TIME OF SAMPLING STRATEGIES FOR ASPHALT PAVEMENT QUALITY ASSURANCE

Brenda McCabe¹, Simaan AbouRizk², Members, Jim Gavin³

ABSTRACT

The cost of quality assurance programs in asphalt road construction is high in part because of the need for daily testing. This paper demonstrates the methodology used to investigate the timing of sampling and its effect on the quality assurance program. Fourteen lots over two highway paving projects were tested twice, during construction on a daily basis, and after construction was complete. Three quality measures were compared, namely asphalt content, degree of compaction, and aggregate gradation. The test data during construction was obtained from two sources: the consultant's quality assurance (DQA) and the contractor's quality control (CQC). One-way ANOVA tests and T-tests were used to compare the means. In addition, the arithmetic difference between the means was reviewed to understand how the time of sampling might affect contractor payment adjustments.

Key words: quality assurance, quality control, time of sampling, asphalt pavement, post-construction sampling

BACKGROUND

Government agencies at all levels 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. Several issues were considered including the number of samples taken [McCabe et al. 1999], the time at which the samples were taken, and alternative nondestructive technologies for evaluating Quality Assurance (QA) measures. This paper is concerned with the second of these three studies, namely the time of sample taking.

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 current practice involves the collection of core samples for each lot as it is completed with further testing in a laboratory setting for each of the QA measures.

A study was undertaken to examine the difference between samples extracted at the completion of construction and samples taken during construction. As the cost of the

¹ Assistant Professor, Department of Civil Eng'g, University of Toronto, Toronto, ON, Canada, M5S 1A4

² Professor, Department of Civil and Environmental Eng'g, University of Alberta, Edmonton, AB, Canada, T6G 2G7

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

sampling is related to the frequency of sample-taking and the associated laboratory testing of the samples, QA costs could be reduced by decreasing the number of trips taken to the construction site by QA technicians. For this study, focus was placed on the sampling process for three QA measures: asphalt content, degree of compaction, and aggregate gradation.

This paper describes the methodology used to conduct the study, and is organized in the following manner. The analysis methodology is first reviewed. The results for each of the QA measures are discussed. Finally, conclusions and recommendations are provided.

POST-CONSTRUCTION SAMPLING

Post-construction Quality Assurance (PQA) is a testing methodology whereby the test samples for quality assurance are taken all at once, after construction has completed. The current methodology for sampling is to take samples as each lot is completed, resulting in daily trips on average to the site by QA technicians during construction.

To facilitate the study, sections of a highway undergoing pavement construction were sampled daily (lot-by-lot) during construction in the conventional manner and then again after the completion of construction. Contrary to cost-saving objectives, this study caused an increase in QA costs due to the double sampling. For this reason, only fourteen lots over two projects were included in the study. This relatively small sample will be used to illustrate the methodology and provide insight to the effects of post-construction sampling.

The purpose of this study was to learn if there exists statistically significant differences between the mean test values of the post construction samples compared to the during construction samples for three QA measures, namely asphalt content, pavement compaction and aggregate gradation. Possible sources of variance were considered to be different sampling locations, testing equipment and procedures, testing personnel, and changes in material properties with time (e.g. pavement compaction could increase with time due to traffic). All test procedures were conducted according to AT&U standards. The study was not designed to investigate the relative contribution of any of possible sources of variance.

Data Characteristics

Data from two 1996 projects containing both during-construction and post-construction quality assurance results were provided by AT&U for the analysis. Six lots were sampled for PQA from Project A, and eight lots from Project B. In each case, data from during-construction sampling were from two sources: the consultant's during-construction quality assurance (DQA), and the contractor's quality control (CQC). Both the contractor's and the consultant's during-construction results were compared to the post-construction results.

