

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Application of Sampling Strategies for Hot-Mix Asphalt Infrastructure: Quality Control-Quality Assurance Sampling; Specification for Performance Test Requirements

Bor-Wen Tsai<sup>1</sup>, Jiangmiao Yu<sup>1,2</sup> and Carl L. Monismith<sup>1</sup>

<sup>1</sup>*University of California at Berkeley,*

<sup>2</sup>*South China University of Technology,*

<sup>1</sup>*U.S.A.*

<sup>2</sup>*P.R.C.*

## 1. Introduction

Due to the lack of a rational, effective, and systematic quality control-quality assurance (QC/QA) methodology, the nonconformity of construction quality with design requirements for public works, especially for civil engineering infrastructure systems, can result in increased expenditures over time. Thus, development of a rational QC/QA methodology to ensure that the construction quality complies with the design requirements should have a high priority. The limited sample size constrained by the consideration of cost and time may result in the misjudgement that the construction quality does not meet the design requirements.

In this chapter, the effects of sampling size, sampling strategies, and acceptance/rejection criteria for QC/QA projects using statistically based decision making in hot-mix asphalt (HMA) construction are presented. Also, there has developed an increased interest recently in ensuring that the HMA as placed will meet certain performance requirements by measuring the actual performance parameters on test specimens prepared from in situ samples rather than from surrogate values such as asphalt content and aggregate gradation. Examples include direct measures of mix permanent deformation characteristics and fatigue characteristics, mix stiffness, and degree of compaction as measured by air-void content.

Determination of sample size is primarily based on an acceptable error level for a performance parameter specified by the agency. It is not uncommon to base quality assurance by many agencies on three samples. Through the  $t$  distributions, discussion is presented as to why it is not appropriate to take only this number of samples for quality-assurance. Based only on three samples in a large project, the agency will have insufficient power to reject the null hypothesis given that the null hypothesis is false unless the project quality delivered by the contractor is extremely poor so that the agency is confident enough to reject the project.

In addition to providing a general introduction to fundamental statistics and hypothesis testing, two case studies are used to clarify the relationships among sampling size, sample strategies, and performance specifications (or acceptance/rejection criterion). These include the following:

(1) A QC/QA case study is used to illustrate a methodology to determine strategies for a sampling scheme and selection of sample size for QC/QA for HMA construction to ensure that the acceptable level of a mix parameter is obtained with the same risk to the contractor and the agency. A sampling scheme and sampling size based on statistical simulation of a fixed length of a one-lane-width placement of HMA are discussed. Sample size is based on the combination of the sample size of the contractor and that of the agency to balance the risk to both organizations which will result in a mix that will meet the minimum performance requirement. An example is presented for the placement of 15,000 tons of HMA according to the California Department of Transportation (Caltrans) QC/QA requirements. For this total tonnage, the contractor and agency are assumed to perform a specific number of performance tests using the California stabilometer methodology for QC and QA.

(2) A QA case study is used to illustrate the application of the use of uniform design (UD) as a sampling strategy to ensure that the most representative sampling scheme can be achieved with a specified sample size. A sampling scheme using uniform design and sampling size through statistical simulation of a fixed length of a two-lane-width placement of HMA with several segregation data patterns is discussed. Based on the simulation, a QA guideline is developed by inspecting the accuracy of sample mean and the precision of sample standard deviation criteria combined with the application of the UD table is proposed and verified with two full scale pavement sections by measured air-void contents (measure of degree of compaction).

## **2. Case I: quality control-quality assurance sampling strategies for hot-mix asphalt construction**

The effects of sampling strategies and size on statistically based decision making in hot-mix asphalt (HMA) construction are presented. For sample sizes agreed upon by the agency and the contractor, an acceptable level for an HMA mix parameter is determined with risk balanced between the two organizations. With increased emphasis on specific performance requirements, the use of performance tests on HMA specimens prepared from in situ samples is developing. Examples include direct measures of mix stiffness and permanent deformation characteristics. A measure of rutting resistance, the stabilometer S-value, is used by the California Department of Transportation (Caltrans) for quality control-quality assurance (QC/QA) projects. Although the S-value was used for this simulation because extensive tests were available, this approach is applicable to any performance measures already in use, such as HMA thickness or compacted air-void content. A sampling scheme and sampling size through statistical simulation of a fixed length of a one-lane-width placement of HMA are discussed. Sample size is based on the combination of the sample size of the contractor and that of the agency to balance the risk to both organizations and results in a mix that meets the minimum performance requirement.

### **2.1 Hypothesis testing of inequality**

The acceptance or rejection of the null hypothesis,  $H_0$ , is referred to as a decision. Therefore, a correct decision is made in situations where (1) the  $H_0$  is correctly accepted if  $H_0$  is true and (2) the  $H_0$  is correctly rejected if the  $H_0$  is not true. As shown in the following for a decision based on a sample, when the null hypothesis is valid, the probability  $\alpha$  of erroneously rejecting it is designated as the Type I error (or seller's risk), i.e.,  $\alpha = P\{\text{Type I error}\} =$

$P\{\text{reject } H_0 \mid H_0 \text{ is true}\}$  ; when the null hypothesis is not true, the probability  $\beta$  of erroneously accepting it is named the Type II error (or buyer’s risk), i.e.,  $\beta = P\{\text{Type II error}\} = P\{\text{fail to reject } H_0 \mid H_0 \text{ is false}\}$ .

	Truth about the population	
	$H_0$ True	$H_0$ Not True
Reject $H_0$	Type I Error ( $\alpha$ )	Correct Decision
Accept $H_0$	Correct Decision	Type II Error ( $\beta$ )

The power is defined as the probability  $1 - \beta$  of correctly rejecting  $H_0$  if  $H_0$  is not true, i.e.,  $1 - \beta = P\{\text{reject } H_0 \mid H_0 \text{ is false}\}$ . Hence, from the viewpoint of the agency (the buyer), it is necessary to have the power as high as possible; likely, from the perspective of the contractor (the seller), the Type I error should be as minimum as possible.

### 2.1.1 Testing inequality $\mu \geq C_s$ and size of test $\alpha$

The objective is to test the null hypothesis that  $H_0 : \mu \geq C_s$  from the viewpoint of the contractor. The relevant  $t$  statistic is given by  $t = \frac{\hat{\mu} - C_s}{SE(\hat{\mu})}$ , where  $\hat{\mu}$  is the sample mean of the stabilometer tests,  $C_s$  is the minimum specification limit for the stabilometer test, and  $SE$  the standard error. The critical region for the  $t$  test of size  $\alpha$  of the null hypothesis is given by  $t \leq -t_{1-\alpha, n-p}$ , where  $n = n_1 + \dots + n_p$  and  $p$  the number of laboratories. In other words, the  $t$  statistic lies in the acceptance region  $t > -t_{1-\alpha, n-p}$  if and only if

$$(\hat{\mu} - C_s) / SE(\hat{\mu}) > -t_{1-\alpha, n-p} \Rightarrow \hat{\mu} > C_s - t_{1-\alpha, n-p} SE(\hat{\mu}) \tag{1}$$

Note that the critical region for the  $t$  test of size  $\alpha = 0.05$  of the null hypothesis  $H_0 : \mu \geq C_s$  can be given by  $t \leq -t_{1-\alpha, n-p} \cong -\Phi(1 - \alpha) = -\Phi(0.95) = -1.64485$ , if  $n - p \gg 0$ , where  $\Phi$  is the distribution function of a standard normal distribution. The size of test  $\alpha = 0.05$  represents that at most a 5% chance is allowed to erroneously reject a valid null hypothesis; that is, there is a 95% chance that  $H_0$  is accepted if  $H_0$  is valid.

### 2.1.2 Test power, sample size, and operating-characteristic curve

Suppose that the hypothesis is not true, that is,  $\mu < C_s$  (the opposite of  $\mu \geq C_s$ ). Then the power as shown by Stone (Stone, 1996) is:

$$\begin{aligned} 1 - \beta &\cong P\left(\frac{\hat{\mu} - C_s}{SD(\hat{\mu})} \leq -z_{1-\alpha}\right) \\ &= P\left(\frac{\hat{\mu} - \mu}{SD(\hat{\mu})} \leq -z_{1-\alpha} + \frac{C_s - \mu}{SD(\hat{\mu})}\right) \\ &= \Phi(-z_{1-\alpha} + \delta) \end{aligned}$$

where  $\delta = \frac{|\mu - C_S|}{SD(\hat{\mu})}$  and  $z_{1-\alpha}$  is quantile of a standard normal distribution. For the specified  $\alpha$  and  $\beta$  levels under the null hypothesis that  $H_0: \mu \geq C_S$ , Table 1 lists the  $SD(\hat{\mu})$ , test power, and required sample sizes for the case with the agency and the contractor. The test power equation shown in Table 1 indicates that the power of testing a null hypothesis is actually a standard normal distribution function in terms of the test of size  $\alpha$ ,  $d$  ( $d = \frac{|\mu - C_S|}{S_p}$ ), and the number of tests. Figure 1 plots power versus  $d$  with  $\alpha = 0.05$  at various numbers of tests, designated as the operating-characteristic curves. Several observations can be addressed in the following:

1. With the same number of tests and power level, increasing  $\alpha$  will decrease the value of  $d$ ; alternatively, at the same number of tests and a fixed value of  $d$ , increasing  $\alpha$  will increase the power.
2. At fixed levels of  $\alpha$  and  $\beta$ , increasing the number of tests will reduce the value of  $d$ .
3. For  $n=4$  to ensure that the test power is greater than 0.95,  $d > 1.645$ , i.e.,  $d \cong \frac{|\hat{\mu} - C_S|}{S_S} > d_{0.95}$ . In other words, if the sample mean of tests  $\hat{\mu}$  is either  $\hat{\mu} > C_S + d_{0.95} \cdot S_S$  or  $\hat{\mu} < C_S - d_{0.95} \cdot S_S$ , then the agency has enough power to confidently accept or reject the null hypothesis  $H_0: \mu \geq C_S$ . If  $\hat{\mu}$  lies in the range of  $(C_S - d_{0.95} \cdot S_S, C_S + d_{0.95} \cdot S_S)$ , then the agency does not have enough power with  $n = 4$ . Thus the number of tests has to be increased to reach the same level of power.
4. The test power approaches the test of size  $\alpha$  as  $d \rightarrow 0$ .

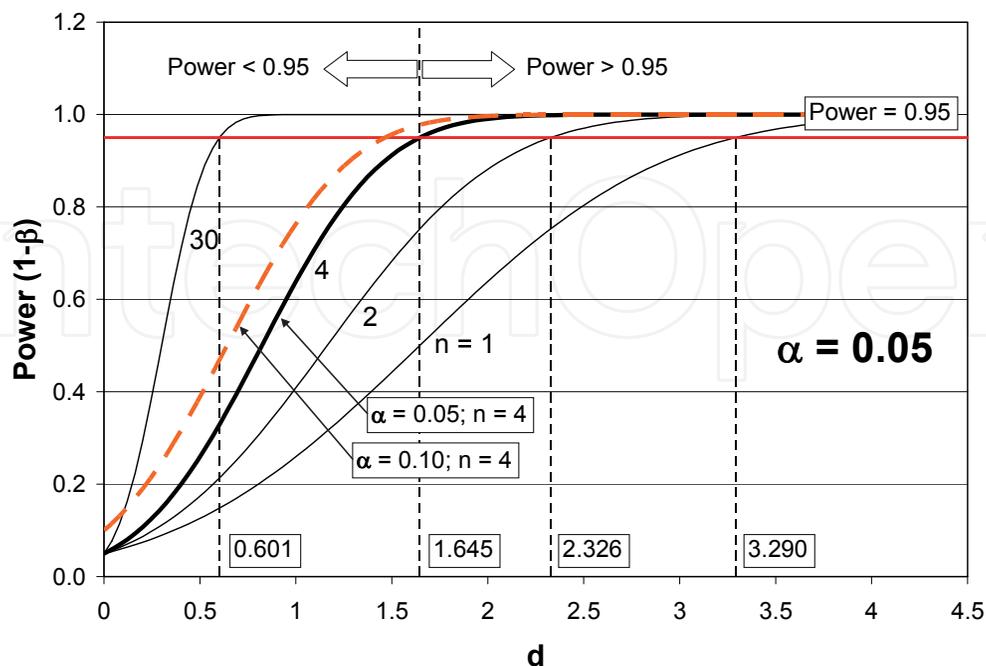


Fig. 1. Operating-characteristic curves with  $\alpha = 0.05$ .

