis of the Means for Aggregate Designations

Comparison	10mm Topsize		16mm Topsize	
	$\bar{X}_4 - \bar{X}_5$	$\bar{X}_3 - \bar{X}_5$	$\bar{X}_4 - \bar{X}_5$	$\bar{X}_3 - \bar{X}_5$
[1]	[2]	[3]	[4]	[5]
Mean Difference	0.04	-0.03	-0.05	0.03
Std. Deviation	0.31	0.41	0.36	0.79
Maximum Absolute Difference	1.55	2.87	1.25	1.8
Number of Comparisons = 30 for each sieve size				

Summary for Analysis of Aggregate Gradation

The analysis of aggregate gradation is unique in that the gradation is documented to the nearest 1%, and a deviation of the lot mean by 1% from the specification warrants a penalty to the contractor. This affected the calculation of the minimum number of samples, which is strongly dependent upon the width of the specified sample mean, w , and the standard deviation of the samples. Results showed that to achieve 85% confidence that the population standard deviation is equal to or lower than that of the sample, 10 samples per lot would have to be used. However, the analysis of the various sample means indicated overwhelmingly that the means of three-, four-, and five-sample lots are equal. Confidence in the standard deviation does not appear to affect the price adjustment to the contractor.

Conclusion

There were two objectives in this analysis. The first was to determine the effect on the confidence in the test results of changing the number of samples tested per lot. This was measured by comparing the lot means using different sample sizes.

Table 8 contains a summary of the comparative test results. The analysis has shown that three or four samples may be used per lot to determine the mean of either asphalt content, compaction or aggregate gradation. It was found that the means are statistically the same as the mean derived from the current practice of five samples per lot. A high level of confidence was used to evaluate the standard deviation of the means to address concerns related to contractor confidence in the variability of the means.

The second objective was to determine the effect a change to the sampling procedure might have on the payment adjustments to the contractors. The mean difference between the means of four- or three-sample tests and the traditional five-samples was approximately zero in all cases, indicating there was no tendency to over- or underestimate the mean with varying sample size. This further implies the overall effect of changing sample sizes on the risk to the contractor of increased penalties and the risk to AT&U of increased bonuses on the contractor through price adjustments is negligible.

These two conclusions have met the technical aspects of the objectives, however,

practical challenges to changing the sample size should be recognized. These challenges differ from the statistical confidence in that statistical confidence is a numeric value that does not take into consideration the human concept of risk. Although there is proof that the change to smaller sample sizes would not affect the lot mean nor the payments to the contractor, confidence of both AT&U and the contractors in the new system must be gained. Therefore, as an interim means to implement changes in the standard testing protocols, it is recommended that five samples continue to be taken, but testing is conducted on only four of them to determine the lot mean values. The fifth sample could be used where disputes arise between the contractor and the owner. After time, if confidence is gained by both parties in using a reduced number of samples, the standard number of samples taken could be reduced to four.

Table 8: Summary of Sample Size Analysis

Comparison [1]	Asphalt Content [2]	Percent Compaction [3]	Aggregate Gradation [4]
Minimum Number of Samples Required			
$P(S \leq \sigma) = 50\%$	3	2	4
$P(S \leq \sigma) = 85\%$	6	5	10
Percent of Successful Tests (Minimum 90% Required)			
5 vs. 4 samples	92%	87%	99%
5 vs. 3 samples	92%	97%	99%
Mean Difference			10mm, 16mm
5 vs. 4 samples	0.00861	0.134	0.04, -0.05
5 vs. 3 samples	0.00223	-0.049	-0.03, 0.03
Standard Deviation of Difference			
5 vs. 4 samples	0.0403	1.269	0.31, 0.36
5 vs. 3 samples	0.0965	1.186	0.41, 0.79
Maximum Absolute Difference			
5 vs. 4 samples	0.124	4.615	1.55, 2.87
5 vs. 3 samples	0.235	3.473	1.25, 1.8

REFERENCES

Cowden, DJ, 1957, *Statistical Methods in Quality Control*, Prentice-Hall, Inc., Englewood Cliffs NJ, pp. 63

Green, SB, and Salkind, NJ, 1997, *Using SPSS for Windows: Analyzing and Understanding Data*, Prentice Hall, Upper Saddle River, NJ

Palsat, D, McMillan, C, 1994, "End product specifications - eight years of experience by Alberta Transportation and Utilities", 1994 International Road Federation Conference Proceedings, Alberta Transportation and Utilities, Calgary, Canada

Rilett, LR, 1998, "Effect of variability on end product specifications", *Journal of Construction Engineering and Management*, Vol. 124, No. 2, pp. 139-145

APPENDIX A - NOTATION

The following symbols were used in this paper.

c	=	value dependent on gamma function
H_0	=	null hypothesis
H_1	=	alternate hypothesis
n	=	number of samples
$P[]$	=	probability function of terms inside square brackets
S	=	sample standard deviation
w	=	width of the confidence interval
$\bar{X}$	=	sample mean
z	=	standardized normal distribution value
α	=	level of confidence
Γ	=	gamma function
μ	=	population mean value
π	=	pi
σ	=	population standard deviation

Subscripts

$3, 4, 5$	=	value taken from a sample size of three, four or five respectively
-----------	---	--