Research Hypothesis

Two hypotheses were tested:

$$H_0: \mu_{DQA} = \mu_{PQA}$$

$$H_1: \mu_{DQA} \neq \mu_{PQA}$$

and

$$H_0: \mu_{CQC} = \mu_{PQA}$$

$$H_1: \mu_{CQC} \neq \mu_{PQA}$$

where:

μ_{DQA} is the mean of the samples from during-construction quality assurance testing conducted by the project consultant commissioned by AT&U

μ_{CQC} is the mean of the samples from during-construction quality control collected by the contractor performing the construction work

μ_{PQA} is the mean of post-construction samples collected by an independent consultant commissioned by AT&U. (Note that this is not the same consultant as the project consultant)

Analysis Approach

Number of Samples Required

The minimum number of lots, n , required for 95% confidence that the difference in the means of during-construction and post-construction sampling fall within an acceptable width or range can be evaluated by Equation 1.

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

where:

$z_{\alpha/2}$ is the standardized normal distribution value for the desired confidence

σ is the population standard deviation which is estimated by the sample

w is the width of the differences covering both positive and negative deviation at which point it is determined that the difference will not significantly affect the contractor's payment adjustment.

In this case, the objective is to determine, with a confidence of 95%, how many lots are required to accurately determine the difference between post construction and during construction means. The number of lots required is dependent on the value of the interval or width, w , that is acceptable for the differences between the means, and the standard deviation, σ , of that difference. The width of the acceptable difference will determine the number of lots required for analysis. The value of z is equal to 1.96 for a confidence of 95%.

Analysis of Variance (ANOVA)

Analysis of variance, or ANOVA, is a method of testing the null hypothesis that two or more group means are equal in the population, by comparing the sample variance estimated from the group means to that estimated within the groups. In this case, ANOVA was used to compare the means of the two sample sets for each of the lots tested. The ANOVA method consists of a series of calculations that define the variance measures of the sample components, shown in Table 1, where N = total number of samples, and t = number of treatments being compared. If the test of significance is large, then there is

sufficient reason to believe that the null hypothesis can be rejected by the data.

Table 1: ANOVA Calculations

Statistic [1]	Equation [2]	Degrees Of Freedom [3]
Correction term (C)	$C = [\sum(\text{all observations})]^2/N$	$df(C) = 1$
Sum of Squares (SS)	$SS = \sum[(\text{each observation})^2] - C$	$df(SS) = N-1$
Sum of Squares due to Treatment (SSTr)	$SSTr = [\sum(\text{treatment sum})^2 / t] - C$	$df(SSTr) = t-1$
Sum of Squares due to Error (SSE)	$SSE = SS - SSTr$	$df(SSE) = (N-1)-(t-1) = N-t$
Mean Sum of Squares due to Error (MSE)	$MSE = SSE/df(SSE) = s^2$	
Mean Sum of Squares due to Treatment (MSTr)	$MSTr = SSTr / df(SSTr)$	
Test of Significance (F statistic)	$F \text{ statistic} = MSTr/MSE$	

For example, consider five samples each from two populations ($t=2$, $N=10$) shown in Table 2. The null hypothesis, H_0 , is that the two means are equal. According to the equations set out in Table 1, the values required to perform the ANOVA analysis are as shown.

Table 2: Example Data

Treatment [1]	Samples					Sum [7]
	i [2]	ii [3]	iii [4]	iv [5]	v [6]	
A	3.00	3.10	2.90	3.05	2.99	15.04
B	3.50	3.55	3.40	3.49	3.51	17.45
						32.49

$$\begin{aligned}
 df(C) &= 1 \\
 df(SS) &= N-1 &= 9 \\
 df(SSTr) &= t-1 &= 1 \\
 df(SSE) &= df(SS)-df(SSTr) &= 9-1 &= 8
 \end{aligned}$$

$$\begin{aligned}
 C &= [\sum(\text{all observations})]^2/N &= [32.49]^2/10 &= 105.56 \\
 SS &= \sum [(\text{each observation})^2] - C &= 106.175 - 105.560 &= 0.615 \\
 SSTr &= [\sum(\text{treatment sum})^2 / t] - C &= [(15.04^2 + 17.45^2) / 5] - 105.56 &= 0.581
 \end{aligned}$$

SSE	=SS - SSTr	= 0.615-0.581	= 0.034
MSE	=SSE/df(SSE)	= 0.034/8	= 0.0043
MSTr	=SSTr / df(SSTr)	= 0.581/1	= 0.581
F	=MSTr/MSE	= 0.581/0.0043	= 135.12

Because F is large, the null hypothesis (that the means of the two treatments are equal) can be rejected.