<b>The agency and the contractor:</b>	
$n_1 = k \cdot n_2, 0 < k \leq 1$	
<b>(<math>n_1</math> : number of QA samples; <math>n_2</math> : number of QC samples.)</b>	
$SD(\hat{\mu})$	$\frac{S_p}{p} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{S_2}{2} \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$
Test Power $1 - \beta$	$\Phi\left(-z_{1-\alpha} + 2 \cdot d \cdot \sqrt{\frac{k \cdot n_2}{1+k}}\right)$
Sample size	$n_2 \approx \frac{1+k}{k} \cdot \left(\frac{z_{1-\beta} + z_{1-\alpha}}{2 \cdot d}\right)^2$
Minimum Requirement of contractor	$\hat{\mu} \geq C_S - \frac{t_{1-\alpha, n_1+n_2-2}}{2} \cdot \sqrt{\frac{1+k}{k \cdot n_2}} \cdot S_p$
Upper and Lower bounds of Agency	$\hat{\mu} \geq C_S + \frac{z_{1-\alpha} + z_{1-\beta}}{2} \cdot \sqrt{\frac{1+k}{k \cdot n_2}} \cdot S_p$ (upper bound)
	$\hat{\mu} \leq C_S - \frac{z_{1-\alpha} + z_{1-\beta}}{2} \cdot \sqrt{\frac{1+k}{k \cdot n_2}} \cdot S_p$ (lower bound)

Note:

The pooled sample variance,  $S_p^2$ , is defined as,

$$S_p^2 = \frac{\sum_{i=1}^{n_1} (x_{1,i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2,i} - \bar{x}_2)^2 + \dots + \sum_{i=1}^{n_p} (x_{p,i} - \bar{x}_p)^2}{n_1 + n_2 + \dots + n_p - p}; \text{ if } p = 1, \text{ then } S_p = S_S.$$

$z_{1-\alpha}$  and  $z_{1-\beta}$  are quantiles of a standard normal distribution;  $\Phi$  is the distribution function of a standard normal distribution.

$$d = \frac{|\mu - C_S|}{S_2}; \mu = \frac{\mu_1 + \mu_2}{2}$$

Table 1. Test power, required sample size, minimum requirement of contractor, and upper and lower bounds of agency.

### 2.1.3 Size of test $\alpha$ and power $1 - \beta$

For the contractor, under the null hypothesis  $H_0 : \mu \geq C_S$ , the acceptance region for the  $t$  test of size  $\alpha$  is given by Equation 1, that is,  $\hat{\mu} \geq C_S - t_{1-\alpha, n-p} \cdot SE(\hat{\mu})$ .

For the agency, as noted earlier, the power of a test under the null hypothesis is given by

$$1 - \beta \cong \Phi(-z_{1-\alpha} + \delta) \Rightarrow z_{1-\beta} \cong -z_{1-\alpha} + \delta$$

where  $\delta = \frac{|\mu - C_S|}{SD(\hat{\mu})} \cong \frac{|\hat{\mu} - C_S|}{SE(\hat{\mu})}$ ;  $z_{1-\alpha}$  and  $z_{1-\beta}$  are quantiles of a standard normal distribution.

Therefore, to satisfy the power requirement of the agency,  $\delta \geq z_{1-\beta} - (-z_{1-\alpha})$ , i.e.,

$$\frac{|\hat{\mu} - C_S|}{SE(\hat{\mu})} \geq z_{1-\alpha} + z_{1-\beta}$$

$$\Rightarrow \hat{\mu} \geq C_S + (z_{1-\alpha} + z_{1-\beta}) \cdot SE(\hat{\mu}), \text{ or } \hat{\mu} \leq C_S - (z_{1-\alpha} + z_{1-\beta}) \cdot SE(\hat{\mu}) \quad (2)$$

The  $C_S + (z_{1-\alpha} + z_{1-\beta}) \cdot SE(\hat{\mu})$  will be designated as the upper bound and  $C_S - (z_{1-\alpha} + z_{1-\beta}) \cdot SE(\hat{\mu})$  the lower bound of  $1 - \beta$  power.

It should be noted that (1) if  $\beta = 0.5$ , then  $z_{1-\beta} = 0$  and (2)  $z_{1-\alpha} \cong t_{1-\alpha, n-p}$ , if  $n - p \gg 0$ .

Thus, Equation 1 is equivalent to the lower bound of Equation 2. Based on Equation 1, the minimum requirement of the contractor, and Equation 2, the upper and lower bounds of power requirement of the agency, the case of the agency and the contractor is defined in Table 1.

Figure 2 illustrates plots of the upper and lower bounds at various power levels of the agency and the minimum requirement of the contractor under  $H_0 : \mu \geq 37$  in terms of  $\mu$  and sample size,  $n_2$ . The minimum requirements of the contractor in Figure 2 are plotted based on the  $t$ - distribution and standard normal distribution. It will be noted that the two curves coincide after  $n_2 = 10$ . From Table 1 and Figure 2, two observations can be made:

1. It is very important to recognize that the minimum requirement of the contractor actually matches the lower bound of 0.5 power of the agency.
2. The distance enclosed by the upper and lower bounds at a specified power level decreases with smaller  $S_p$ , larger  $\alpha$  and  $\beta$ , larger  $k$  ( $0 < k \leq 1$ ), and, more importantly, larger sample size.

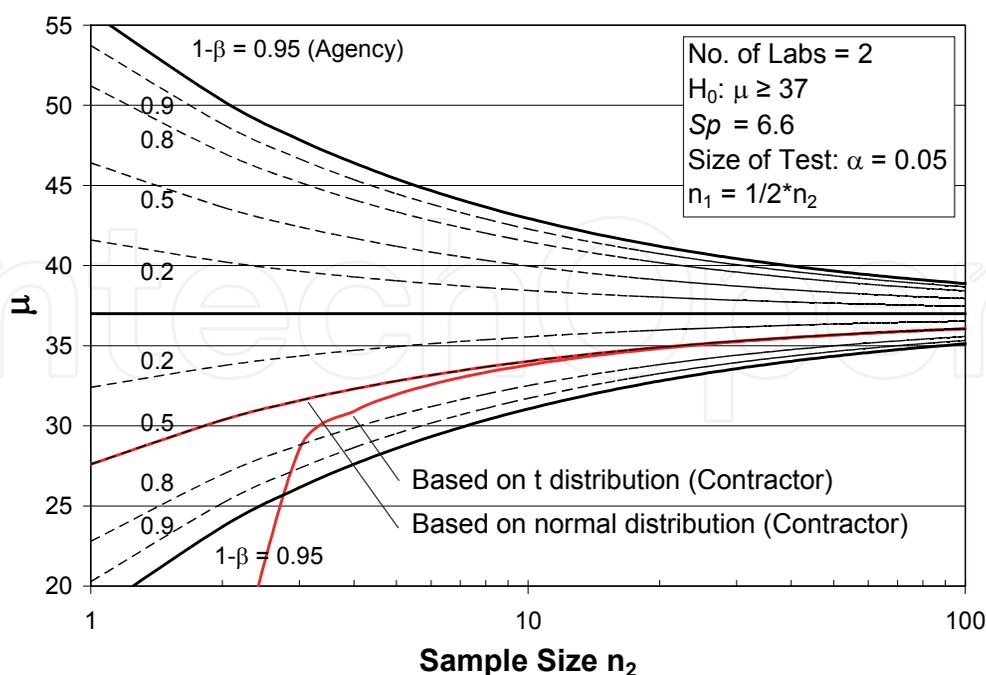


Fig. 2. Minimum stability requirements of the contractor and power requirement of the agency under the same null hypothesis.

## 2.2 QC/QA demonstration example

In this demonstration example 15,000 tons of HMA will be placed on 20 sublots (750 tons per subplot). The contractor is required to conduct 20 tests ( $n_2$ ), i.e., one test per subplot. The number of tests conducted by the agency ( $n_1 = k \cdot n_2$ ) will include the minimum required by the agency according to Caltrans specifications, i.e,  $k = 0.1$  (2 tests in this case); in addition, determinations will be made for four tests ( $k = 0.2$ ), six tests ( $k = 0.3$ ), and eight tests ( $k = 0.4$ ). The minimum stabilometer S-value has been set at 37 (Type A HMA) (California Department of Transportation [CALTRANS], 2007), and a standard deviation  $S_p$  is used for the S-value for tests between two laboratories of 6.6 (Paul Benson, private communication transmitting analyses of stabilometer test results for periods 1967– 1970 and 1995-1999). The demonstration example will include sampling consistency between QC and QA, sampling stabilization of  $S_p$ , and minimum requirements for both the agency and the contractor.

To conduct the sampling size simulation, several assumptions were made:

1. Lane width: 12 ft (3.66 m),
2. Unit weight of HMA - 145 lb/ft<sup>3</sup> (2,323 kg/m<sup>3</sup>),
3. HMA layer thickness - 8 in. (20 cm), and
4. One stability sample is represented by a 4 × 4-in (10 × 10-cm) square with each square assigned a normalized stability value.

For these assumptions, the 15,000 tons of HMA will produce a section ~26,000 ft (7,925 m). long and 12 ft (3.66 m) wide. This results in a  $N(0,1)$  stability population of  $12 \times 3 \times 26,000 \times 3 = 2,808,000$  data points to generate three types of data patterns as schematically shown in Figure 3: (1) random pattern, (2) transverse strip pattern with 40 vertical strips, and (3) longitudinal strip pattern with 6 horizontal strips. The  $N(0,1)$  distribution is separated by the points of quantiles into several intervals, e.g., 6 intervals for transverse strip pattern or 4 intervals for longitudinal strip pattern as shown in Figure 3. These intervals are then permuted to vary randomly across the  $x$ -direction or the  $y$ -direction of a lane of HMA paving. Those points within the interval are also randomly distributed over the transverse strip or the longitudinal strip.

The sampling scheme used was illustrated in Figure 4 with cases of  $M \times N$  cells ( $N$  [ $y$ -direction] = 1;  $M$  [ $x$ -direction] = 10, 20, 30, 40, 50, 100, 200, and 500). That is, one random QC sample from each cell and one random QA sample from one random cell of  $n_1 = k \cdot n_2$  random transverse strips. A total of 8 cases were simulated over three data patterns. Each case, per data pattern, was simulated 200 times.

To verify the minimum sampling size for an HMA paving strip is to show (1) no apparent difference of sampling consistency between the contractor (QC) and the agency (QA) and (2) stabilization of the pooled sample estimate of standard deviation of stability value,  $S_p$ . (Tsai & Monismith, 2009).

In each sampling simulation, the normalized stability values form a distribution with mean and standard deviation; hence, when repeated 200 times, the standard deviations will form another distribution. For each case, the standard deviations of the standard deviation distributions (SDSD) were calculated for QC and QA respectively. The difference of SDSD between QA and QC were used as an index to represent the sampling consistency between the agency (QA) and the contractor (QC).

Likewise, for each simulation, the  $S_p$  was calculated based on the equation in Table 1; hence, when repeated 200 times, the standard deviation of the  $S_p$  distribution will be used to inspect its stability over the  $M \times N$  domain.