Analysis of the Means

ANOVA analysis was used to compare the three means (PQA, DQA, and CQC) simultaneously. For this project, SPSS 8.0 [Green and Salkind 1997] software was used to perform the ANOVA analysis. A level of confidence of $\alpha=0.05$ was used for the analysis. This means that the results may be viewed with a $1-\alpha$, or 95% confidence. In all cases, the data has been confirmed to be approximately normal through standard tests for normality. If the ANOVA results indicated the lot means for the three samples were equal, then no further analysis was done. However, if the ANOVA results indicated that the means were not equal, then independent T-tests were performed between PQA and DQA as well as between PQA and CQC on those lots to determine where the inequality was observed.

The percentage of comparisons that indicated the means were equal (referred to as successful comparisons) was targeted at 85%. In other words, if more than 85% of the lot mean comparisons were successful, then the means in general could be viewed as equal. The value of 85% was based on a subjective level of comfort and then checked by calculating the theoretical probability of achieving this target using the binomial distribution.

Assuming that 95% of the population means are equal (95% confidence), the number of groups tested is n , and 85% of the tests undertaken must be successful for the means to be considered equal, then the number of successful ANOVA analysis required is k , where $k = 0.85 n$. The theoretical likelihood of k or more successful comparisons occurring is:

$$P(\text{Success}) = \sum_{m=k}^n \binom{n}{m} * P^m * Q^{n-m} \quad [2]$$

where $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ evaluates the number of combinations of n items, taken m at a time with m ranging between k and n . P is the population frequency of the successful tests with equal means ($P=0.95$), and Q is the frequency of the unsuccessful tests where $Q = 1-P$.

For example, if 14 lots are tested ($n=14$), then the required number of successful analysis is $k=0.85*14=12$. The probability of observing 12 of 14 successful tests assuming 95% confidence that the means are equal is 97% as evaluated using [2]. The expectation of observing 85% successful tests is therefore conservative.

One may also look at what a successful test means in terms of acceptable differences in the means. Equation 3 provides some insight.

$$p = P \left(\frac{|\bar{X}_{PQA} - \bar{X}_{DQA}|}{\sigma \sqrt{\frac{2}{5}}} \leq z_{\alpha/2} \right) = 1 - \alpha \quad [3]$$

where

$\bar{X}_{PQA}$ is the mean of the post-construction samples

$\bar{X}_{DQA}$ is the mean of the consultant's QA samples

σ is the standard deviation of the samples

z is the normalized statistic for normal distributions

α is the level of confidence

$P()$ is the probability that the difference of the means term would be equal to or less than z .

Therefore, with a confidence of 95% ($\alpha=0.05$), and consequently $z_{\alpha/2}=1.96$, the difference between the means:

$$|\bar{X}_{PQA} - \bar{X}_{DQA}| \leq z_{\alpha/2} * \sigma \sqrt{\frac{2}{5}} \Rightarrow |\bar{X}_{PQA} - \bar{X}_{DQA}| = 1.240\sigma \quad [4]$$

Using the average standard deviation over all of the aggregate gradation lots and sieve sizes of 0.994, the theoretical difference between the aggregate gradation means is 1.233 using [4].

In addition to the ANOVA analysis, the arithmetic difference between the PQA lot means and the during-construction lot means was evaluated. If the mean difference is approximately zero, then the PQA means are neither consistently greater than nor consistently less than the during-construction lot means, and the effect on the contractors' payment adjustments would be negligible. The standard deviation provides indication of the spread of the differences. Finally, the maximum absolute difference is the greatest difference between the PQA and the during-construction means found in the analysis without regard to whether it was higher or lower.

Analysis Results for Asphalt Content

Asphalt content is the weight proportion of asphalt binder relative to the weight of dry aggregate used, expressed as a percent. Normally, specifications require the measured asphalt content to be within $\pm 0.3\%$ of target values. Payment adjustments are applied for pavement material outside of this tolerance zone. Of the fourteen lots used in the analysis, one was eliminated because the variance of PQA was not equal to the variance of either the DQA or the CQC samples. (Equal variance is an assumption of ANOVA and if this assumption is not met, then the results may not be reliable.) Of the remaining thirteen, 62% (8 of 13) of the CQC and 85% (11 of 13) of the DQA results indicated the means were equal. Therefore, the DQA and the PQA means could be considered equal but the CQC and PQA could not.

Table 3 shows other relevant information about the samples. The mean difference

between the sample averages is approximately zero in both comparisons implying that the PQA sample means were neither consistently higher nor lower than the means of the standard test taken during construction. The mean differences, 0.0061% and -0.045%, are significantly smaller than the $\pm 0.3\%$ variance allowed in the specification and should have negligible effect on contractor payment adjustments. The maximum difference is also shown without indication of whether it is a positive or negative value but is within two standard deviations in both cases.

Table 3: Analysis of Post-Construction Sampling for Asphalt Content

[1]	$\mu_{PQA} - \mu_{DQA}$ [2]	$\mu_{PQA} - \mu_{CQC}$ [3]
Percent Successful	85%	62%
Mean Difference	0.0061	-0.045
Std. Deviation	0.179	0.200
Maximum Absolute Difference	0.268	0.334

If we want to determine the number of lots required to have 95% confidence that the mean difference between PQA and DQA is $0 \pm 0.1\%$, then by Equation 1, using $w=0.2$ and $\sigma=0.179$ results in $n=12.3$, or 13 lots. This sample was therefore sufficient.

Analysis Results for Percent Pavement Compaction

Percent Pavement Compaction is expressed as a percentage of the laboratory standard compaction (Marshall method) for the specified pavement mix. The specified percent compaction is 97% with payment adjustments provided for pavements compacted to a higher (bonus) or lower (penalty) standard. Of the fourteen lots used in the analysis, thirteen, or 93% of the CQC and twelve, or 86% of the DQA results indicated the means were equal. The results, when compared to the previously set minimum of 85%, show in both cases that the means may be considered equal. The actual differences between the during-construction values and the post-construction values are shown in Table 4. The maximum difference is also shown without indication of whether it is a positive or negative value.

Table 4: Analysis of Post-Construction Sampling for Percent Compaction

[1]	$\mu_{PQA} - \mu_{DQA}$ [2]	$\mu_{PQA} - \mu_{CQC}$ [3]
Percent Successful	86%	93%
Mean Difference	0.383	0.526
Std. Deviation	0.817	0.929
Maximum Absolute Difference	1.84	2.28

Finding the means equal was a rather surprising result. The authors expected the post-construction compaction means to be significantly higher than the means from samples taken during construction. Note, however, that the mean difference shown in Table 4 is above zero in both cases. This indicates that the DQA compaction results were consistently lower than the PQA by 0.383 Percent of Compaction and lower than the CQC by 0.526 Percent of Compaction on average. Although the ANOVA analysis did not find the difference statistically significant, this may affect the contractors' payment adjustments such that payments to the contractor would increase.

By Equation 1, the number of samples required to have 95% confidence that the mean of the differences will fall within a range of $0 \pm 0.5\%$ so $w=1\%$ using $\sigma=0.817$ is $n=10.2$ or 11 lots

Analysis Results for Aggregate Gradation

Aggregate gradation is expressed as a percent of material passing a series of specified sieve sizes. Fourteen lots, each consisting of nine sieve tests, were utilized in the analysis. Of the 120 analysis performed, nine were rejected because the samples did not have approximately equal variances. As mentioned, ANOVA requires an assumption of equal variance between the samples. If this is not the case, the ANOVA results are not reliable, therefore the samples were removed. Of the remaining 111, 77% of the PQA vs. CQC were successful, and 82% of the PQA vs. DQA were successful i.e. indicated the means of the gradation sampling were equal. This is a surprisingly low result. Further investigation showed that the analysis involving comparisons between the CQC and the PQA on one of the projects contained a high number of rejections (33%) of the null hypothesis, i.e. failures to find the means were equal. If these data are removed from the analysis, over 80% of the means are equal of the remaining data. However, this is still not sufficient to meet our 85% minimum limit.