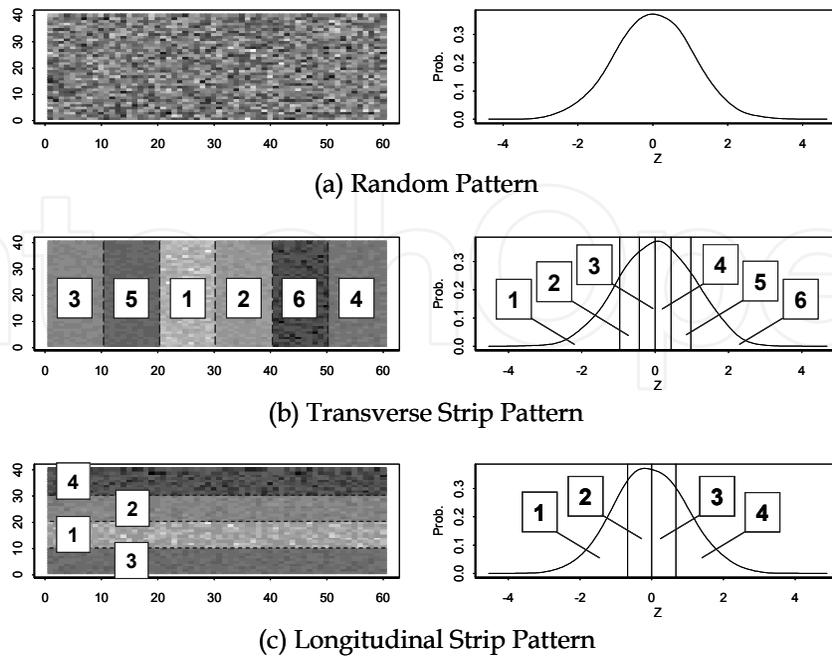


Fig. 3. Schematic illustration of three data patterns: (a) random pattern, (b) transverse strip pattern, and (c) longitudinal strip pattern.

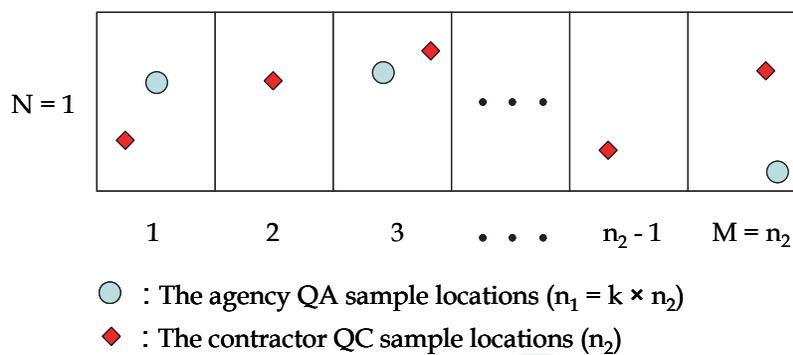


Fig. 4. Sampling scheme.

Figure 5a illustrates the simulation results for sampling consistency between QC and QA at various  $k$  values in terms of global smoothed line over three different data patterns. As would be expected, the sampling consistency between QC and QA increases as the  $k$  value increases. Figure 5b indicates that sampling stabilization of  $S_p$  depends only on the contractor’s sampling size,  $n_2$ , rather than the  $k$  value.

From a series of operating-characteristic curves for the four  $k$  values and two  $\alpha$  values (5% and 10%), the values in Table 2 were determined for the required minimum value of  $S$ , termed  $\mu_{min}$ . With Figure 6a as an example, under the condition that  $\alpha = 5\%$ ,  $n_2 = 20$ ,  $k = 0.2$ , and power = 0.95,  $d$  has to be greater than 0.902 to satisfy the agency’s power requirement; that is,  $\hat{\mu}$  has to be greater than 42.95 so that the agency has power 0.95 to clearly accept the contractor’s mix. Figure 6b shows a smaller  $d$  (0.803) will be obtained when the  $\alpha$  value is increased to 10%.

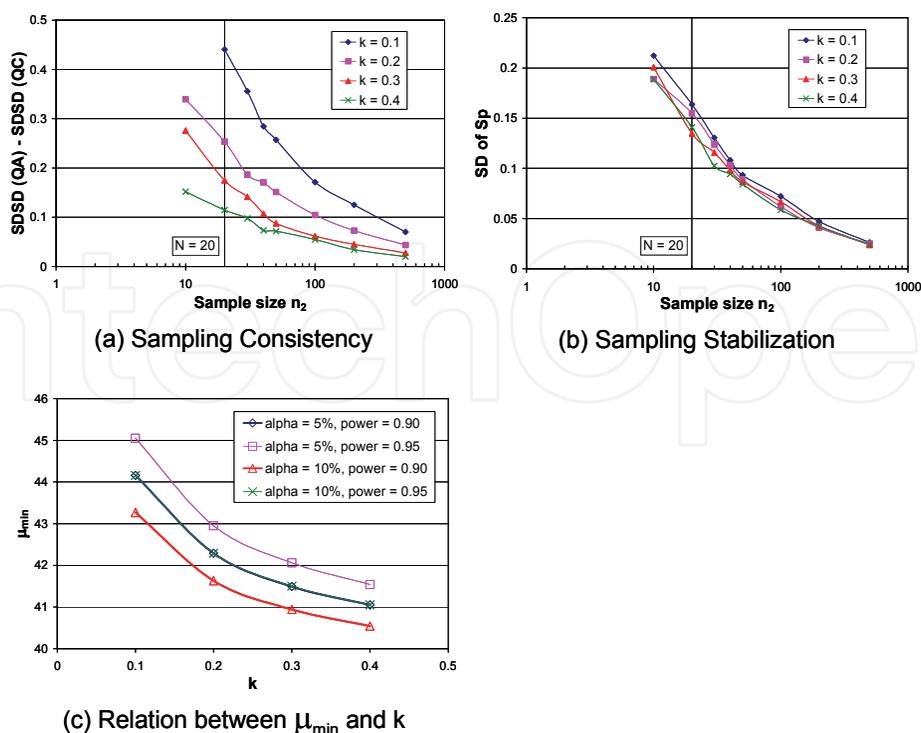


Fig. 5. Summary of simulation results: (a) sampling consistency; (b) sampling stabilization; and (c) relationship between  $k$  and  $\mu_{\min}$ .

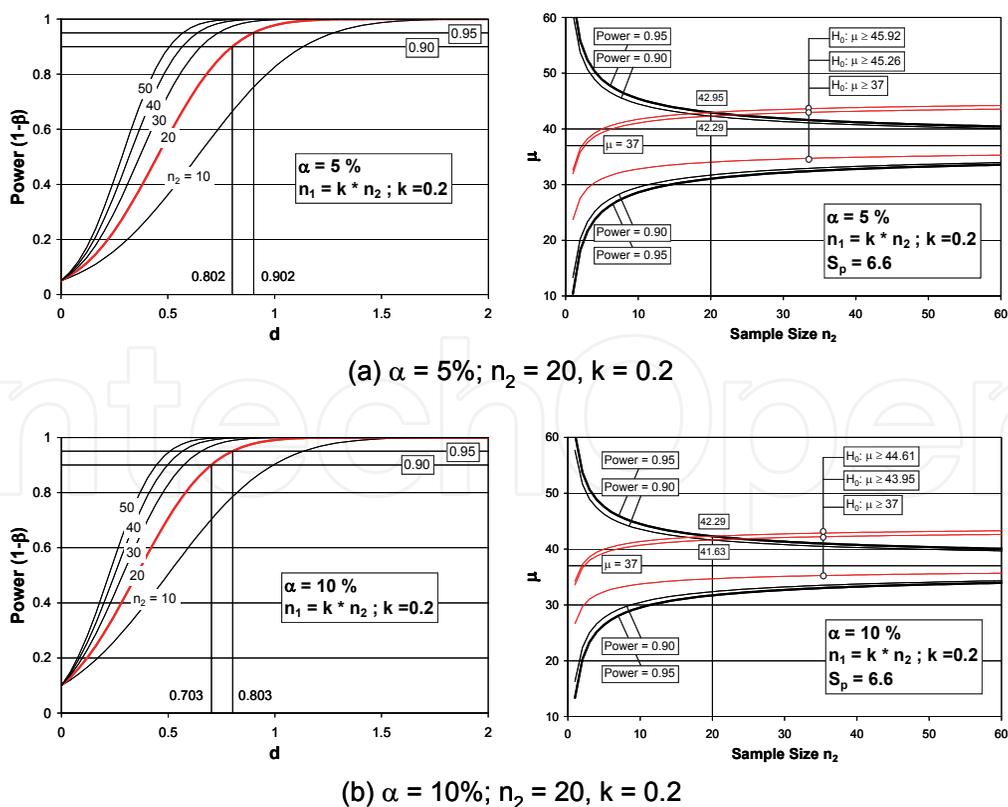


Fig. 6. Examples of operating-characteristic curves and  $\mu_{\min}$  required to meet the agency's power requirement and the contractor's minimum requirement: (a)  $\alpha = 5\%$  and (b)  $\alpha = 10\%$ .

Figure 5c illustrates the relationship between  $k$  and the minimum S-value. It is apparent that an increase of  $k$ -value reduces the value of  $\mu_{\min}$ . It is interesting to observe that the curve of  $\alpha = 5\%$  and power = 0.90 is exactly the same as the curve of  $\alpha = 10\%$  and power = 0.95. From Figure 5c, it is also shown that the higher  $\mu_{\min}$ -criterion is needed if both the agency and the contractor require a high power level and a low  $\alpha$ -level, whereas if both the agency and the contractor require a low power level and a high  $\alpha$  level, then the  $\mu_{\min}$  criterion can be much smaller.

$\alpha$ (Contractor)	Power $1-\beta$ (Agency)	$k$	$d$	Acceptance $\mu_{\min}$	Target Hypothesis (Contractor)
5%	0.90	0.1	1.085	44.16	$H_0^* : \mu \geq 48.19$
		0.2	0.802	42.29	$H_0^* : \mu \geq 45.26$
		0.3	0.680	41.49	$H_0^* : \mu \geq 44.02$
		0.4	0.613	41.05	$H_0^* : \mu \geq 43.31$
	0.95	0.1	1.220	45.05	$H_0^* : \mu \geq 49.07$
		0.2	0.902	42.95	$H_0^* : \mu \geq 45.92$
		0.3	0.766	42.06	$H_0^* : \mu \geq 44.58$
		0.4	0.688	41.54	$H_0^* : \mu \geq 43.81$
10%	0.90	0.1	0.951	43.27	$H_0^* : \mu \geq 46.41$
		0.2	0.703	41.63	$H_0^* : \mu \geq 43.95$
		0.3	0.597	40.94	$H_0^* : \mu \geq 42.91$
		0.4	0.537	40.54	$H_0^* : \mu \geq 42.31$
	0.95	0.1	1.086	44.16	$H_0^* : \mu \geq 47.30$
		0.2	0.803	42.29	$H_0^* : \mu \geq 44.61$
		0.3	0.682	41.50	$H_0^* : \mu \geq 43.46$
		0.4	0.613	41.05	$H_0^* : \mu \geq 42.81$

Note:

Null hypothesis:  $H_0 : \mu \geq 37$ .

$n_1 = k \cdot n_2$  ( $0 \leq k \leq 1$ ), where  $n_1$  is the number of tests of agency;  $n_2$  the number of tests of contractor.

$d = \frac{|\mu - C_S|}{S_p}$ , where  $\mu = \frac{\mu_1 + \mu_2}{2}$ ,  $\mu_1$  the average stabilometer value from

agency;  $\mu_2$  the average stabilometer value from contractor;  $C_S = 37$ ;  $S_p = 6.6$ .

$$\mu_{\min} = d \cdot S_p + C_S.$$

Table 2. Acceptance  $\mu_{\min}$  values and target hypotheses for contractor with combinations of various  $\alpha$  levels, power levels, and  $k$ - values.

### 3. Case II: HMA sampling strategies using uniform experimental design for quality assurance

The application of using uniform design (UD) as a sampling strategy for quality assurance (QA) ensures that the most representative and unbiased sampling scheme can be achieved with the sample size based on an acceptable error level of a hot-mix asphalt (HMA) parameter specified by the agency. Through statistical simulations and demonstration of air-void measurements of two field pavement sections, a QA guideline combined with the UD sampling scheme was developed to justify construction quality using the sample mean and sample standard deviation criteria. This approach can also be applied to any performance measure already in use.

#### 3.1 Uniform experimental design

Statisticians have developed a variety of experimental design methods for different purposes, with the expectation that use of these methods will result in increased yields from experiments, quality improvements, and reduced development time or overall costs. Popular experimental design methods include full factorial designs, fractional factorial designs, block designs, orthogonal arrays, Latin square, supersaturated designs, etc. One relatively new design method is called Uniform Design (UD). Since it was proposed by Fang and Wang in the 1980s (Fang, 1980; Fang et al., 2000; Wang & Fang, 1981), UD has been successfully used in various fields, such as chemistry and chemical engineering, quality and system engineering, computer sciences, survey design, pharmaceuticals, and natural sciences, etc.