The actual differences between the during-construction values and the post-construction values are shown in Conclusion

This paper demonstrated the methodology used to compare time of sampling for three pavement quality assurance measures, namely, asphalt content, degree of compaction and aggregate gradation. Time of sampling was compared for during-construction sampling and post-construction sampling. A summary of the analysis result is shown in Table 6.

Table 5. The mean difference for the PQA vs. DQA is close to zero at 0.073%, where the PQA vs. CQC shows a higher average difference of 0.458%. Since adjustments are generally made with deviations of 1% or more from the specified gradation, the effect of taking the samples post-construction would be negligible to the contractor's payment adjustments. Again, using Equation 1 to calculate the number of samples for 95% confidence that the difference of the means is $0 \pm 0.3\%$ so $w=0.6\%$ using

$\sigma=1.014$ is $n=43.9$ or 44 lots.

Conclusion

This paper demonstrated the methodology used to compare time of sampling for three pavement quality assurance measures, namely, asphalt content, degree of compaction and aggregate gradation. Time of sampling was compared for during-construction sampling and post-construction sampling. A summary of the analysis result is shown in Table 6.

Table 5: Analysis of Post-Construction Sampling for Aggregate Gradation

[1]	$\mu_{PQA} - \mu_{DQA}$ [2]	$\mu_{PQA} - \mu_{CQC}$ [3]
Percent Successful	82%	77%
Mean Difference	0.073	0.458
Std. Deviation	1.014	1.432
Maximum Absolute Difference	2.6	4.15

Both asphalt content and percent compaction comparisons between the DQA and PQA met the minimum criteria of 85% successful comparisons with 85% and 86%, respectively, and the means could be considered statistically equal. Although the aggregate gradation fell just below the standard with 82% successful tests, the mean difference was very close to zero and would likely have a negligible effect on contractor payment adjustments. The mean difference for the percent compaction was greater than zero at 0.383. This is not surprising as compaction increases as the construction equipment and traffic continues to drive over the road surface before samples are taken. The effect of this result on the contractor would be an increased compaction and therefore potentially higher bonus payment to the contractor.

Table 6: Summary for the Analysis of Post-Construction Sampling

	Percent of Successful Tests (Minimum 85% Required)		Mean Difference	
	PQA vs. DQA	PQA vs. CQC	PQA vs. DQA	PQA vs. CQC
Asphalt Content	85%	62%	0.018	-0.032
Percent Compaction	86%	83%	0.383	0.526
Aggregate Gradation	82%	77%	0.073	0.458

None of the comparisons between the DQA and CQC met the 85% criteria, and the mean difference for the percent compaction and aggregate gradation were not greater than zero at 0.526 and 0.458 respectively. Although the purpose of this study was not to

identify the sources of variance between the time of sampling, the differences between the CQC and PQA results are interesting.

The results of this study show a promising direction for the reduction of quality assurance costs. Further research should be undertaken to identify the causes of the variance in during- and post-construction sampling.

REFERENCES

Green, SB, and Salkind, NJ, (1997), *Using SPSS for Windows: Analyzing and understanding data*, Prentice-Hall, Englewood Cliff, NJ

McCabe, B, AbouRizk, S, and Gavin, J, (1999), "Sample Size Analysis for Asphalt Pavement Quality Control", ASCE Journal of Infrastructure Systems, Vol.5, No.4, pp. 118-123

APPENDIX A - NOTATION

c	=	value dependent on	distribution value
γ	=	gamma function	
H_0	=	null hypothesis	α = level of confidence
H_1	=	alternate hypothesis	Γ = gamma function
n	=	number of samples	μ = population mean value
$P()$	=	probability function of	π = pi
terms inside brackets			σ = population standard deviation
S	=	sample standard deviation	
w	=	width of the confidence interval	Subscripts
$\bar{X}$	=	sample mean	3, 4, 5 = value taken from a sample size of three, four or five respectively
z	=	standardized normal	