Generally speaking, uniform design is a space-filling experimental design that allocates experimental points uniformly scattered in the domain. The fundamental concept of UD is to choose a set of experimental points with the smallest discrepancy among all the possible designs for a given number of factors and experimental runs.

Suppose that there are  $s$  factors in an experiment. Without loss of generality we can assume that the experimental domain is the unit cube  $C^s = [0,1]^s$  (after making a suitable linear transformation). The aim is to choose a set of  $n$  experiment points  $P = \{x_1, \dots, x_n\} \subset C^s$  that is uniformly scattered on  $C^s$ . Let  $M$  be a measure of uniformity of  $P$  such that the smaller  $M$  corresponds to better uniformity. Let  $Z(n,s)$  be the set of sets of  $n$  points on  $C^s$ . A set  $P^* \in Z(n,s)$  is called a uniform design if it has the minimum  $M$ -value over  $Z(n,s)$ , i.e.,  $M(P^*) = \min_{P \in Z(n,s)} M(P)$ .

Many different measures of uniformity have been defined. However, the centered  $L_2$ -discrepancy ( $CD_2$ ) is regarded as one of the most commonly used measures in constructing the UD tables, the reason is that the  $CD_2$  considers the uniformity not only of  $P$  over  $C^s$ , but also of all the projection uniformity of  $P$  over  $C^u$  which is the  $u$ -dimensional unit cube involving the coordinates in  $u$ ,  $P_u$  is the projection of  $P$  on  $C^u$ . Hickernell gave an analytical expression of  $CD_2$  as follows (Fang & Lin, 2003):

$$CD_2(P) = \left[ \left( \frac{13}{12} \right)^s - \frac{2}{n} \sum_{k=1}^n \prod_{j=1}^s \left( 1 + \frac{1}{2} |x_{kj} - 0.5| - \frac{1}{2} |x_{kj} - 0.5|^2 \right) + \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n \prod_{i=1}^s \left( 1 + \frac{1}{2} |x_{ki} - 0.5| + \frac{1}{2} |x_{ji} - 0.5| - \frac{1}{2} |x_{ki} - x_{ji}| \right) \right]^{\frac{1}{2}}$$

where  $x_k=(x_{k1}, \dots, x_{ks})$  is the  $k$ -th experimental point,  $s$  is the number of factors in an experiment,  $n$  is the number of runs.

One of the most noteworthy advantages of the uniform design is that it allows an experiment strategy to be conducted in a relatively small number of runs. It is very useful when the levels of the factors are large, especially in some situations in which the number of runs is strictly limited to circumstances when factorial designs and orthogonal arrays can not be realized in practice.

Given that the strength of uniform design is that it provides a series of uniformly scattered experiment points over the domain, this homogeneity in two factors has physically become the spatial uniformity of sampling from a pavement section in  $x$  and  $y$  directions. The application of uniform design resulted in the generation of a sampling scheme with a UD table consisting of pairs of  $(x, y)$  coordinates.

### 3.2 Fundamental statistics

If  $\bar{x}$  is the sample mean of a random sample of size  $n$  from a normal population,  $X \sim N(\mu, \sigma^2)$ , then  $Z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$  has a standard normal distribution. A  $100(1 - \alpha)\%$

confidence interval (CI) can be defined as (Figure 7a),  $p\left(-z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$ .

Hence, If  $\mu$  and  $\sigma$  are specified, a  $100(1 - \alpha)\%$  confidence interval on  $\bar{x}$  can be then given by

$$\mu - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (3)$$

It can be assumed that the error  $E = |\bar{x} - \mu|$  is equivalent to  $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$  (Figure 7b). Then the required sample size will be

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 \quad (4)$$

That is to say, if  $\bar{x}$  is used as an estimate of  $\mu$ , we can be  $100(1 - \alpha)\%$  confident that the error  $|\bar{x} - \mu|$  will not exceed a specified amount  $E$  when the sample size is  $n = (z_{\alpha/2} \cdot \sigma / E)^2$  (Montgomery & Runger, 2010). If the specified error level is selected as the fraction of standard deviation of  $N(\mu, \sigma^2)$  distribution, i.e.,  $E = |\bar{x} - \mu| = k\sigma$ , where  $k \geq 0$ , then the Equation 4 can be simplified as  $n = (z_{\alpha/2} \cdot \sigma / E)^2 = (z_{\alpha/2} / k)^2$ . It should be noted that  $z_{\alpha/2} = 1.9600$  if  $\alpha = 0.05$ ;  $z_{\alpha/2} = 1.6449$  if  $\alpha = 0.10$ .

The same argument of sample mean can also be applied to sample standard deviation  $s$ . Let  $X_1, X_2, \dots, X_n$  be a random sample of size  $n$  from a normal distribution  $N(\mu, \sigma^2)$ , and let

$s^2$  be the sample variance. Then the random variable  $X^2 = \frac{(n-1)s^2}{\sigma^2}$  has a chi-square ( $\chi^2$ )

distribution with  $n - 1$  degrees of freedom. As shown in Figure 7c, we may write

$$p\left(\chi_{\alpha/2, n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{1-\alpha/2, n-1}^2\right) = 1 - \alpha$$

k $\left( E =  \bar{x} - \mu  \right)$ $(= k\sigma)$	Sample <sup>1</sup> Size	Mean		Standard Deviation		
		Two-Sided <sup>2</sup>		Two-Sided <sup>3</sup>		One-Sided <sup>4</sup>
		$\frac{z_{\alpha/2}}{\sqrt{n}}$	$\sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}}$	$\sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}}$	$\sqrt{\frac{\chi_{1-\alpha, n-1}^2}{n-1}}$	
1.0	4	0.9800	0.2682	1.7653	1.6140	
0.9	5	0.8765	0.3480	1.6691	1.5401	
0.8	7	0.7408	0.4541	1.5518	1.4487	
0.7	8	0.6930	0.4913	1.5125	1.4176	
0.6	11	0.5910	0.5698	1.4312	1.3530	
0.5	16	0.4900	0.6461	1.3537	1.2909	
0.4	25	0.3920	0.7188	1.2807	1.2318	
0.3	43	0.2989	0.7868	1.2128	1.1764	
0.2	97	0.1990	0.8587	1.1411	1.1174	
0.1	385	0.0999	0.9293	1.0707	1.0591	
0.62	10	0.6198	0.5478	1.4538	1.3711	
0.44	20	0.4383	0.6847	1.3149	1.2596	
0.36	30	0.3578	0.7439	1.2556	1.2114	
0.31	40	0.3099	0.7788	1.2208	1.1829	
0.28	50	0.2772	0.8025	1.1971	1.1636	
0.25	60	0.2530	0.8199	1.1798	1.1493	

Note:

$$\text{Sample size is calculated by } n = \left( z_{\alpha/2} \cdot \sigma / E \right)^2 = \left( z_{\alpha/2} / k \right)^2 .$$

The two-sided 100(1-α)% confidence interval of sample mean is calculated by

$$\mu - z_{\alpha/2} \cdot \sigma / \sqrt{n} \leq \bar{x} \leq \mu + z_{\alpha/2} \cdot \sigma / \sqrt{n} .$$

The two-sided 100(1-α)% confidence interval of sample standard deviation is calculated by

$$\sqrt{\chi_{\alpha/2, n-1}^2 / (n-1)} \cdot \sigma \leq s \leq \sqrt{\chi_{1-\alpha/2, n-1}^2 / (n-1)} \cdot \sigma .$$

The one-sided 100(1-α)% confidence interval of sample standard deviation is calculated by

$$s \leq \sqrt{\chi_{1-\alpha, n-1}^2 / (n-1)} \cdot \sigma .$$

Table 3. Summary of 95% confidence intervals of sample mean and sample standard deviation at various error levels and sample sizes for a  $N(0, 1)$  distribution.

If  $s^2$  is the sample variance from a random sample of  $n$  observations from a normal distribution with known or specified variance  $\sigma^2$ , then a two-sided 100(1-α)% confidence interval on  $s$  is

$$\sqrt{\frac{\chi_{\alpha/2, n-1}^2}{n-1}} \cdot \sigma \leq s \leq \sqrt{\frac{\chi_{1-\alpha/2, n-1}^2}{n-1}} \cdot \sigma$$

As for the one-sided 100(1-α)% upper confidence bound as shown in Figure 7d, we may write

$$p \left( \frac{(n-1)s^2}{\sigma^2} \leq \chi_{1-\alpha, n-1}^2 \right) = 1 - \alpha$$

then the confidence upper bound on  $s$  is

$$s \leq \sqrt{\frac{\chi_{1-\alpha, n-1}^2}{n-1}} \cdot \sigma \tag{5}$$

Table 3 summarizes the 95% confidence interval of sample mean and sample standard deviation at various error levels and sample sizes. Notice that the sample size listed in Table 3 was rounded to its ceiling value.

Figure 8a plots the sample size versus the specified error ( $E = |\bar{x} - \mu|$ ) in terms of standard error ( $\sigma$ ) with 95% confidence interval. The two-sided 95% confidence interval on the sample mean and the one-sided 95% upper confidence bound on the sample standard deviation of a  $N(0, 1)$  distribution, as a function of sample size, can be illustrated as shown in Figures 8b and 8c, respectively.

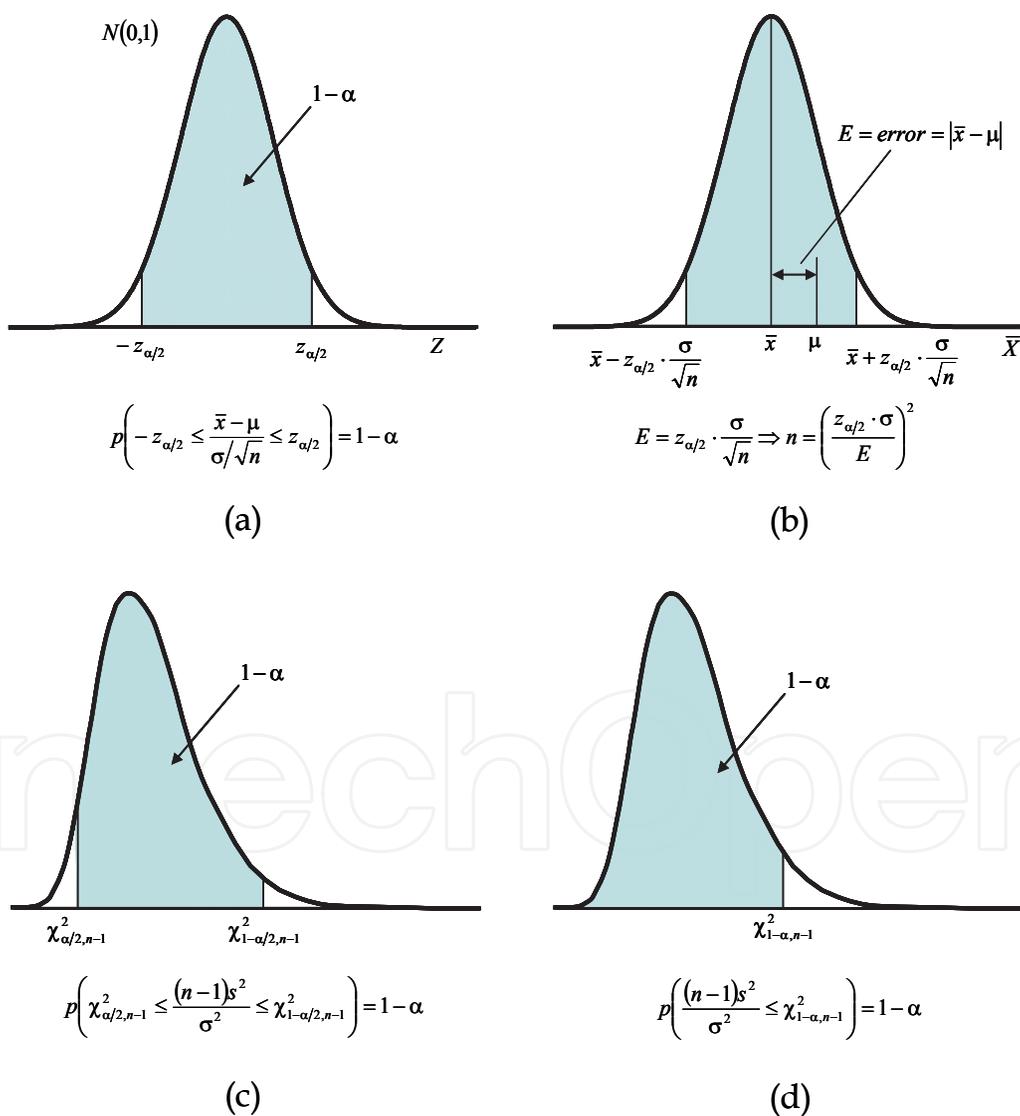


Fig. 7. (a) 100(1- $\alpha$ )% confidence interval of  $N(0, 1)$  distribution, (b) sample size determination with a specified error level, (c) 100(1- $\alpha$ )% two-sided confidence interval of  $\chi^2$  distribution, and (d) 100(1- $\alpha$ )% one-sided confidence interval of  $\chi^2$  distribution.

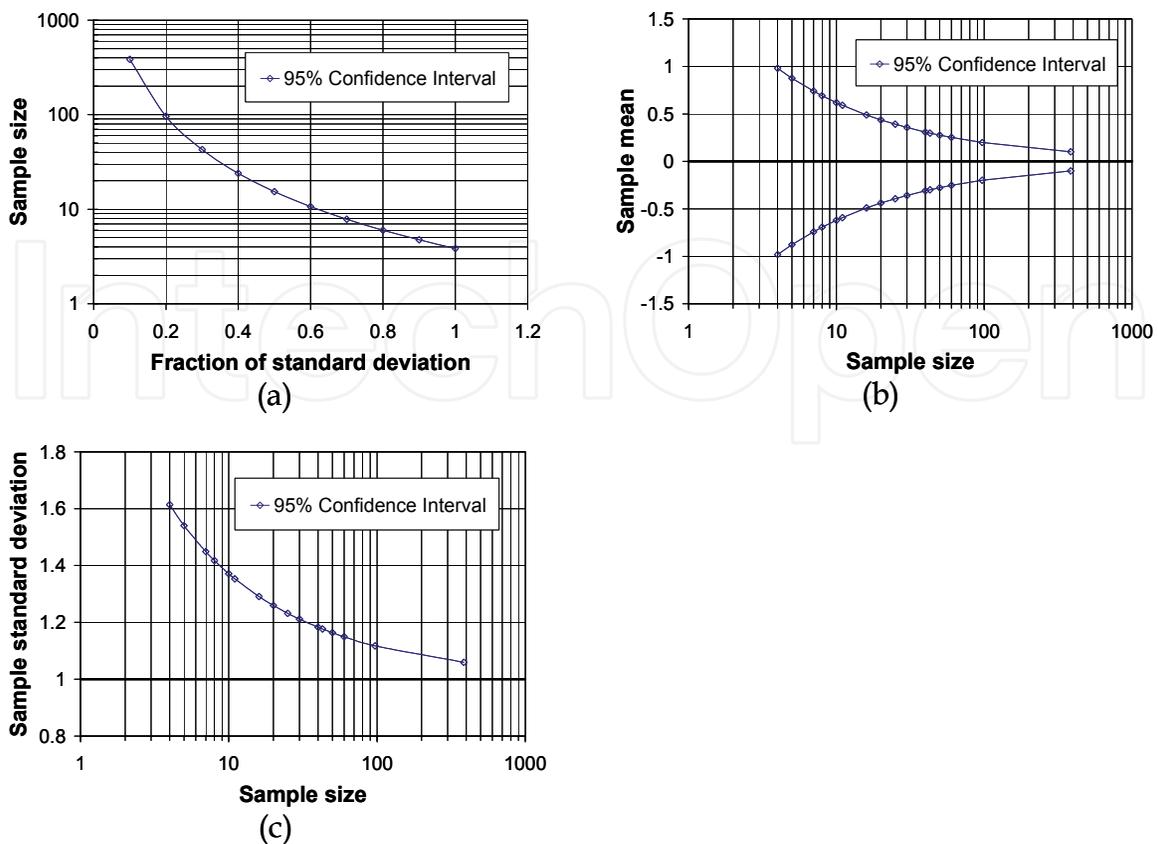


Fig. 8. (a) Sample size versus fraction of standard deviation, (b) 95% two-sided confidence interval of sample mean, and (c) 95% one-sided upper confidence bound of sample standard deviation.

### 3.3 Sampling scheme and size simulation

In this approach, it was assumed that the air-void contents on a project can be represented by a standard normal  $N(0, 1)$  distribution. The data from the  $N(0, 1)$  distribution were used to generate five data patterns: random pattern, central segregation pattern, bilateral segregation pattern, central-bilateral segregation pattern, and block segregation pattern (Figure 9). The reasons for selecting these pattern types are as follows:

1. Random pattern: non-segregation, with ideal construction quality.
2. Central segregation pattern: the gap between two augers of an asphalt paver makes coarse aggregate concentrated near the center of the paved area.
3. Bilateral segregation pattern: the gap between the auger and the lateral board of the asphalt paver makes coarse aggregate concentrated near the bilateral regions of the paved area, or provides less compaction of the side area.
4. Central-bilateral segregation pattern: a combined situation of patterns 2 and 3.
5. Block segregation pattern: as demonstrated in gradation segregation, temperature segregation, uneven compaction, etc.

The segregation horizontal strips as shown in Figures 9b, 9c, and 9d were randomly generated using the data in the shaded area of the  $N(0, 1)$  distribution, which represent higher air-void contents. In the block segregation pattern (Figure 9e), the  $N(0, 1)$  distribution was divided into 6 intervals and the data of each interval were randomly distributed into blocks of pavement sections.

The prospective road section was divided into  $n(X)$  ( $x$ -direction)  $\times$   $n(Y)$  ( $y$ -direction) cells. The  $n(X)$  represents the number of intervals in the  $x$ -direction.  $N$  points were then assigned to these  $n(X) \times n(Y)$  cells. Hence, a sampling scheme was defined by  $n(X)$ ,  $n(Y)$ , and  $N$ . For instance,  $x30y6n30$  represents 30 runs that were assigned to 30 cells of the  $30 \times 6$  cells based on the UD table. The sampling schemes considered in this study were combinations of various numbers of  $n(X)$  and  $n(Y)$ —that is,  $n(X) = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, 60$  and  $n(Y) = 1, 2, 3, 4, 6$ —and  $N = n(X)$ ; however, the cases with  $n(Y) > n(X)$  were excluded, resulting in a total of 62 cases. Each case was assigned a UD table with minimum  $CD_2$  value. Figures 10a through 10c respectively illustrate the example sampling schemes (i.e., UD tables),  $x10y6n10$ ,  $x30y6n30$ , and  $x60y6n60$ , from the uniform design. These sampling schemes are on the same scales of a 900 ft  $\times$  24 ft (274 m  $\times$  7.32 m) pavement section. The black rectangle cell physically represents the area of which one measure should be sampled randomly.

For this sampling simulation, a total of  $2700 \times 72$  points with a standard normal distribution of air-void contents were used to generate five data patterns with the following assumptions:

1. Lane width: 24 ft (7.32 m).
2. Time frame of construction: 1 hour with 900 ft (274 m) of HMA placed, i.e., paver speed = 15 ft/min. (4.57 m/min.).
3. One air-void sample is represented by a 4  $\times$  4-in. (10  $\times$  10-cm) square with each square assigned a normalized air-void value.

Each type of sampling scheme per data pattern was simulated 200 times. For each simulation, the sample mean and sample standard deviation were calculated. It should be noted that the data of each simulation were randomly drawn from the cells specified in the UD table with replacement. Consequently, the distributions of the sample mean and standard deviation were generated after 200 simulations. The boxplot was then utilized to characterize the location and dispersion of sample means and standard deviations.

The boxplot illustrates a measure of location (the median [solid black dot or white strip]), a measure of dispersion (the interquartile range IQR [lower quartile: left or bottom-edge of box; upper quartile: right or top-edge of box]), and the possible outliers (data points with light circle or horizontal line outside the 1.5 IQR distance from the edges of box; the most extreme data points within 1.5 IQR distance are marked with square brackets) and also gives an indication of the symmetry or skewness of the distribution.

The Trellis graph introduced by Cleveland in 1993 (Cleveland, 1993) is a graphical way of examining high-dimensional data structure by means of conditional one-, two-, and three-dimensional graphs. As an example, we would like to determine how the sample mean distribution depends on  $n(X)$ ,  $n(Y)$ , and the data pattern. To inspect this graphically, the simulation results can be split up into groups and can be plot separately as opposed to blurring the effects in a single graph. The Trellis graph of boxplots presented in Figures 11 and 12 was arranged in such a way that each panel consists of all the  $n(Y) = 1, 2, 3, 4, 6$  cases (i.e., 5 boxplots in each panel), each row is made by all the  $N = n(X) = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, 60$  cases (i.e., 13 panels in a row) with the same data pattern, and each column has 5 panels (i.e., 5 data patterns) with the same  $n(X)$ . Thus, for each individual column, the effects of data pattern and  $n(Y)$  can be examined at the specified  $n(X)$ ; for each individual row, the effects of  $n(X)$  and  $n(Y)$  can be inspected at the specified data pattern. The Trellis graph was categorized by  $n(X)$ ,  $n(Y)$ , and five data patterns.

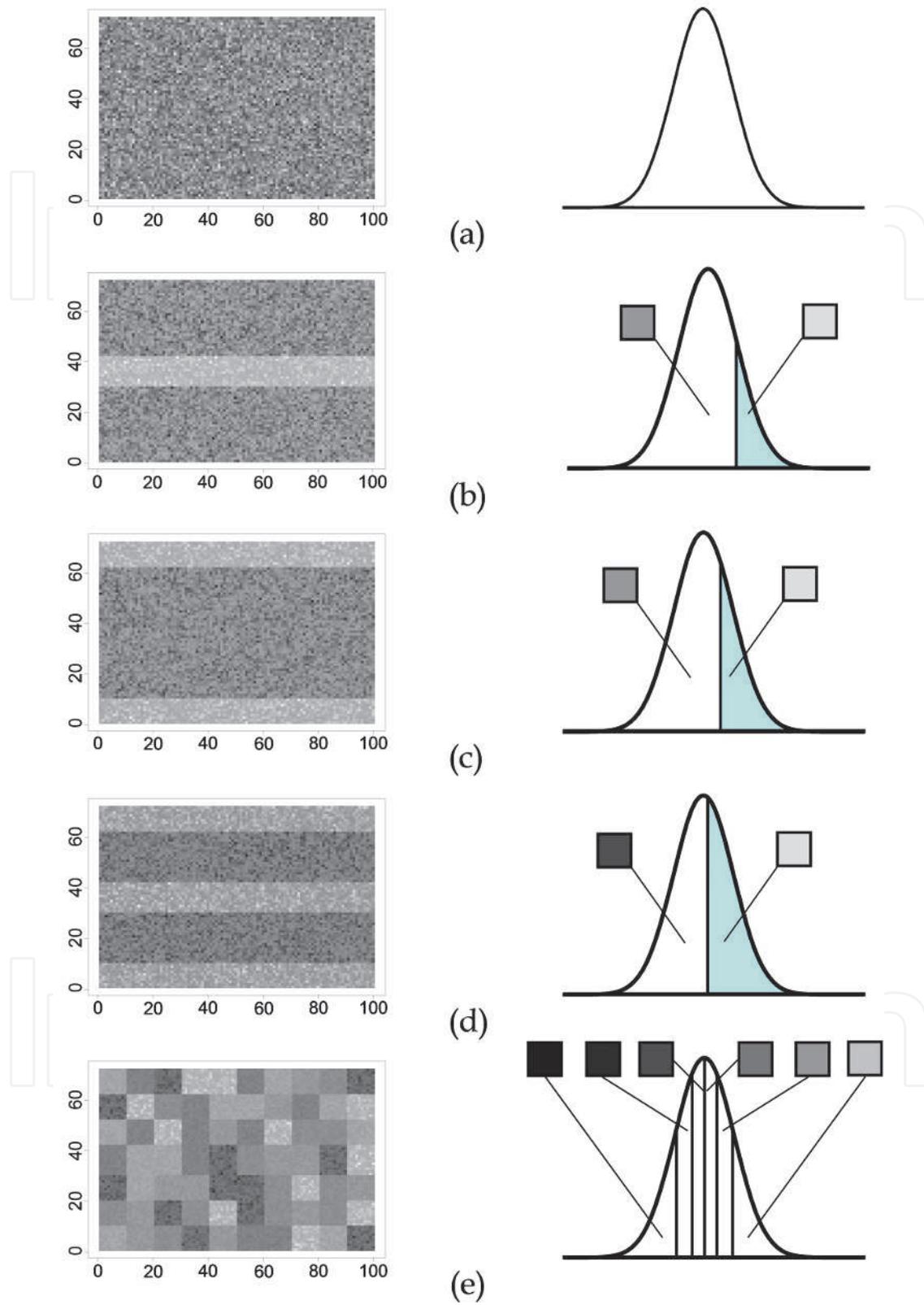


Fig. 9. Schematic illustration of five data patterns: (a) random pattern, (b) central segregation pattern, (c) bilateral segregation pattern, (d) central-bilateral segregation pattern, and (e) block segregation pattern.

The Trellis graphs of the boxplots shown in Figures 11 and 12 summarize respectively the simulation results of sample means and sample standard deviations. Several observations from the Trellis graphs can be made:

1. As  $n(X)$  increases, i.e., increase of  $N$ , the variations of sample mean and standard deviation reduce exponentially regardless of data patterns.
2. For the segregation data patterns 2, 3, and 4, the increase of  $n(Y)$  does benefit the decrease of variation per  $n(X)$  and per data pattern. However, no apparent decrease of variation on random or block segregation patterns was perceived. This implies that the UD table provides a uniform sampling strategy. From the perspective of practice, it is suggested that  $n(Y)$  should be as large as possible to include all the possible data patterns.
3. It should be noted that the distributions of sample standard deviation at small  $n(X)$ s exhibit unsymmetrical and skewed distributions due to the intrinsic properties of  $\chi^2$  (chi-square) distribution. From the point of view of HMA construction, the  $100(1-\alpha)\%$  one-sided upper bound is suggested for the judgment of sample standard deviation. As a result, the smaller the sample standard deviation, the more uniform the construction quality of HMA. Also, from Figures 8b, 8c, 11, and 12, it is apparent that the change of variation decreases sharply at the beginning and the rate of change of variation stabilizes after  $N = 20 \sim 30$ .

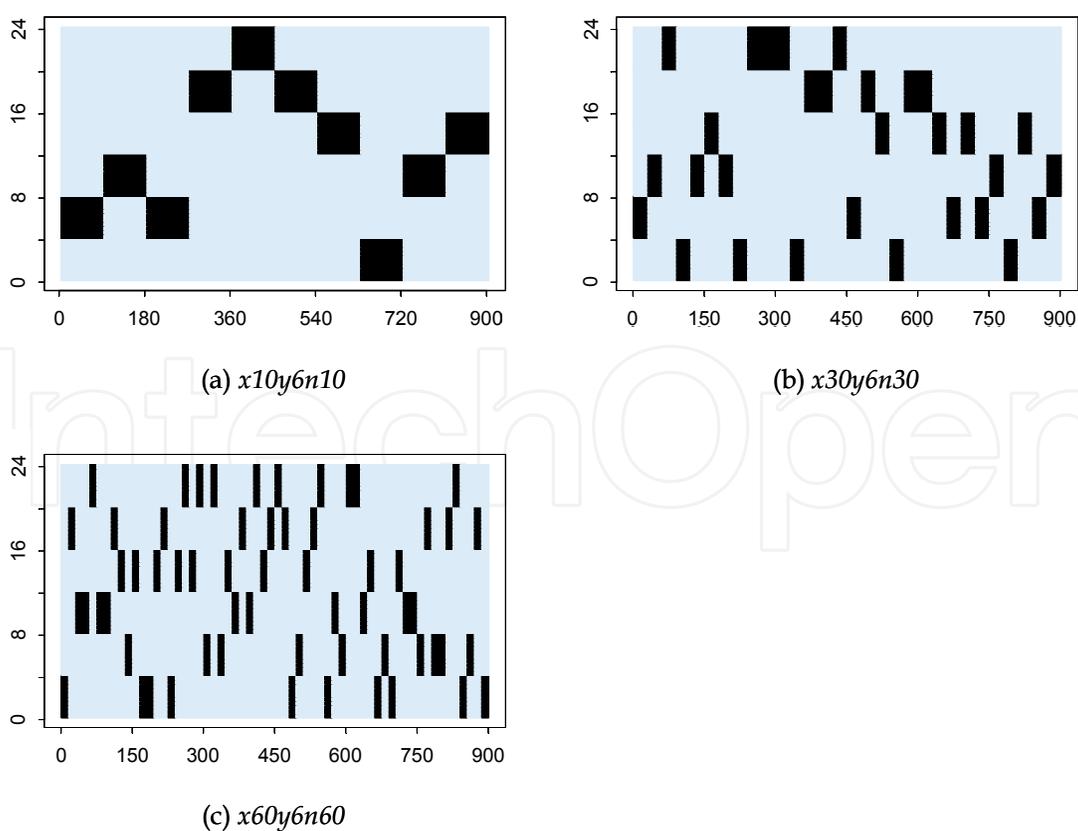


Fig. 10. Examples of UD tables (a)  $x10y6n10$ , (b)  $x30y6n30$ , and (c)  $x60y6n60$ .

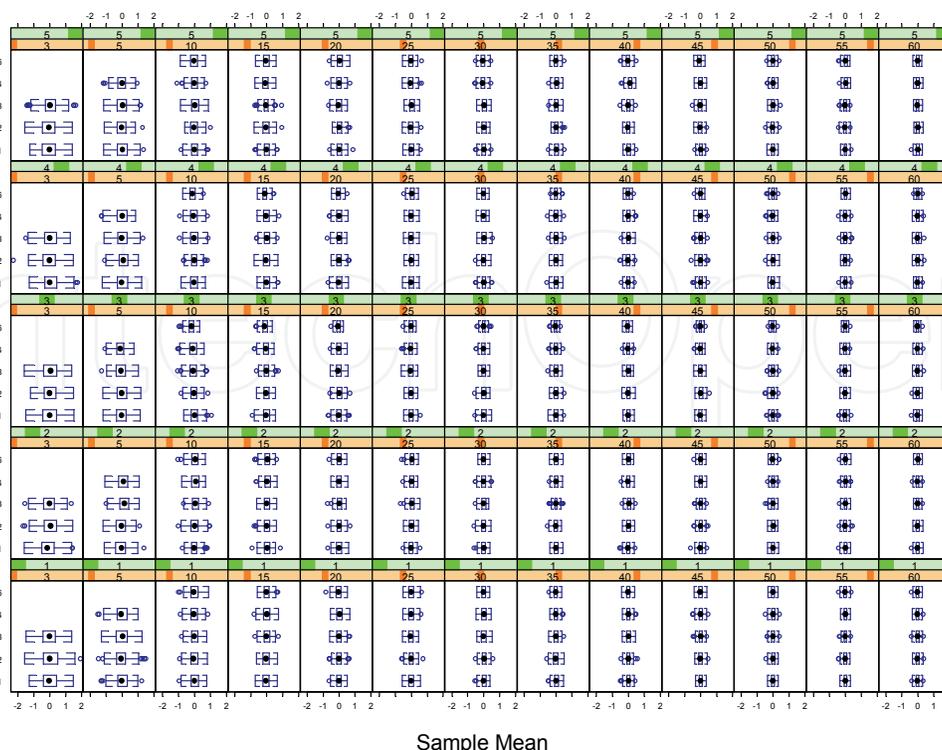


Fig. 11. Trellis graph of boxplots of sample mean categorized by  $n(X)$ ,  $n(Y)$ , and five data patterns with  $N = n(X)$ .

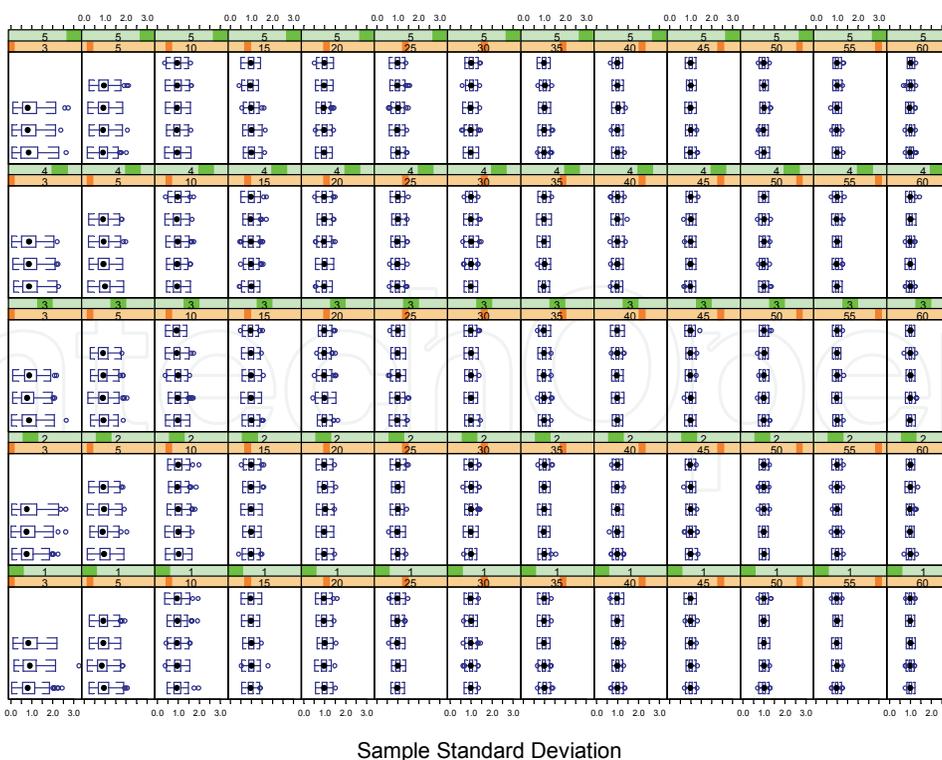


Fig. 12. Trellis graph of boxplots of sample standard deviation categorized by  $n(X)$ ,  $n(Y)$ , and five data patterns with  $N = n(X)$ .

### 3.4 UD demonstration example using two field sections

In this demonstration example, the percent air-void content data of two field pavement sections each with 164 ft (50 m) in length and 36 ft (11 m) in width were acquired by the Pavement Quality Indicator (PQI), which is a non-nuclear density measurement device calibrated with core samples. The percent air-void content was taken by a 3.3 × 3.3-ft (1 × 1-m) square. These two pavement sections served as the "testing sections" of which the paving operation, compaction pattern/effort, and other construction details were verified and corrected (if necessary) by the contractor. Several performance tests were comprehensively conducted by the agency to guarantee that the pavement quality of the whole project met the specifications afterwards. The material properties of two pavement sections, AC-13 and AC-20, are as follows.

	Pavement Section AC-13	Pavement Section AC-20
<b>Asphalt mix type</b>	Asphalt concrete with nominal maximum aggregate size (NMAS) 13 mm	Asphalt concrete with NMAS 20 mm
<b>Binder type</b>	SBS modified binder (equivalent to PG76-22)	
<b>Aggregate type</b>	Diabase (fully crushed, hard rock)	Granite (fully crushed)
<b>Design binder content</b>	5.6%	4.8%
<b>Target air-void content</b>	$N(\mu, \sigma^2) = N(5, 1)$ : mean 5%; standard deviation 1%.	
<b>Acceptable air-void content range</b>	$5 \pm 2$ %, i.e., $P(3 \leq AV \leq 7) = 0.95$ of a $N(5, 1)$ distribution	

The measured percent air-void contents are illustrated in Figures 13a and 13b respectively for the AC-13 and AC-20 pavement sections. As can be seen from the figures, the AC-13 section presents high air-void content on the section edges and seems to have a wide variation of air-void content. The AC-20 section appears to have more uniform distribution of air-void content.

To illustrate the proposed QA approach, it was decided that 20 points (20 runs) will be sampled to ensure that the agency is 95% confident that the error  $|\bar{x} - \mu|$  will not exceed  $0.44\sigma$ , i.e., 0.44 percent (Table 3). Two UD tables (Figures 13c and 13d) were generated for both sections which are subdivided into 10 ( $x$ -direction) by 11 ( $y$ -direction), i.e.,  $x10y11n20$ . In this case study, the sampling for each UD table was conducted only once. Figures 13e and 13f summarized the sampled, measured, and specified distributions of air-void content. Several findings can be addressed in the following:

1. The sampled distribution based on the UD table matches the measured distribution reasonably well: AC-13 sampled  $N(6.29, 1.40^2)$  versus AC-13 measured  $N(6.18, 1.43^2)$ ; AC-20 sampled  $N(5.41, 1.22^2)$  versus AC-20 measured  $N(5.12, 1.24^2)$ .
2. The sample mean, 6.29, of AC-13 section is outside the 95% CI (4.56, 5.44) (Table 3); therefore, it is identified as an "inaccurate" distribution. The sample standard deviation, 1.40 exceeds the 95% one-sided upper bound 1.26 (Table 3); thus, it is designated as an "imprecise" distribution. As a result, the construction quality of AC-13 section is not acceptable because of its "inaccurate" and "imprecise" distribution.
3. On the contrary, the construction quality of AC-20 section is not rejected because of its "accurate" and "precise" distribution: the sample mean 5.41 lies in the 95% CI although

on the high side; the sample standard deviation 1.22 is a slightly less than the 95% one-sided upper bound 1.26.

Accordingly, a proposed QA sampling guideline can be provided by the agency in the following steps:

1. Specify an error level ( $E$ ) of sample mean in terms of standard deviation of the specified distribution  $N(\mu, \sigma^2)$ , i.e.,  $E = |\bar{x} - \mu| = k\sigma$ .
2. Specify the  $\alpha$  value to construct a  $100(1 - \alpha)\%$  confidence interval.
3. Determine the sample size based on the Equation 4,  $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$ .
4. Generate a uniform design table (UD table) as the sampling scheme, the  $X$  factor should have  $n$  (sample size) levels, i.e.,  $N = n(X) = n$ . It is suggested the  $Y$  factor should have at least 3 levels per lane, i.e.,  $n(Y) \geq 3$  per lane.
5. Randomly take the measurement from each  $(x, y)$  cell specified in UD table.
6. Check the sample mean  $\bar{x}$  from  $n$  observations. If  $\mu - \frac{z_{\alpha/2}}{\sqrt{n}} \cdot \sigma \leq \bar{x} \leq \mu + \frac{z_{\alpha/2}}{\sqrt{n}} \cdot \sigma$  (Equation 3), then we accept the sample mean; otherwise, the sample mean has been rejected and the agency has to reject the project.
7. Check the sample standard deviation  $s$  if step 6 has been satisfied. If  $s \leq \sqrt{\frac{\chi^2_{1-\alpha, n-1}}{n-1}} \cdot \sigma$  (Equation 5), then we accept the sample standard deviation; otherwise the project should be rejected because of non-uniformity of construction quality.

#### 4. Why is it inappropriate to take only three samples?

It is not uncommon for agencies to base QA on three samples. However, the following discussion using  $t$  distribution is presented showing why it is inappropriate to take only this number of samples for quality assurance. When estimating the mean of a normally distributed population with unknown mean  $\mu$  and unknown variance  $\sigma^2$ , the  $t$  distribution should be applied especially with small sample size. Let  $X_1, X_2, \dots, X_n$  be a random sample from a normal distribution with unknown mean  $\mu$  and unknown variance  $\sigma^2$ . The random variable  $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$  has a  $t_{n-1}$  distribution with  $n - 1$  degrees of freedom,

where  $\bar{X}$  is the sample mean and  $S$  the sample standard deviation. Now if  $\mu = \mu_0$ , then

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - \mu_0 - \delta + \delta}{S/\sqrt{n}} = \frac{\bar{X} - (\mu_0 + \delta)}{S/\sqrt{n}} + \frac{\delta\sqrt{n}}{S}$$

If  $\delta = 0$ , then it is called the central  $t_{n-1}$  distribution. When the true value of the mean is  $\mu = \mu_0 + \delta$ , the distribution of  $T_0$  is termed the noncentral  $t_{n-1}$  distribution with the noncentrality parameter  $\delta\sqrt{n}/S$ . Based on the definition of type II error ( $\beta$ ):  $P\{\text{fail to reject } H_0 \mid H_0 \text{ is false}\}$  under the hypothesis testing  $H_0 : \mu = \mu_0, H_1 : \mu < \mu_0$ , the Type II error is

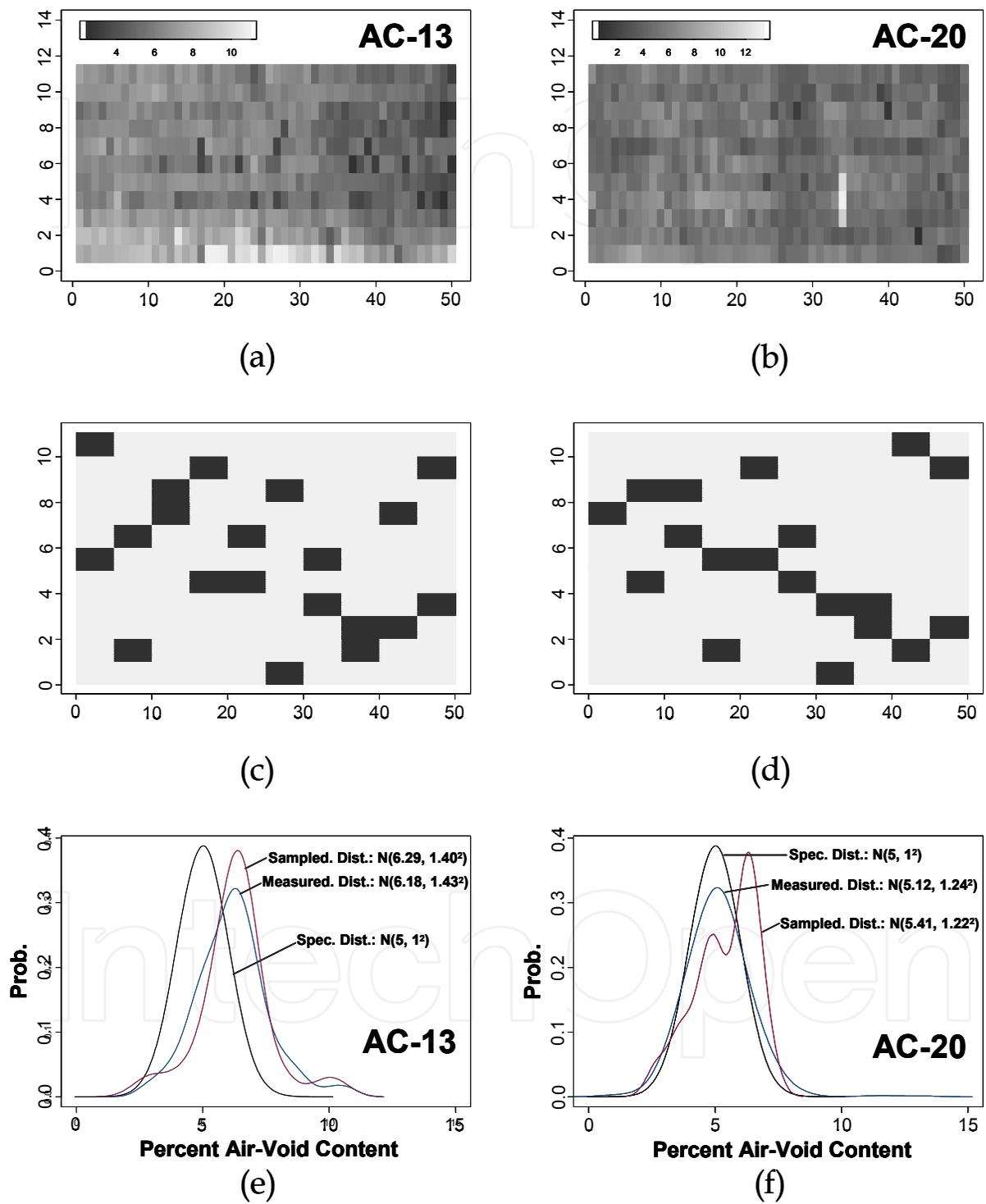


Fig. 13. Image plots of air-void measures for sections (a) AC-13 and (b) AC-20; UD tables  $x10y11n20$  for sections (c) AC-13 and (d) AC-20; the specification, measured, and sampled distributions for sections (e) AC-13 and (f) AC-20

made only if  $T_0 \geq -t_{\alpha, n-1}$  [shown in Figures 14a and 14b respectively for the situations that (a)  $\delta > 0$  and (b)  $\delta < 0$ ] where  $T_0$  is noncentral  $t_{n-1}$  distribution. Hence, we have the probability of Type II error

$$\beta = 1 - T_{n-1} \left( -t_{\alpha, n-1} - \frac{\delta\sqrt{n}}{\sigma} \right), \text{ i.e., power} = 1 - \beta = T_{n-1} \left( -t_{\alpha, n-1} - \frac{\delta\sqrt{n}}{\sigma} \right), \quad (6)$$

where  $T_{n-1}$  is the distribution function with  $n - 1$  degrees of freedom. From Figure 14, it is apparent that the more positive  $\delta$  value the larger the  $\beta$  value, i.e., the smaller the power; on the contrary, the more negative  $\delta$  value the smaller the  $\beta$  value, i.e., the larger the power.

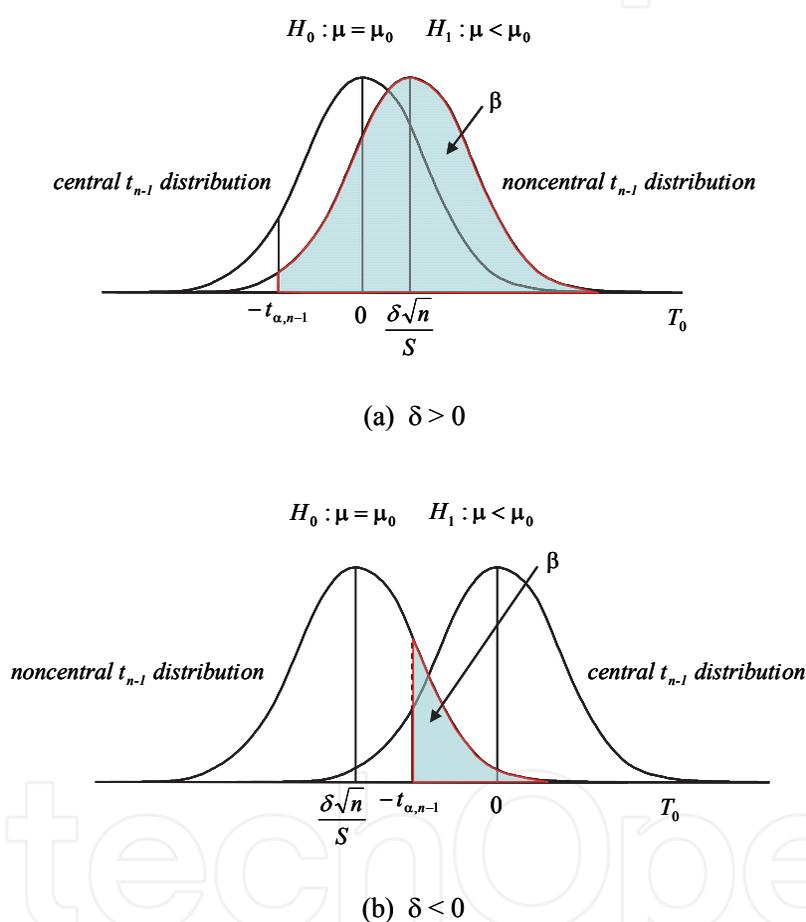


Fig. 14. Definition of type II error ( $\beta$ ) of a  $t$  distribution under the hypothesis testing:  $H_0 : \mu = \mu_0, H_1 : \mu < \mu_0$  for the situations that (a)  $\delta > 0$  and (b)  $\delta < 0$ .

Equation 6 indicates that power is a function of  $\alpha, n,$  and  $\delta/S$ . Figure 15 plots power versus  $\delta/S$  at various sample sizes. Under the hypothesis testing  $H_0 : \mu = \mu_0, H_1 : \mu < \mu_0, \alpha = 0.05$  and  $n = 3$ , the interpretation of Figure 15 is that one will have power greater than 0.8 to reject the null hypothesis if  $\delta/S \leq -2.30$ ; on the other hand, if  $\delta/S \geq -2.30$ , then the agency has insufficient power to reject the null hypothesis that  $H_0 : \mu = \mu_0$ . It should be noted that, to increase sample size from three to five, the agency will have power greater than 0.8 if  $\delta/S \leq -1.37$ ; that is, the agency can detect smaller mean difference from  $2.30S$  down to  $1.37S$  by

increasing two samples. In sum, by taking only three samples out of a project, the agency will have insufficient power to reject  $H_0 : \mu = \mu_0$  given that  $H_0$  is false unless the quality of the project delivered by the contractor is so poor that the agency is confident enough to reject the project.

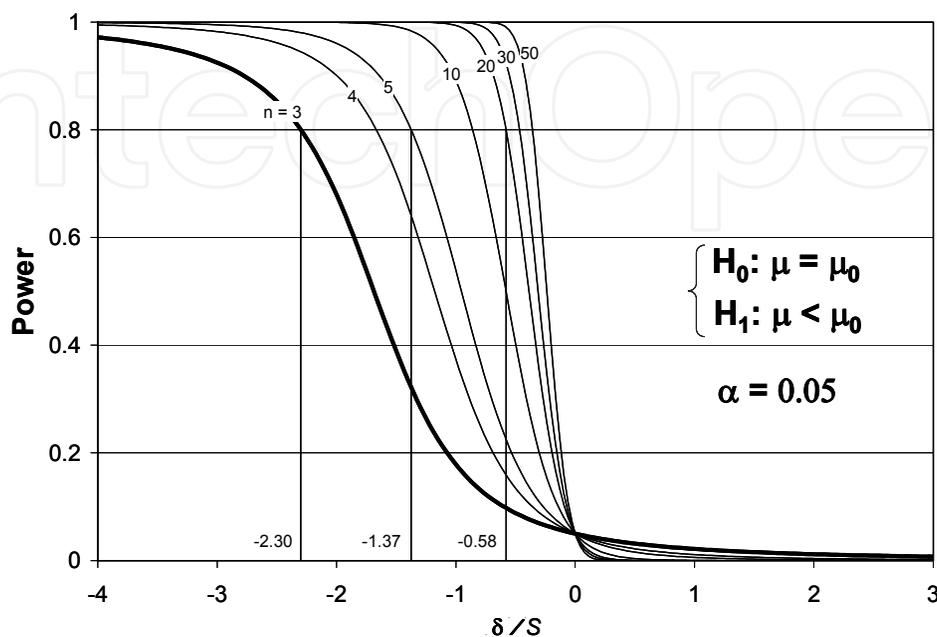


Fig. 15. Power versus  $\delta/S$  curves at different sample sizes for the one-sided  $t$ -test at a significance level  $\alpha = 0.05$ .

## 5. Findings and conclusions

For the Case I study, an attempt has been made to illustrate an approach and the extent of testing required using a performance test to insure reasonable quality in as-placed HMA. Stabilometer S-value test results were used in this example since extensive data were available. It should be emphasized that the same approach could be applied using other test parameters to control the quality of the as-constructed mix.

Based on stabilometer test results, the brief discussion of hypothesis testing, and the simulation results of sampling scheme and size, the following observations and suggestions are offered:

1. Cooperation between the agency and the contractor is essential. It is necessary to have the testing process, test equipment, test results, and specimen preparation as consistent as possible between the two organizations.
2. The sampling simulation of the Case I demonstration example suggests that the sample size required to stabilize the sampling consistency and sampling stabilization is around 50 ~ 70 for the placement of 15,000 tons HMA.
3. Likely, sampling as noted (2) is perhaps impractical. However, increasing the sample size is actually beneficial for both the agency and the contractor since it reduces the potential for dispute and guarantees the quality of the constructed mix. By extension, it is advisable for the agency to provide incentives to encourage the contractor to increase sampling size and testing.

4. To ensure the success of the proposed QC/QA guidelines, the contractor's minimum value of the testing null hypothesis must exceed that required by the agency.
5. From the Caltrans case study, the  $\mu_{\min}$  criterion depended not only on the contractor's  $\alpha$  value and the agency's power level as expected but also on the  $k$  value that the agency would select for use. The  $\mu_{\min}$  criterion can be smaller if both the agency and the contractor require low power level and high  $\alpha$  level and/or the agency increases the  $k$  value.

A concluding general observation relates to the concern for developing longer lasting pavement at this period of time because of increased costs of both pavement materials and increased traffic that must be accommodated. The added costs of testing by both the contractor and the agency are a very small proportion of the total costs associated with long lasting pavements. Accordingly an "attitude adjustment" for both parties relative to QC and QA testing would enhance long-term pavement performance.

From above discussion of Case II for determining sample size, simulation results of the sampling size and sampling scheme using UD tables, along with a demonstration example, the following observations and suggestions are offered:

1. It is important to recognize that the agency can be  $100(1-\alpha)\%$  confident that the error  $|\bar{x} - \mu|$  will not exceed a specified amount  $E$  if and only if the sample size is  $n = (z_{\alpha/2} \cdot \sigma / E)^2$ . The variations of sample mean and sample standard deviation for the 900 ft HMA paving simulation (Figures 8, 11, and 12) suggests that the minimum sample size required to stabilize the variation is around 20 ~30.
2. The UD table not only provides the most representative sampling scheme with the sample size for a given specified error level by the agency but also minimizes the possible effect of the underlying data pattern. Moreover, the UD table gives the agency a more unbiased "random" sampling scheme that can be followed in the quality assurance process.
3. The sample mean and sample standard deviation criteria proposed in the QA guideline demonstrates the accurate/inaccurate and precise/imprecise concept of sampling outcomes. If the sample mean is located in the range of  $100(1-\alpha)\%$  confidence interval, then it is accurate. Precision is a term to describe the degree of data dispersion; if the sample standard deviation is less than the  $100(1-\alpha)\%$  one-sided upper bound, then it is precise. The case study presents a very good example of an inaccurate/imprecise case of the AC-13 field section and an accurate/precise case of the AC-20 field section. The quality of a project can only be accepted if and only if these criteria have been fulfilled simultaneously.
4. The proposed QA guideline with the introduction of the UD table is relatively simple, practical, and robust. The sample mean and sample standard deviation criteria are rational enough for both the agency and the contractor to agree upon.
5. It should be emphasized that the proposed QA approach could be applied with other performance measurement parameters to control the quality of the as-constructed mix, such as thickness, stabilometer testing as used in California, performance testing of fatigue and rutting, etc. Moreover, the decision-making based on this proposed QA approach can also be a basis for pay factor determination.
6. By taking only three samples out of a project, the agency will have insufficient power to reject  $H_0 : \mu = \mu_0$  given that  $H_0$  is false unless the quality of the project delivered by the

contractor is so poor that the agency is confident enough to reject the project. However, by increasing sample size from three to five, the agency can detect smaller mean difference from  $2.30S$  down to  $1.37S$  by simply increasing two samples.

7. It is likely that the proposed sampling size is impractical. In this regard, the alternative is to establish a "testing section" similar to those in the case study and follow the proposed QA approach with the minimum sampling size (at least greater than 20) to ensure that the compaction pattern/effort, paving operation, and other construction details are appropriate to guarantee that the pavement quality meets the specifications.

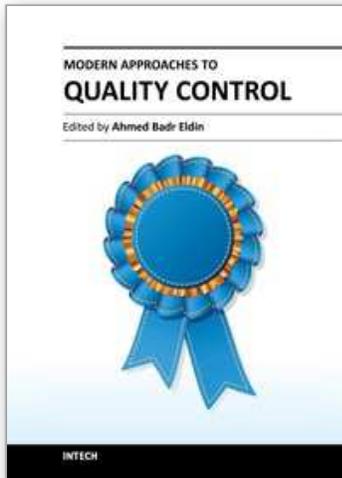
## 6. Acknowledgments

The research associated with the first case study was conducted as a part of the Partnered Pavement Research Program supported by the California Department of Transportation (Caltrans) Division of Research and Innovation. Special thanks go to Mr. Kee Foo of Caltrans who provided the stability data from Caltrans projects. The contents of this paper reflect the views of the authors who are responsible for the facts and accuracy of the information presented and do not reflect the official views of the policies of the State of California or the Federal Highway Administration.

The field data associated with the second case study was sponsored by the Ministry of Transport of the People's Republic of China. The contents of this paper reflect the views of the authors who are responsible for the facts and accuracy of the information presented and do not reflect the official views of the policies of the Ministry of Transport of the People's Republic of China.

## 7. References

- California Department of Transportation (Nov. 2007). *Standard Specifications*, Sacramento, Calif., USA
- Cleveland, W.S. (1993). *Visualizing Data*, Hobart Press, ISBN 978-0963488404, Summit, NJ, USA
- Fang, K.T. (1980). The Uniform Design: Application of Number Theoretical Methods in Experimental Design. *Acta Mathematicae Applagatae Sinica*, Vol.3, pp. 353-372, ISSN 1618-3832
- Fang, K.T.; Lin, D.K.J, Winker, P. & Zhang, Y. (2000). Uniform Design: Theory and Application. *Technometrics*, Vol.42, No.3, pp. 237-248, ISSN 1537-2723
- Fang, K.T. & Lin, D.K.J (2003). Uniform Experimental Designs and Their Applications in Industry. In: *Handbook of Statistics 22*, Khattree, R. & Rao, C.R., pp. 131-170, ISBN 0444-506144
- Montgomery, D.C. & Runger, G.C. (2010). *Applied Statistics and Probability for Engineers*, John Wiley & Sons, Inc., ISBN 978-0-470-05304-1, USA
- Stone, C.J.A. (1996). *Course in Probability and Statistics*, Duxbury Press, ISBN 0-534-23328-7, Pacific Grove, Calif., USA
- Tsai, B.-W. & Monismith, C.L. (2009). Quality Control - Quality Assurance Sampling Strategies for Hot-Mix Asphalt Construction. In *Transportation Research Board: Journal of the Transportation Research Board*, No.2098, pp. 51-62, ISSN 0361-1981
- Wang, Y. & Fang, K.T. (1981). A Note on Uniform Distribution and Experimental Design. *KeXue TongBao*, Vol.26, pp. 485-489, ISSN 0250-7862



## **Modern Approaches To Quality Control**

Edited by Dr. Ahmed Badr Eldin

ISBN 978-953-307-971-4

Hard cover, 538 pages

**Publisher** InTech

**Published online** 09, November, 2011

**Published in print edition** November, 2011

Rapid advance have been made in the last decade in the quality control procedures and techniques, most of the existing books try to cover specific techniques with all of their details. The aim of this book is to demonstrate quality control processes in a variety of areas, ranging from pharmaceutical and medical fields to construction engineering and data quality. A wide range of techniques and procedures have been covered.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Bor-Wen Tsai, Jiangmiao Yu and Carl L. Monismith (2011). Application of Sampling Strategies for Hot-Mix Asphalt Infrastructure: Quality Control-Quality Assurance Sampling; Specification for Performance Test Requirements, Modern Approaches To Quality Control, Dr. Ahmed Badr Eldin (Ed.), ISBN: 978-953-307-971-4, InTech, Available from: <http://www.intechopen.com/books/modern-approaches-to-quality-control/application-of-sampling-strategies-for-hot-mix-asphalt-infrastructure-quality-control-quality-assura>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen