ANALYSIS OF QA PROCEDURES AT THE OREGON DEPARTMENT OF TRANSPORTATION

Final Report

SPR 650

ANALYSIS OF QA PROCEDURES AT THE OREGON DEPARTMENT OF TRANSPORTATION

Final Report

SPR 650

by

David N. Sillars, PhD, PE
Todd Scholz, PhD, PE
Matt Hallowell, PhD
Oregon State University

for

Oregon Department of Transportation
Research Section
200 Hawthorne Ave. SE, Suite B-240
Salem OR 97301-5192

and

Federal Highway Administration
400 Seventh Street, SW
Washington, DC 20590-0003

June 2010

Technical Report Documentation Page

1. Report No. FHWA-OR-RD-10-21	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ANALYSIS OF QA PROCEDURES AT THE OREGON DEPARTMENT OF TRANSPORTATION		5. Report Date June 2010 6. Performing Organization Code
7. Author(s) David N. Sillars, Todd Scholz, Matt Hallowell		8. Performing Organization Report No.
9. Performing Organization Name and Address Oregon State University Department of Civil, Construction and Environmental Engineering 220 Owen Hall Corvallis, Oregon 97331		10. Work Unit No. (TRAIS) 11. Contract or Grant No. SPR 650
12. Sponsoring Agency Name and Address Oregon Department of Transportation Research Section 200 Hawthorne Ave. SE, Suite B-240 Salem, OR 97301-5192	and Federal Highway Administration 400 Seventh Street, SW Washington, DC 20590-0003	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This research explored the Oregon Department of Transportation (ODOT) practice of Independent Assurance (IA), for validation of the contractor's test methods, and Verification, for validation of the contractor's Quality Control (QC) data. The intent of the project was to discover whether adjusted or additional processes for comparison between ODOT's test results and the contractor's test results may be available to improve confidence regarding whether the contractor's test results are reliable. It was found that ODOT utilizes a combination IA/Verification process that uses the comparison of single results for its IA and Verification decisions. Based on statistical principles, published literature, FHWA guidance, and a small case study, recommendations were made that ODOT establish the breadth of systematic testing bias and that the IA and Verification processes be enhanced to include statistically-based comparison tests, including the often-used t-test and F-test.

17. Key Words QA/QC		18. Distribution Statement Copies available from NTIS, and online at http://www.oregon.gov/ODOT/TD/TP_RES/	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 86 + appendices	22. Price

Technical Report Form DOT F 1700.7 (8-72) Reproduction of completed page authorized ♻ Printed on recycled paper

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS					APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	When You Know	Multiply By	To Find	Symbol	Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH					**LENGTH**		
in	inches	25.4	millimeters	mm	mm	millimeters	0.039	inches	in
ft	feet	0.305	meters	m	m	meters	3.28	feet	ft
yd	yards	0.914	meters	m	m	meters	1.09	yards	yd
mi	miles	1.61	kilometers	km	km	kilometers	0.621	miles	mi
		AREA					**AREA**		
in^2	square inches	645.2	millimeters squared	mm^2	mm^2	millimeters squared	0.0016	square inches	in^2
ft^2	square feet	0.093	meters squared	m^2	m^2	meters squared	10.764	square feet	ft^2
yd^2	square yards	0.836	meters squared	m^2	m^2	meters squared	1.196	square yards	yd^2
ac	acres	0.405	hectares	ha	ha	hectares	2.47	acres	ac
mi^2	square miles	2.59	kilometers squared	km^2	km^2	kilometers squared	0.386	square miles	mi^2
		VOLUME					**VOLUME**		
fl oz	fluid ounces	29.57	milliliters	ml	ml	milliliters	0.034	fluid ounces	fl oz
gal	gallons	3.785	liters	L	L	liters	0.264	gallons	gal
ft^3	cubic feet	0.028	meters cubed	m^3	m^3	meters cubed	35.315	cubic feet	ft^3
yd^3	cubic yards	0.765	meters cubed	m^3	m^3	meters cubed	1.308	cubic yards	yd^3
NOTE: Volumes greater than 1000 L shall be shown in m^3.									
		MASS					**MASS**		
oz	ounces	28.35	grams	g	g	grams	0.035	ounces	oz
lb	pounds	0.454	kilograms	kg	kg	kilograms	2.205	pounds	lb
T	short tons (2000 lb)	0.907	megagrams	Mg	Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)					**TEMPERATURE (exact)**		
°F	Fahrenheit	(F-32)/1.8	Celsius	°C	°C	Celsius	1.8C+32	Fahrenheit	°F

*SI is the symbol for the International System of Measurement

ACKNOWLEDGEMENTS

The author would like to thank the members of the ODOT Research Unit, and Technical Advisory Committee:

Norris Shippen	Research Coordinator	ODOT Research Unit
Vincent Van Der Hyde	Senior Research Analyst	ODOT
Cole Mullis	Quality Assurance Engineer	ODOT
John Svadlenak	Transportation Economist	ODOT
Mark Winn	QCCS	ODOT
Anthony Boesen		FHWA

Especially helpful in this process was Cole Mullis, whose professional interest in establishing a viable organizational process to improve quality assurance has provided insight into many issues explored in this study.

DISCLAIMER

ANALYSIS OF QA PROCEDURES AT THE OREGON DEPARTMENT OF TRANSPORTATION

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 PURPOSE OF RESEARCH 2
1.2 METHODOLOGY 2

2.0 LITERATURE REVIEW 3

2.1 OVERVIEW 3
2.2 QA/QC CONCEPTS..3
  2.2.1 Background...3
  2.2.2 Historical perspective 5
  2.2.3 Overview of QA/QC..5
  2.2.4 Specification types..6
2.3 QA/QC USE IN U.S. TRANSPORTATION AGENCIES..6
  2.3.1 Quality Control..6
    2.3.1.1 State DOT procedures for Quality Control 8
    2.3.1.2 Preferred QC process 8
  2.3.2 Acceptance and Independent Assurance 9
    2.3.2.1 Introduction...9
    2.3.2.2 State DOT procedures..10
  2.3.3 D2S limits testing 11
  2.3.4 Statistically-based comparison of means..11
    2.3.4.1 Sampling 12
    2.3.4.2 Randomization 12
    2.3.4.3 Independence 13
  2.3.5 Overview of statistical procedures 14
    2.3.5.1 Introduction and assumptions..14
    2.3.5.2 Split sample comparisons 15
    2.3.5.3 Independent sample comparisons..18

3.0 QA PROCEDURES AT ODOT...29

3.1 INTRODUCTION 29
  3.1.1 General discussion..29
  3.1.2 ODOT quality assurance program...30
    3.1.2.1 Random sample procedure 32
    3.1.2.2 Sampling parameters and procedure..32
3.2 ODOT MATERIAL TEST PROCEDURES AND GUIDELINES....................................34
  3.2.1 Hot Mixed Asphalt Concrete (HMAC) 34
    3.2.1.1 Quality Control..34
    3.2.1.2 Independent Assurance (IA) and Verification.....................................34
  3.2.2 Aggregates – base, subbase, and shoulders 36

3.2.2.1 Quality Control 36
3.2.2.2 Independent Assurance (IA) and Verification 36
3.2.3 Portland Cement Concrete (PCC): 37
3.2.3.1 Quality Control 37
3.2.3.2 Independent Assurance (IA) and Verification 37
3.2.4 Earthwork 38
3.2.4.1 Quality Control 38
3.2.4.2 Independent Assurance (IA) and Verification 38
3.2.5 Very small project quantity guidelines 39
3.2.6 Assessment of ODOT Procedures for Independent Assurance and Verification 40
3.2.7 Sampling quantities 41
3.3 PROCESS EVALUATION 42
3.3.1 NCHRP Study Results 42
3.3.2 ODOT Case Studies 43
3.3.2.1 Oregon 18: Oregon Coast Highway to Oldsville Road Section 43
3.3.2.2 Oregon 18: Fort Hill to Wallace Bridge Section 44
3.3.2.3 Case study evaluations 44
3.3.2.4 IA evaluation: Paired t-test 49
3.3.2.5 Verification evaluation: F-test and t-test 52
3.3.2.6 Case study assessment 54
3.4 SUMMARY 54

4.0 PROCESS RECOMMENDATIONS 57

4.1 SAMPLING 57
4.2 INDEPENDENT ASSURANCE 59
4.3 VERIFICATION 60
4.3.1 Statistical verification procedures 60
4.3.2 Establishing minimum sample quantities for statistical evaluation 62
4.3.3 Sampling frequency 64
4.4 VERIFICATION TESTING SPECIFICATIONS 68
4.5 ELECTRONIC METHODS OF QA DATA ANALYSIS 68
4.5.1 Overview 68
4.5.2 Electronic QA analysis at ODOT 69
4.6 MATERIALS TO CONSIDER FOR IMPROVED VERIFICATION PROCEDURES 69
4.7 IMPLEMENTATION STEPS 70
4.8 SUMMARY 70

5.0 CONCLUSION 73

6.0 REFERENCES 75

APPENDICES

APPENDIX A: SAMPLE CONTROL CHART
APPENDIX B: FEDERAL REGISTER QUALITY ASSURANCE PROCEDURES FOR CONSTRUCTION
APPENDIX C: CRITICAL VALUES FOR THE *T*-TEST
APPENDIX D: CRITICAL VALUES FOR *F*-TEST

APPENDIX E: CASE STUDY DATA
APPENDIX F: CALIFORNIA VERIFICATION FLOWCHART
APPENDIX G: SOUTH CAROLINA METHOD FOR VERIFICATION OF CONTRACTOR HMA ACCEPTANCE TEST RESULTS
APPENDIX H: COLORADO F- AND T-TEST STATISTICAL METHOD FOR HMA VOIDS ACCEPTANCE
APPENDIX I: SURVEY OF QA/QC SOFTWARE

List of Figures

Figure 2.1: Typical arrangement of QA/QC programs 4
Figure 3.1: ODOT representation of QA functions 30
Figure 3.2: ODOT QA/QC process 31
Figure 3.3: Case study key constituents control charts 46
Figure 3.4: Split sample results variances 50
Figure 4.1: Recommended ODOT QA/QC process 58

List of Tables

Table 2.1: Comparison of Independent Assurance and Quality Control 6
Table 2.2: D2S limits for Independent Assurance evaluation (Burati, et al.) 11
Table 2.3: Example split sample data 18
Table 2.4: Example asphalt content tests 21
Table 2.5: Rank transformation 26
Table 2.6: Rank transformation of Table 2.4 27
Table 3.1: Independent Assurance tolerances 33
Table 3.2: Testing frequencies and requirements for HMAC 35
Table 3.3: Testing frequencies and requirements for aggregate 37
Table 3.4: Testing frequencies and requirements for PCC (structural) 38
Table 3.5: Testing frequencies and requirements for embankment 39
Table 3.6: Very small project quantity table 39
Table 3.7: Project size parameters 41
Table 3.8: Testing frequencies (per MFTP guidelines, Section 4(D)) 41
Table 3.9: Minimum number of QA samples, based on project size 42
Table 3.10: Case study constituent pay factor weights 45
Table 3.11: Coast Hwy/Oldsville IA and Verification results 48
Table 3.12: Ft Hill/Wallace Br IA and Verification results 49
Table 3.13: Case study IA paired t-test results 51
Table 3.14: Case study Verification t-test results 53
Table 4.1: AASHTO R-9 α recommendations 62
Table 4.2: Sample size versus likelihood of detecting a three-fold difference in standard deviations (F-test) 63
Table 4.3: Sample size versus likelihood of detecting a two-standard deviation difference in means (t-test) 64
Table 4.4: Current scheme: 10% Verification sampling--likelihood of detecting a two-standard deviation difference in means 65
Table 4.5: Scheme A: Minimum of 3 Verification samples, except for very small projects 66
Table 4.6: Scheme A: Likelihood of detecting a two-standard deviation difference in means 66
Table 4.7: Scheme B: Minimum of 6 Verification samples, except for very small projects 66
Table 4.8: Scheme B: Likelihood of detecting a two standard deviation difference in means 67
Table 4.9: Material category compliance characteristics 70

1.0 INTRODUCTION

In July 2005 the Federal Highway Administration (FHWA) conducted a stewardship review of the Oregon Department of Transportation's (ODOT's) Quality Assurance Program. The review proposed that ODOT's process of validating the contractor's test results was not in compliance with the Code of Federal Regulations (CFR). The issue noted in the report was that ODOT needed to incorporate a statistically-based analysis method to analyze the contractor's ongoing test results against the agency's results. FHWA suggested using the F-test and t-test for validating contractor test results.

There are several procedures used in the U.S. for both test method assurance and process testing verification, depending on the agency's goals in the Verification process and whether independent or split samples are used. ODOT's current procedures use split samples in the independent assurance process and couples those results with process verification. The first evaluation is to compare between the contractor and agency results to determine if they are within specified tolerances. If they are within tolerance, the difference is considered acceptable. If they are not within tolerance, an investigation is conducted to determine the source of the difference. This portion of the process is test method Quality Assurance (QA), i.e., the differences in results are attributable to differences in test methods and equipment. The second part of the process is to compare the test results to the specifications for the material in question. If the contractor and agency test results meet specification and the previous check for test method assurance is acceptable, then ODOT infers that the contractor's separate Quality Control (QC) test results are verified (to the extent that the QC results are also within specification); and those QC results are used to determine project payment. If either of the results does not meet specification, an investigation is completed and the project manager must decide the appropriate course of action.

This research project examines other systems of Quality Assurance (QA), compares those systems against ODOT's current QA system. It then provides recommendations to add to ODOT's current QA procedures, the purpose of which are to gain further insight into potential systematic differences between the contractor's and ODOT's test results. These recommendations are intended as a supplement, not a replacement, to ODOT's current methods. Further, the recommendations are intended to capture data in a manner that provides insight into what factors may be causal in any contractor/ODOT differences found. Implementing the recommendations of this report will give ODOT a better understanding regarding the quality of materials being incorporated into its construction projects and improve its ability to make decisions related to the acceptance of quality of materials and workmanship. These efforts will also assist the department in responding to FHWA's stewardship review.

1.1 PURPOSE OF RESEARCH

The objectives of this project were to:

- Evaluate ODOT's current system and suggest a statistically valid procedure to analyze and compare the following:
  - Split sample results for Independent Assurance (IA) (test method verification);
  - Sample results for validation of contractor Quality Control (QC) test results (process verification);
- Evaluate for which products / test procedures to apply the statistical process;
- Suggest an electronic process for conducting the statistical analysis; and
- Develop guidelines for conducting the analysis and interpreting and using the results.

1.2 METHODOLOGY

The organization of this report reflects the following research methodology:

A Literature Review and Background Investigation is included in Chapter 2, which was conducted to gain an understanding of the acceptable statistical analysis procedures, their requirements, and their limitations;

Documentation and Assessment of the ODOT process is described in Chapter 3; included are notes regarding comparison against processes discovered in the literature review;

Process Recommendations are provided in Chapter 4 for the following three processes to suggest the best statistical approach and to suggest which products / test procedures are best suited for enhanced QA analysis:

- The contractor's ongoing QC test results,
- Split sample results for test method verification, and
- Sample results for validation of contractor's QC test results (process verification).

Electronic Processes are reviewed and recommendations provided that can be used by ODOT to establish an institutionalized process for conducting the analytical procedures recommended by this project.

2.0 LITERATURE REVIEW

2.1 OVERVIEW

The literature review of this section is intended to provide a discussion of existing methodologies and techniques utilized in highway construction QA/QC procedures. This literature review chapter covers the investigation of two topics:

- QA/QC concepts; and
- QA/QC procedures related to U.S. highway construction.

The purpose of this document is to provide a background on Quality Assurance and Quality Control (QA/QC), an overview of the key components and techniques of a well-structured program, and the preferred methods identified by literature. This project reviewed literature such as publications from state departments of transportation (DOTs) and the Federal Highway Administration (FHWA), journal publications, and relevant books. It was found that significant prior study has occurred, and it has been well-documented and summarized in National Cooperative Highway Research Program (NCHRP) and FHWA publications.

This literature review utilizes definitions and terms that are frequently used in the national publications. Acknowledged within these publications is that there are variants of these terms and definitions across the many states. Although the fundamental concepts are the same, ODOT has slightly different terminology and definitions than some of those presented in this chapter; those terms and definitions are presented in Chapter 3 – QA Procedures at ODOT. The reader is asked to primarily consider the concepts involved and focus on the principles that provide confidence in the quality assurance process.

2.2 QA/QC CONCEPTS

2.2.1 Background

Quality Assurance and Quality Control, known as QA/QC, is a compound term that refers to the tasks associated with ensuring that a company or agency provides an acceptable product according to the specifications (Dixon 2003). To ensure that the public receives a high-quality product that is delivered in the most efficient, economical, and satisfactory manner possible, state DOTs must have a well-structured and advanced method of ensuring quality in all aspects of the design and construction processes. Quality Assurance and Quality Control (QA/QC) are inter-related processes that are used to ensure that the quality of the product is equal to or greater than specified in the contract documents.

Quality Assurance (QA) is defined as the planned and systematic actions necessary to provide confidence that a product or facility will perform satisfactorily in service, according to the Transportation Research Board (TRB 2005). The three elements of Quality Assurance, as described in the national literature, are illustrated in Figure 2.1.

Quality Control (QC) involves the actions and considerations necessary to assess and adjust production and construction processes so as to control the level of quality being produced in the end product. Typically, Quality Control is the responsibility of the producer of the product (e.g., the contractor), whereas Quality Assurance is typically the responsibility of the agency (e.g., ODOT).

Acceptance is used to evaluate the acceptability of material or construction by taking samples and making measurements or observations of the samples. Acceptance also refers to the testing of samples to ensure that the specified performance characteristics have been met. A product or process is accepted only if the samples meet or exceed the specified performance or quality standards. Acceptance is typically performed by the agency. In its most basic definition, Quality Assurance is a combination of QC and Acceptance, where QC is used for the control of the process and Acceptance is used to assess the quality of the product. In some cases, Independent Assurance is added to the Quality Assurance process.

Independent Assurance is a management tool that requires an objective party to provide an independent assessment of the product and/or the reliability of the tests obtained from the Quality Control and Acceptance testing. The party performing Independent Assurance should ideally not be responsible for Quality Control or Acceptance, and Independent Assurance should not be used to justify product acceptance (Dixon 2003). The definitions provided above are consistent with the Transportation Research Board's (TRB's) glossary of QA/QC terms for highway construction (TRB 2005).

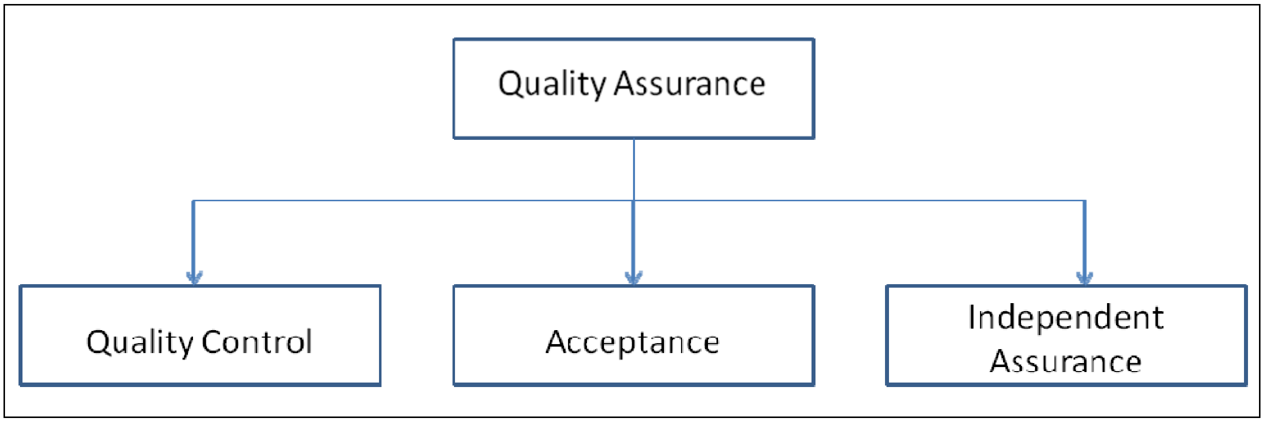


Figure 2.1: Typical arrangement of QA/QC programs

Quality Assurance (QA) is ultimately performed in an effort to decide whether to accept or reject a product, including in some cases what percentage reduction in payment to apply to work that is of reduced quality (called a "pay factor"). When an independent agency does not directly perform QA testing, then Independent Assurance must be instituted, including statistical tests to establish whether a significant difference exists between the agency's QA tests and the contractor's QC tests. In such an arrangement the contractor must first perform QC by collecting and testing samples to measure its conformance to the specifications (AASHTO 1984). Secondly, the DOT validates the contractor's testing process by obtaining samples and conducting separate testing using either the DOT's or a consultant's labs. Lastly, the DOT must validate the contractor's QC results by independent sampling, confirming that there is no statistically significant difference between the DOT's and the contractor's test results (Burati, et al. 2004). This latter process is called Verification.

According to Caltrans, there is a significant amount of misinformation and confusion regarding QA/QC procedures in state DOTs (Caltrans 2002). This confusion has resulted in a lack of uniformity among state DOTs. In fact, the procedures used by some state DOTs have been described as ineffective and impractical. Because of the general level of confusion in the industry, this literature review will cover the basics of a QA/QC program, indicating the processes that are highly recommended.

2.2.2 Historical perspective

Two FHWA publications provide insight to the origin and necessity of a well-formed QA/QC program (Burati, et al. 2003; Burati, et al. 2004). According to these documents, approximately 90 percent of all State Highway Agencies (i.e., DOTs) and most federal agencies have an established QA program. These programs are by no means identical; however, these state and federal programs share common characteristics and definitions. The evolution to institutionally-defined QA took place over the course of several decades beginning in 1956. It grew out of the American Association of State Highway Officials' (AASHTO) Road Test and the many analyses that emanated from that study (Weed 1986). The high variability found in this study led to the creation of the first QA and Acceptance plans. Prior to this study, specifications were mainly materials and methods specifications known as "prescription" or "recipe" standards that did not reference specific quality requirements (Burati, et al. 2003). The first QA specifications focused mainly on what is now the modern-day quality control. These specifications evolved into true QA when the need to separate Quality Control from Acceptance was recognized.

A detailed description of the history of QA/QC in the transportation construction industry can be found in the *National Cooperative Highway Research Program (NCHRP) Synthesis 38* "Statistically-Oriented End-Result Specifications" (NCHRP 1976).

2.2.3 Overview of QA/QC

As previously indicated, QA and QC are used jointly, but they are distinct concepts. While originally considered to be one concept, QA has evolved into a separation between Independent Assurance (IA) and Acceptance, and Quality Control. As evidenced by the previous discussion, it becomes clear that QC is merely one component of a comprehensive QA program. According to Dixon (2003), QC is also called process control and refers to the actions and considerations necessary to assess construction materials and processes. In other words, QC is applied to ensure the adequate quality of the final product. The differences between QA and IA and Accpetance are illustrated in Table 2.1. This table has been adapted from Dixon (2003) and Burati, et al. (2004).

Table 2.1: Comparison of Independent Assurance and Quality Control

	Quality Control	Independent Assurance and Acceptance
Definition	Making the quality of a product what it should be	Checking the that the QC test results are reliable
Responsibility	Producer/Contractor	Agency
Interrelationship	Validates Quality Assurance	Validates Quality Control results

As QA and QC have become separated, so also have Acceptance and Independent Assurance (IA) become typically considered as separate functions. IA is a management tool that requires an objective party that is not directly responsible for the product or process to ensure that the testing process is reliable (AASHTO 1996). With Acceptance, a product or process is accepted and payment made only if the condition exists where a product or process meets the specified performance and quality requirements. Included within the Acceptance function is Verification, which is the process of determining the accuracy of QC tests results by examining the data (TRB 2005).

2.2.4 Specification types

QA/QC efforts are necessary to ensure that the end product meets the specifications in the contract documents. In the construction of highway projects there are three basic types of specifications: method specifications, end result specifications, and quality assurance specifications (Dixon 2003). Methods specifications (i.e., materials and methods specifications) require the contractor to follow step-by-step procedures using specified materials and methods to place a given material. End result specifications require the contractor to take entire responsibility for supplying the product and its construction; these specifications are also called performance specifications, wherein the means and methods of achieving the end result are irrelevant (Benson 1995). Finally, quality assurance specifications are a combination of method specifications and end result specifications. With quality assurance specifications, the contractor is responsible for the Quality Control, and the agency (e.g., ODOT) is responsible for Acceptance. Because the dominant specification type in highway construction is the quality assurance specification, and because this is the dominant specification type used in Oregon, this specification type will be assumed in this report.

2.3 QA/QC USE IN U.S. TRANSPORTATION AGENCIES

2.3.1 Quality Control

As previously defined, Quality Control (QC) involves the actions and considerations needed to assess production and construction processes so as to control the level of quality of the end product. QC has two distinct parts: the QC processes and the quality characteristics to be measured. Typically, Quality Control is the responsibility of the contractor, and this process should be able to quickly identify nonconforming materials during construction (Burati, et al. 2003).

Generally, the agency requires the contractor to have its own QC plan, because it is vital for the contractor to understand the activities required to produce, test, and inspect acceptable material. There are two main approaches that agencies use to specify the QC requirements for a project. The first option is to stipulate the minimum QC requirements and properties that the QC program must contain as a part of the project specifications. The other is for the agency to stipulate the requirements and properties that must be tested (Burati, et al. 2004). The latter is known as a performance standard, while the former is generally referred to as a prescriptive standard. Regardless of the method used to specify quality requirements, the program must be set up in such a way that materials are sampled, tested, and compared in an objective, rigorous, and defendable fashion.

In order for a contractor to institute an acceptable QC plan, the contractor may need to incorporate many elements (Burati, et al. 2004). Such elements include the following:

- QC plan acceptable for submission to an agency;
- Employment of qualified technicians;
- Use of a qualified laboratory;
- Understanding and statement of the properties and materials to be tested;
- Maintenance of control charts and a statement of the properties that will be plotted and how often;
- Statement of the action criteria that will be used to identify “out of control” production;
- Listing of the procedures to follow when an “out of control” product is identified; and
- Identification and recording of the parties responsible for correcting an “out of control” product.

Sampling and evaluating the data is perhaps the most important aspect of the Quality Control process. When the agency stipulates the QC plan requirements, as is the norm in the U.S. transportation construction industry, the agency plays virtually no role in the QC process. In other words, the contractor is responsible for all activities associated with sampling and analyzing the samples. The project specifications only stipulate the quality requirements that must be satisfied and the frequency and quantity of samples.

When establishing the action limits for control, available data must be analyzed to determine the ability to meet a specified target. This is achieved through the use of control charts that use historical project data to determine if a given sample is within acceptable tolerances and how such compliance varies over time. If a sample is outside a specified range, the sample is considered to be “out of control” and action must be taken as specified in the contract documents. Importantly, the QC samples must be collected and analyzed for the specific quality characteristic in the same manner and under the same general conditions that have been used in collecting the prior data. Sample consistency is of the utmost importance when applying control charts (Burati, et al. 2004).

2.3.1.1 State DOT procedures for Quality Control

Burati, et al. (2004) reviewed the QA/QC procedures implemented by the 50 state DOTs. In this study they found that 11 states require the use of control charts on all projects as a part of their Acceptance/QA process. The eleven states are as follows: California, Florida, Idaho, Illinois, Maryland, Minnesota, New York, Texas, and Wisconsin. Each of these states follows the preferred process outlined in the following subsection of this report. The remaining 39 states all use control charts as part of their QC process to varying degrees, including ODOT.

Many contractors that work with State DOTs are reluctant to adopt statistical process control (i.e., control charts). NCHRP found that contractors are reluctant to maintain control charts because:

> They conduct only the minimal tests required;
> The results on the charts are often plotted at a convenient time rather than immediately, obviating the ability to react to an out-of-control product;
> They use simplistic and less effective types of control charts called, “run charts”;
> They do not establish effective control limits;
> They use specification limits for control limits (this is sometimes an agency requirement);
> They do not react when a product is found to be out-of-control; and
> They use agency Acceptance test results for their QC. (Hughes 2005)

2.3.1.2 Preferred QC process

As indicated previously, the preferred process of quality control involves the use of control charts and, hence, historical data. An example of a generic control chart is provided in Appendix A. There are many sources that define the appropriate procedures for creating and using a control chart. The following sources provide specific guidance for control charts: Dixon (2003); Gentry and Yrjanson (1987); Gibria (1975); Burati, et al. (2004); and Weed (1984).

The control limits of a control chart are based on the variability of the specific process or material (Gibria 1975). As a general rule, if the contractor’s work processes produce variability consistently outside the control limits created from the contract specifications, then it is obvious that the contractor cannot consistently meet the specification requirements (Burati, et al. 2004). When the contractor can consistently meet the control limits based on the specification requirements, the ongoing QC data can be used to create a control chart that helps to identify rogue (i.e., “out of control” samples) that may indicate a threat to overall quality, and that require immediate action to ensure that the entire process has not moved out of control. It is of the utmost importance for overall quality that the variability of the contractor's processes be within acceptable limits prescribed in the specifications. Since this information is critical, the data used for the creation and population of a control chart must be unbiased.

According to Burati, et al. (2004), historical control chart data may frequently be biased. To ensure that valid project data is used, the data must have been obtained through randomization. Significant factors regarding randomization include the importance of non-systematic data collection; various techniques used to achieve randomization are covered in detail in statistics text books such as Ramsey (2002).

2.3.2 Acceptance and Independent Assurance

As previously discussed, the term Quality Assurance (QA) involves Quality Control (QC), Acceptance and Independent Assurance (IA). This section discusses Acceptance and IA procedures and requirements as published by the Federal Highway Administration. Two central documents (both part of the same, extended study)— Optimal Procedures for Quality Assurance Specifications and Evaluation of Procedures for Quality Assurance Specifications—provide state DOTs with information to conduct a statistically-valid, rigorous, and defendable comparison of contractor samples and agency samples (Burati, et al. 2003; Burati, et al. 2004). As discussed earlier, the adoption of these FHWA-suggested techniques varies among the state DOTs.

2.3.2.1 Introduction

As with QC procedures, there is no single prescriptive method of Acceptance and Independent Assurance that applies to all situations. However, there are several procedures that have proven to be effective in the Acceptance process. A brief overview of the various Acceptance procedures, as discussed in Burati, Weed, et al. (Burati, et al. 2004), has been included below:

> "If the primary function is to ensure that the contractors do not totally disregard quality, then the presence of an agency inspector accompanied by a minimal amount of Acceptance testing may be sufficient. The limited effort, however, will not really allow the agency to distinguish between good and poor construction material. To do this will require additional random sampling and testing along the lines of what has traditionally been done or greater.
>
> "If the agency wants a sound, statistically-based plan that will enable them to determine with a low degree of risk the quality levels that the contractor is providing, then even larger sample sizes will be required.
>
> "If the agency wants to provide sufficient information to use as the input into some of the elaborate performance models that are now under development, they will require considerably more sampling and testing than have traditionally been done by agencies. In this age of competition for limited resources, it seems unlikely that many agencies will be willing to commit to this level of sampling and testing on all of their projects. This level of testing may be limited to selected projects that might be used to help to develop agency-specific performance

> models. However, the calibration and updating of such models would still require an ongoing level of testing that agencies have been unwilling to maintain in the past. It seems unlikely that they will be willing and able to do so in the future."

Most commonly, one of two entities performs Acceptance testing. In most state DOTs, Acceptance testing is performed by the Agency or an Agency's subcontractor. In other arrangements, the testing is the responsibility of an independent third party. This arrangement is referred to as Independent Assurance (IA). In a preferred scenario, the Agency performs the Acceptance in-house. However, if the Agency finds, through an internal assessment, that there is a lack of qualified personnel, then the responsibility of assurance may be transferred to a third party (CFR 1995). For the purpose of this report, the remaining discussion will focus on Agency-performed Acceptance and Independent Assurance. Independent Assurance, as defined in the ODOT lexicon, is considered to be a function of the agency itself.

When performing Acceptance and Independent Assurance (IA) testing, two types of samples may be taken—split samples and independent samples. The definitions from the TRB QA/QC Glossary are provided below:

> **Split sample** – A sample that has been divided into two or more portions representing the same material. Split samples are sometimes taken to verify the acceptability of an operator's test equipment and procedure. This is possible because the variability calculated from differences in split test results is comprised solely of testing variability.
>
> **Independent sample** – A sample taken without regard to any other sample that may also have been taken to represent the material in question. An independent sample is sometimes taken to verify an Acceptance decision. This is possible because the data sets from independent samples, unlike those from split samples, each contain independent information reflecting all sources of variability, i.e., materials, sampling, and testing.

The purpose of Acceptance and Independent Assurance testing is to provide confidence that a product or facility will perform satisfactorily in service and to confirm quality in the case of variable payment provisions. In order to achieve this objective, the Agency must confirm that the materials and processes implemented in construction are adequate to achieve an end product of equal or greater quality than what is specified in the contract documents. QC focuses on process control, whereas Acceptance and Independent Assurance are used to ensure that the contractor's QC test values are true (i.e., the values obtained through the contractor's testing are representative of the actual material properties in place).

2.3.2.2 State DOT procedures

When using contractor test results in the Acceptance decision, 23 CFR 637 (see Appendix B) requires that Verification testing (a part of the Acceptance function) be done by the agency (Hughes 2005). The type of verification currently being used varies greatly from state to state and region to region. The use of consultants is widespread as a means to comply with 23 CFR 637 while coping with personnel reductions. More than 75% of the state DOTs in the NCHRP study (Hughes 2005) stated that they use consultants for Verification testing.

There are two main procedures used among state DOTs to compare the tests performed by the agency to those performed by the contractor. The first procedure, introduced and briefly described below, is referred to as D2S Limits; it involves comparing the mean value of the contractor's tests with an agency test. This is the most basic comparison and is no longer recommended by the FHWA. The second procedure is utilization of statistically-based comparison tools, using probability-based tools such as the F-test and t-test, also explained further below.

2.3.3 D2S limits testing

Burati, Weed, et al. (Burati, et al. 2003) provide an overview of the D2S procedures used to compare a single agency test to a number of contractor tests. In this method, a single agency test is typically compared to between 5 and 10 contractor tests. The maximum allowable difference between the average of the contractor's test and the agency test must fall within the allowable D2S intervals similar to those defined in Table 2.2. The value X in this table refers to the mean, and R refers to the range of values of the contractor tests. The Oregon Department of Transportation typically implements a similar method for independent assurance, evaluating whether the difference between single contractor and ODOT tests from a split sample fall within the limits noted in Table 3.1, located in the next chapter. Burati, et al. (2003) recognize that this is not an effective approach for Verification; therefore this method is not recommended.

Table 2.2: D2S limits for Independent Assurance evaluation (Burati, et al.)

Number of Contractor Tests	Allowable Interval
10	X ± 0.91 R
9	X ± 0.97 R
8	X ± 1.05 R
7	X ± 1.17 R
6	X ± 1.33 R
5	X ± 1.61 R

2.3.4 Statistically-based comparison of means

This section of the report discusses the suggested methods of data sampling to ensure that: (1) the data is unbiased and representative, (2) the t-tools and F-test can be used to effectively analyze the data, and (3) that the appropriate scope of inference of the conclusions drawn from the tests may be achieved.

The more complex, most powerful, and generally preferable method of comparison between contractor and agency test results is the combined use of the F-test and the t-test. These statistical tests are simple and easy to perform, yet are justifiable and mathematically-sound. In other words, the F-test and t-test can be used to make statistically-based inferences and can be used to make objective and defendable decisions. Detailed descriptions, definitions, and examples of these tests will be covered in the subsequent section of this report.

For Independent Assurance of split samples, the Paired t-test is the desired method (Burati, et al. 2003). This test uses the differences between pairs of tests and determines whether the average difference is statistically different from zero. Therefore, it is the difference within pairs of samples that is being tested. In this methodology, the variability in the results is the direct result of the testing procedure. The differences between pairs are calculated and, using a series of mathematical steps, a t-statistic is calculated. This t-ratio is then compared with a critical t-value obtained from the t-distribution. If the value is less than or equal to this critical value, one may infer that there is not a statistically-significant difference between the contractor's test values and the test values observed by the agency.

For Verification of independent samples, a different t-test is used, which involves testing against the pooled variance of both sets of samples. When performing the t-test one must ensure that several assumptions of the data are met. First, each data point must be independent of the others. In other words, the value of one data point cannot be dependent on the value of another. Second, the data must be approximately normal. For the analysis of transportation data, this means that no outliers or significant trends in the data should be visible. Lastly, the data samples (i.e., agency and contractor) must have an approximately equal variance. The F-test, described in detail later in this report, can be used to compare variances. If the assumptions cannot be met by a given data set there are alternatives to the t-test and F-test that are also statistically-valid but are a bit more complex. A few of these alternative methods are also covered below.

On the surface, the t-test seems to be mathematically-complex. In practice, however, the t-test is very easy and can be done in seconds by a trained individual using Microsoft Excel or similar data management program.

2.3.4.1 Sampling

The method by which samples are collected is a major determinant of the rigor of a sample comparison and the allowable scope of inference. Various characteristics of the sampling process, such as the number of samples taken, the method of selecting samples from the greater population, and the assignment of various samples to their treatment groups all impact the interpretation of statistical results and the conclusions that can be made. Therefore, it is of the utmost importance that sampling is performed in a structured, strategic, and formal method.

2.3.4.2 Randomization

When designing an experiment or comparison, such as QA/QC in a state DOT, the methodology implemented to collect samples is of the utmost importance. In fact, the sampling methods implemented often determine the entire nature of the comparison. For example, in a randomized comparison the investigator controls the acquisition of samples by groups (i.e., contractor or agency samples) with a chance mechanism such as a flip of a coin. A simple random sample from a large population may be a subset of the population selected in such a way that every subset is afforded the same chance of being selected. Alternatively, observational studies involve a non-random assignment of samples to their respective groups; observation studies are usually conducted where quality concern is focused on a small area or small quantity of material.

The motivation behind random sampling is that the scope of inference for a randomized study may be extended to the entire population from which the samples were obtained. In other words, inferences to the greater population should only be drawn from random sampling studies, but not otherwise (Ramsey and Schafer 2002). In a QA/QC study, where the ultimate objective is to determine the quality of the entire roadway (the population), randomization is critical.

The theory behind randomization is that the effect of confounding variables is minimized. Confounding variables are related both to group membership and the outcome. In a QA/QC study, if samples are not randomly selected from the roadway, the inferences from the comparison cannot be extended to the entire roadway or batch.

For QA/QC efforts, one may find that a stratified random sample is preferable. A stratified random sample may be performed when the population is first divided into subgroups and a random sample is selected from each subgroup (or sublot). This procedure is used to make sure all randomly selected samples are not concentrated in one section of the area to be sampled. This allows all strata (e.g., batches) that make up the entire population to be represented.

Drawing a random sample may be achieved through one of many different techniques. When the sample size is very small, the simplest way to randomize is to simply roll a die. Other methods include the use of a random number table, generating a random number using a spreadsheet such as Microsoft Excel or by using a simple pocket calculator. Weed (Weed 1984) provides an excellent, but perhaps outdated, discussion of simplistic randomization techniques.

2.3.4.3 Independence

Regardless of the statistical test implemented in a study, the samples (i.e., material sample, roadway core, etc.) must be independent of one another. Samples are said to be independent when the value of one observation (i.e., one sample) is not controlled or related to any other observations. Two types of independence violation are serial effects and cluster effects. Serial effects occur when measurements are taken over time; samples taken close together in time tend to be more similar than observations collected in distant time points. Similarly, cluster effects occur when data is collected in co-located sub groups such as batches (Ramsey and Schafer 2002). One should note that independence is a requirement for all statistical tests and that departures from independence are unacceptable. Randomization across time and across area are techniques that ensure independence is maximized.

2.3.5 Overview of statistical procedures

The Federal Highway Administration, authors of countless journal publications, other state DOTs, and other nations recognize that adequate verification procedures and valid QA analyses must involve statistical tests. Organizations that implement simplified strategies such as the D2S Limit methodology should re-evaluate their policies. Simple computer programs allow a user to perform an advanced, objective, and justifiable statistical procedure in very little time and without significant training. Knowledge of sampling methods, the theory that drives the statistical procedures, and the simple math that accompanies these tests is important to review, however. This section of the report provides an overview of several of the statistical procedures and identifies and illustrates an example of each method, following much of the discussion provided by Burati, et al. (2003).

When comparing two sets of data, such as contractor and agency test results, independent-sample hypothesis tests are required to evaluate a statistically significant difference between the samples. In QA/QC efforts, statistical tests are required to ensure that both samples (i.e., the contractor's and the Agency's) are from the same population. In such a statistical test the null hypothesis is that the samples are from the same population. In other words, the null hypothesis is that the variability of the two data sets is equal (using the F-test) and the means or medians are equal (using the t-test, permutation test, or Wilcoxon Rank-Sum test).

According to Burati, et al. (2003), when comparing two data sets it is important to compare both the means and the variances of the samples. Fortunately, construction material properties tend to follow a normal distribution, a requisite sample characteristic for most statistical tests. When samples follow a normal distribution the ratios of the variances tend to follow an F-distribution and the means follow a t-distribution. Therefore, tests on these distributions are appropriate. When samples appear to belong to non-normal distributions there are alternative statistical tests that may be performed.

This section presents several statistical tests such as the F-test, t-test, permutation test, and the Wilcoxon Rank-Sum test. The latter two tests are provided for the sake of completeness and generally are only applicable when the base assumptions of normality are violated. In most circumstances, it is sufficient to assume that the data is appropriate for use of the F-test and t-test. When using the F-test and t-test to compare agency and contractor test results, some fundamental characteristics must be true of the samples, as described in the following.

2.3.5.1 Introduction and assumptions

When using the t-tools to compare two samples several assumptions must be validated. The following three assumptions, in addition to independence, must be met. The t-tools are not robust against departures from normality, unequal variance (in most circumstances) or outliers.

2.3.5.1.1 Normality

Simply stated, normality refers to how well a sample distribution approximates a normal distribution. The F-test and t-test require that the samples follow an approximately normal distribution. Significant departures from normality reduce confidence in the statistical conclusions drawn from the tests. Fortunately, test statistics from large samples tend to follow a normal distribution according to the Central Limit Theorem. When the samples appear to belong from populations with non-normal distributions, alternative tests such as the Wilcoxon Rank Sum test must be used. Normality may be assessed by viewing a box plot of each sample.

2.3.5.1.2 Equal variance

When two samples have significantly different variances, problems with the t-tests exist because the t-ratio no longer follows the t-distribution. Statistical theory suggests that the t-tools are fairly robust from departures of equal variance of the samples as long as the sample sizes are roughly the same (Ramsey and Schafer 2002). In QA/QC studies, however, the sample size tested by the Contractor is typically significantly larger than that tested by the Agency. Therefore, the F-test of variance is used to ensure the viability of the t-tests.

2.3.5.1.3 Outliers

The presence of outliers indicates that a sample does not come from a normal population. Outliers have a unique impact on the accuracy of t-tests, the most common and most precise two-sample comparison. Unlike other statistical methods, the t-test involves the comparison of the sample *means*. Therefore, t-tests are not robust to outliers. This is true because one outlier can have a significant impact on the value of the mean, especially true for small samples. When outliers exist, other statistical tests or transformations of the data may be desirable. Tests such as the Wilcoxon Rank-Sum method replace values with ranks, eliminating the influence of outliers but reducing the precision of the resulting statistical comparison.

2.3.5.2 Split sample comparisons

2.3.5.2.1 Paired t-test

Overview: When split samples are used for the Independent Assurance evaluation, and multiple results are available for analysis, then the Paired t-test provides insight into whether there is a systematic difference between the pairs. If differences are not systematic, then they should tend to average toward zero. The Paired t-test examines whether the average difference among the pairs is statistically different from zero.

When conducting a Paired t-test one of the following two conclusions will be drawn:

1. The two sets of data represent different results because the average of the difference between the paired results is greater than is likely to occur from chance if the test results are actually equal; or

2. There is no reason to believe that the results of the two sets of data are different because the average difference between the paired results is not so great as to be unlikely to have occurred from chance.

In a Paired t-test, a t-value is calculated and compared against a value from a table that lists critical values for α using a two-tailed comparison (where α represents the Type I error, or the likelihood that the test will reject an actually acceptable set of paired samples). If the calculated value is less than the critical parameter, then the differences of the results of the paired tests are considered to be within acceptable limits. The t statistic is calculated using Equation 2-1:

$$t = \frac{\left|\overline{X}_d\right|}{\frac{s_d}{\sqrt{n}}} \qquad \textbf{(2-1)}$$

where: $\overline{X}_d$ = the average of the differences between the split sample results;
s_d = the standard deviation of the differences between the split sample test results; and
n = the number of split samples.

Steps required to perform a Paired t-test: The following procedure may be used to determine if a statistically significant difference between the paired results of the split samples exists. The following steps closely follow the procedure provided in Burati, et al.(2003; 2004):

1. Choose a level of significance for the test, α. It is suggested to use α = 0.01 for Independent Assurance comparisons;
2. Compute the differences between the results of the split sample pairs, and calculate the average ($\overline{X}_d$) and the standard deviation (s_d) for the differences;
3. Compute the degrees of freedom ($d.f. = (n-1)$);
4. Compute the t-statistic, using Equation 2-1;
5. Look up the appropriate t_{crit} using the t tables (Appendix C); and
6. Compare the t-statistic with the critical t-ratio. If the t-statistic $\geq t_{crit}$ then decide that the two sets of tests have statistically different means. Alternatively, if the t-statistic $< t_{crit}$ then decide that there is no reason to believe that the differences indicate a statistically significant variation between the ODOT split results and the contractor split results.

Example: The following example (from Burati, et al. (2003)) illustrates the steps requires to perform a t-test for a difference in means between a contractor's tests and Agency tests using asphaltic concrete test data. Please refer to the data in

Table 2.3 for the following example.

1. Choose a level of significance for the test, α. It is suggested to use α = 0.01 for Independent Assurance comparisons;
2. Compute the differences between the results of the split sample pairs, and calculate the average ($\overline{X}_d$) and the standard deviation (s_d) for the differences;

$$\overline{X}_d = \frac{0.60}{10} = 0.06$$

$$s_d = 0.05$$

3. Compute the degrees of freedom ($d.f. = (n-1)$);

$$d.f. = 10 - 1 = 9$$

4. Compute the t-statistic, using Equation 2-1;

$$t = \frac{|0.06|}{\frac{0.05}{\sqrt{10}}} = 3.95$$

5. Look up the appropriate t_{crit} using the *t* tables (Appendix C) for the degrees of freedom:

$$t_{crit} = 3.25 \ ,(\alpha = 0.01, d.f. = 9)$$

6. Compare the t-statistic with the critical t-ratio. If the t-statistic $\geq t_{crit}$ then decide that the two sets of tests have statistically different means. Alternatively, if the t-statistic $< t_{crit}$ then decide that there is no reason to believe that the differences indicate a statistically significant variation between the ODOT split results and the contractor split results:

$$(t = 3.95) \geq (t_{crit} = 3.25)$$

Conclusion: There is reason to believe that the two sets are different and the reason for the difference should be investigated.

Table 2.3: Example split sample data

Sample	Contractor Tests	Agency Tests	Difference
1	5.65	5.75	0.10
2	5.45	5.48	0.03
3	5.50	5.62	0.12
4	5.60	5.58	(0.02)
5	5.53	5.60	0.07
6	5.51	5.55	0.04
7	5.78	5.86	0.08
8	5.40	5.49	0.09
9	5.68	5.67	(0.01)
10	5.70	5.80	0.10
Number			10
Sum			0.60
Average			0.06
Standard Deviation			0.05
t-statistic			3.95

Benefits and limitations: The Paired t-test is an effective and simple test to identify systematic differences across multiple paired results, as may be available for large projects where Independent Assurance, through split sample testing, is applied on a project basis (rather than Independent Assurance applied on a systems, time-based method). It should be noted that not all systematic biases are detected by the split sample method; specifically, a bias that underestimates values early in the sequence, and gradually changes to a condition of overestimation of values late in the sequence would not be detected. However, with relatively few sample sizes, a simple inspection of the resulting differences should easily reveal these sorts of biases. In addition, it is a systematically high or systematically low bias that appears the most troublesome and worth investigation.

2.3.5.3 Independent sample comparisons

2.3.5.3.1 Tests of equal variance (F-test)

Overview: When comparing two independent samples, such as an Agency's sample and a Contractor's sample, a two-sample hypothesis test must be performed, using statistical procedures called the F-test and t-test. Such statistical procedures test for a statistically significant difference between the samples. In randomly selected and randomly assigned independent samples, a difference between the samples is attributed to the treatment effect. A statistically significant difference between the samples (i.e., independent samples) indicates a difference in the testing results performed by the Agency and the Contractor. Ultimately, the presence or absence of a statistical difference between results is what one must determine in a QA/QC comparison.

Independent two-sample t-tests (discussed later) are a statistically recognized method for comparing two independent samples with approximately equal variance obtained from an approximately normal population. The accuracy and, therefore, appropriateness of the t-test is dependent on the variance. The t-statistics do not follow the t-distribution when the variances of the two samples are unequal. This is especially true for small and/or unequal sample sizes. In QA/QC efforts, samples are often both relatively small and of unequal size. Therefore, an F-test for equal variances should always be conducted prior to conducting a t-test. That is, the variances should always be tested before the means.

The objective of conducting an F-test is to test whether a statistical difference exists between the Contractor's tests of quality and the Agency's tests. In other words one performs an F-test to determine if the difference in variability between the two sets of test results is greater than might be expected *by chance* if the samples came from the same population. A comparison of the variance will result in one of the following two conclusions:

- The two sets of data have different variances because the difference between the two sets of test results is greater than is likely to occur from chance if their variances are actually equal; or

- There is no reason to believe the variances are different because the difference is small enough as to likely have occurred from chance (Ramsey and Schafer 2002).

Steps required to perform an F-test: The following procedure may be used to determine if a statistically significant difference between the sample variances exists. The following steps have been created by combining guidance provided in Burati, et al. (2003) and Ramsey and Schafer (2002):

1. Compute the variance for the contractor's tests, s_c^2, and the Agency's tests, s_a^2. Before variance values may be calculated using Equation 2-3, the mean must be calculated using Equation 2-2.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \ldots + x_N}{N} \qquad \textbf{(2-2)}$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \textbf{(2-3)}$$

2. Use a simple ratio to compute F. Note that the larger of the two values must always be used in the numerator in order to obtain a ratio greater than 1, as in Equation 2-4.

$$F = \frac{s_a^2}{s_c^2} \quad \text{or} \quad F = \frac{s_c^2}{s_a^2} \qquad \textbf{(2-4)}$$

3. Identify the level of significance for the test, α. It is suggested to use $\alpha = 0.05$ for QA/QC comparisons. The authors refer the reader to statistical texts, such as Ramsey and Schafer (2002) for a discussion of such Type I error values.
4. Calculate the degrees of freedom for each sample by using Equation 2-5.

$$\begin{aligned} \text{Degrees of Freedom (Contractor)} &= (n_c - 1) \\ &\text{and} \\ \text{Degrees of Freedom (Agency)} &= (n_a - 1) \end{aligned} \qquad \textbf{(2-5)}$$

5. Determine the critical *F* value, F_{crit}, from an F-Table (Appendix D) for the α chosen and the degrees of freedom associated with the variance of each sample. For s_c^2 the degrees of freedom would be $n_c - 1$, and for s_a^2 the degrees of freedom would be $n_a - 1$. Since the concern is whether there is a two-sided difference (either larger or smaller) between the samples, use the F-table from Appendix D which includes values for the two-tailed F-distribution. This means that the F_{crit} values in Appendix D are the same values that would be listed for $\alpha = 0.025$ (i.e., $\alpha = 0.05 / 2 = 0.025$) in a one-tailed distribution.
6. Once the F_{crit} value has been determined from the table, compare the F statistic calculated in Step 2 with the critical value found in Step 5. If $F \geq F_{crit}$ then decide that the two sets of tests have statistically different variability. Alternatively, if $F < F_{crit}$ then decide that there is no reason to believe that the variabilities are statistically different.

Example: The following example (from Burati, et al. (2003)) illustrates the steps required to perform an F-test for a difference in variability between a contractor's tests and Agency tests using asphalt concrete test data (see Table 2.4).

Table 2.4: Example asphalt content tests

Sample	Contractor Tests	Agency Tests
1	6.11	5.95
2	6.14	6.33
3	6.45	6.21
4	6.22	6.55
5	6.11	6.34
6	6.08	6.15
7	6.37	--
8	5.97	--
9	6.01	--
10	5.99	--
11	5.87	
12	5.81	--
Number	12	6
Average	6.09	6.26
Standard Deviation	0.19	0.20
Variance	0.035	0.041

1. Compute the variance for each test, s_c^2 and s_a^2.

$$s_c^{\,2} = 0.035$$
$$s_a^{\,2} = 0.041$$

2. Use a simple ratio to compute F.

$$F = \frac{s_a^2}{s_c^2} = \frac{0.041}{0.035} = 1.17$$

3. Choose a level of significance for the test, α.

$$\alpha = 0.05$$

4. Calculate the degrees of freedom for each sample.

$$Degrees\ of\ Freedom\ \left(Contractor\right) = \left(n_c - 1\right) = 12 - 1 = 11 \text{ (denominator)}$$

$$Degrees\ of\ Freedom\left(Agency\right) = \left(n_a - 1\right) = 6 - 1 = 5 \text{ (numerator)}$$

5. Determine the critical F value, F_{crit}, from an F-Table (Appendix D).

$$F_{crit} = 4.04$$

6. Compare the F statistic from in Step 2 with the critical value from Step 5.

$$1.17 < 4.04$$

Conclusion: there is no reason to believe that the variabilities are statistically different.

Benefits and limitations of the F-test: The major benefit of using the F-test is that the procedure provides statistical, objective evidence for evaluating any difference between the sample variances. The most widely used alternative to the F-test is a visual comparison of box plots of the two samples. Unlike the F-test a review of side-by-side box plots allows the user to quickly evaluate any difference between sample variances without consulting cumbersome tables such as the F-table.

While side-by-side box plots provide a very quick visual check of equal variance, conclusions drawn by visual interpretation are largely subjective and should not be used when the F-test is feasible. Many computer packages, including Microsoft Excel, have simple commands that allow a user to conduct an F-test in a matter of minutes. The major limitation to the F-test is that it requires the user to consult an F-table. When consulting the F-table it is necessary to understand how to read and use the table. Therefore, caution is advised against the use of the F-table by untrained individuals.

Once equal variance has been confirmed, it is appropriate to use the t-tests to test for a difference in the sample means. If the F-test (or side-by-side box plots) indicates that the variances are significantly different, one should investigate why there may be a difference, or one could consider using the Wilcoxon Rank Sum or the permutation tests, both of which are distribution-free.

2.3.5.3.2 T-Test

Overview: If the F-test confirms that there is no statistically significant difference in the sample variances, then the sample appears to be derived from a normal population. If the samples are independent, then the t-test is an appropriate statistical method for testing for a difference in the sample means. Provided that the F-test indicates that the variances are approximately equal, the t-test utilizes the pooled estimate of the standard deviation and the pooled degrees of freedom. The pooled standard deviation is the square root of the pooled variance where the pooled variance is simply the weighted average of the two sample variances (using the degrees of freedom as the weighting factor).

When conducting a t-test one of the following two conclusions will be drawn:

1. The two sets of data have different means because the difference in the sample means is greater than is likely to occur from chance if their means are actually equal; or

2. There is no reason to believe that the means are different because the difference in the sample means is not so great as to be unlikely to have occurred from chance if the means are actually equal.

In a two-sample t-test, a t-statistic is calculated which quantifies the difference in the sample means in the context of the pooled samples and the null hypothesis that the samples are, indeed equal (Ramsey and Schafer 2002).

Steps required to perform a t-test: The following procedure may be used to determine if a statistically significant difference between the sample means exists. The following steps have been created by combining guidance provided in Burati, et al. (2003; 2004) and in Ramsey and Schafer (2002):

1. Choose a level of significance for the test, α. It is suggested to use α = 0.01 for QA/QC comparisons.
2. Compute the sample average for each test set, using Equation 2-2 from the F-test.
3. Compute the pooled degrees of freedom, as in Equation 2-6.

$$\text{Pooled Degrees of Freedom} = (na + nc - 2) \qquad \textbf{(2-6)}$$

4. Compute the pooled variance by using the sample variances obtained in Step 1 of the F-test. The pooled variance is simply a weighted average of the sample variances, where the degrees of freedom are used as weighting factors. The pooled variance may be calculated by using Equation 2-7.

$$s_p^2 = \frac{s_c^2(n_c - 1) + s_a^2(n_a - 1)}{n_c + n_a - 2} \qquad \textbf{(2-7)}$$

5. Compute the t-statistic, using Equation 2-8 for equal variances

$$t = \frac{|X_c - X_a|}{\sqrt{\frac{s_p^2}{n_c} + \frac{s_p^2}{n_a}}} \qquad \textbf{(2-8)}$$

6. Determine the critical t value, $t_{crit,}$ for the pooled degrees of freedom and the significance value using the t-table in Appendix C.
7. Compare the t-statistic with the critical t-ratio. If the t-statistic $\geq t_{crit}$ then decide that the two sets of tests have statistically different means. Alternatively, if the t-statistic < t_{crit} then decide that there is no reason to believe that the means are statistically different.

Example: The following example illustrates the steps required to perform a t-test for a difference in means between a contractor's tests and Agency tests using asphaltic concrete test data. Please refer to the data in Table 2.4 for the following example.

1. Choose a level of significance for the test, α.

$$\alpha = 0.01$$

2. Calculate the sample means.

$$Sample\ mean(contractor) = 6.09$$
$$Sample\ mean(agency) = 6.26$$

3. Compute the pooled degrees of freedom.

$$Degrees\ of\ Freedom(Contractor) = (n_c - 1) = 12 - 1 = 11$$
$$Degrees\ of\ Freedom(Agency) = (n_a - 1) = 6 - 1 = 5$$
$$Pooled\ Degrees\ of\ Freedom = 16 + 6 - 2 = 16$$

4. Compute the pooled variance.

$$s_p^2 = \frac{0.035(12-1) + 0.041(6-1)}{12+6-2} = 0.0369$$

5. Compute the t-statistic.

$$t = \frac{|6.26 - 6.09|}{\sqrt{\frac{0.0369}{12} + \frac{0.0369}{6}}} = 0.095$$

6. Determine the critical t-value from Appendix C.

$$t_{crit}\{\alpha = 0.01, \text{16 deg. of freedom}\} = 2.921$$

7. Compare the t-statistic from Step 5 with the critical t-ratio from Step 6.

$$2.921 > 0.095$$

Conclusion: there is no reason to believe that the means are statistically different.

Benefits and limitations: The t-test is the most accurate two-sample test of means. That is, the t-test most accurately evaluates the probability of a statistically significant difference in the means when all of the assumptions on which the test is based are satisfied. The major limitation of the test is its vulnerability to departures from normality and unequal variance. In fact, differences in variance can have negative impacts on the accuracy of the t-test when the sample sizes are unequal. In the case that there are unequal variances, the Welch's t-test is appropriate. The Welch's t-test is specifically designed for tests of samples with unequal variance. More information about the Welch's t-test can be found in statistic texts such as Ramsey and Schafer (2002). When there are significant departures from normality the Wilcoxon Rank Sum test (not to be confused with the Welch's t-test) is recommended, as it is a distribution-free test. While these tests allow a user to compare samples, the t-test is superior in every way when the normality and equal variance assumptions are met.

2.3.5.3.3 Wilcoxon Rank Sum

Overview: The Wilcoxon Rank-Sum test requires a simple form of transformation known as the rank transformation. Simply stated, the rank transformation requires the user to replace each value with its rank in the combined sample. This transformation is unique in that each transformed value depends on all of the other values and that it completely eliminates the importance of the population distributions (Ramsey and Schafer 2002). This transformation is especially useful in negating the impacts of extreme outliers.

The Wilcoxon Rank Sum test requires the use of an alternate statistic known as the Z-statistic. This statistic, like the t-statistic and F-statistic, is used to determine whether either sample is unusually large or small. Similarly, any disparity may be compared to the possible disparities attributable to random occurrence.

Steps required to perform a Wilcoxon Rank Sum test:

1. Choose a level of significance for the test, α. It is suggested to use α = 0.01 for QA/QC comparisons.
2. Transform the data sets from values to ranks. Note that each value should be replaced by its rank in the combined sample. That is, once all observations have been combined into one collective group, replace each value with its rank in the collective group.
3. Calculate the average and sample standard deviation of ranks from the combined sample by using Equations 2-2 and 2-3, presented above (Ramsey and Schafer 2002).
4. Compute the theoretical null hypothesis mean and standard deviation using Equations 2-9 and 2-10, respectively.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \textbf{(2-9)}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \qquad \textbf{(2-10)}$$

5. Compute the Z-statistic using Equation 2-11.

$$Z_{(X1-X2)} = \frac{(X_1 - X_2) - (\mu_{h1} - \mu_{h2})}{\sqrt{\frac{\sigma_{x1}^2}{n_1} + \frac{\sigma_{x2}^2}{n2}}} \qquad \textbf{(2-11)}$$

6. Compare the Z-statistic with the critical Z-statistic (from tables available in statistics textbooks). If the Z-statistic ≥ Z_{crit} then decide that the two sets of tests have statistically different means. Alternatively, if Z-statistic < Z_{crit} then decide that there is no reason to believe that the means are statistically different.

Wilcoxon Rank Sum example: The following example illustrates the steps requires to perform the test for a difference in means between a contractor's tests and Agency tests using asphalt concrete test data. Please refer to the data in Table 2.4 for the following example.

1. Choose a level of significance for the test, α.

$$\alpha = 0.01$$

2. Transform combined Table 2.4 data.

Table 2.5: Rank transformation

	Original Value	Rank
	6.55	1
	6.45	2
	6.37	3
	6.34	4
	6.33	5
	6.22	6
	6.21	7
	6.15	8
	6.14	9
	6.11	10
	6.11	11
	6.08	12
	6.01	13
	5.99	14
	5.97	15
	5.95	16
	5.87	17
	5.81	18
Number	18	18
Average	6.15	
Standard Deviation	0.20	

3. Calculate the average and sample standard deviation of ranks from the combined sample using Equations 2-2 and 2-3.

$$\overline{X} = 9.5$$

$$s_R = 0.20$$

4. Compute the theoretical null hypothesis mean and standard deviation

$$Mean(T) = 18(9.5) = 171$$

$$SD(T) \;=\; 0.20\sqrt{\frac{12 \; x \; 6}{(12+6)}} = 0.40$$

Table 2.6: Rank transformation of Table 2.4

Sample	Contractor Tests	Agency Tests
1	10.5	16
2	9	5
3	2	7
4	6	1
5	10.5	4
6	12	8
7	3	--
8	15	--
9	13	--
10	14	--
11	17	
12	18	--
Number	12	6
Average	6.09	6.26
Standard Deviation	0.19	0.2

5. Compute the Z-statistic:

$$Z = \frac{6.09 - 6.26}{\sqrt{\frac{0.19^2}{12} + \frac{0.22^2}{6}}} = -1.61$$

6. Determine the critical Z-statistic from tables available in statistics texts:

$$Z_{crit} = -1.55$$

7. Compare Z-statistic from Step 5 with the critical Z-statistic from Step 6:

$$Z < Z_{crit}$$

Conclusion: There is no statistically significant difference between the sample means.

Benefits and Limitations: The major benefit to the Wilcoxon Rank Sum procedure is that it is highly resistant to outliers and other departures from normality. The ranking transformation creates a distribution-free sample. The test is, however, not recommended when the assumptions of the t-tools are met, because the t-tools are significantly more accurate.

2.3.5.3.4 Permutation Test

Overview: A permutation test is any test that finds a p-value as the proportion of regroupings of the collective observations of the two groups ($n_1 + n_2$) that leads to test statistics as extreme as or more extreme than observed. When the test statistic is the difference between the sample means the permutation test provides the exact p-value for a comparison. Other tests, such as the t-test and Wilcoxon Rank Sum test are only approximations of the permutation test.

Steps required to perform a permutation test: The following steps are required when performing a permutation test for a difference in sample means (Ramsey and Schafer 2002):

1. Decide on a test statistic, and compute its value from the two samples.
2. List all of the regroupings of the combined group ($n_1 + n_2$) of size n_1 and n_2 and compute a test statistic for each.
3. Count all of the number of regroupings that produce test statistics at least as extreme as the observed test statistic from Step 1.
4. The p-value is the number found in Step 3 divided by the total number of regroupings.

Due to the relatively high number of regroupings of the example data from Table 2.4 required to perform a permutation test, an example will not be provided in this document. Please consult Ramsey and Schafer (Ramsey and Schafer 2002) pages 95-97 for an example and discussion of permutation tests.

Benefits and limitations: The chief benefit of the permutation test is that the p-value calculated through the permutation process is the exact p-value, not an approximation. The major limitation of the permutation test is that it is labor intensive when the number of samples is large. Additionally, most computer programs and spreadsheets will not perform the permutation test. When the number of samples is very small (e.g., 10 or below) the permutation is feasible.

3.0 QA PROCEDURES AT ODOT

3.1 INTRODUCTION

3.1.1 General discussion

To meet FHWA requirements, state agencies utilize a variety of strategies and practices to comply with the Code of Federal Regulations (23 CFR 637)—see Appendix B. The Quality Assurance program required by the CFR is needed to ensure that the quality of materials incorporated into a federal aid project satisfactorily meets the minimum requirements established in the contract documents. Programs have evolved over the years, and different states establish different guidelines to comply with the requirements written in the CFR (Hughes 2005). Some basic principles that state agencies incorporate into their quality assurance programs include:

- Attributes for QC acceptance;
- Test methods for QC acceptance;
- Point of sampling for QC acceptance;
- Testing frequencies;
- Personnel training/certification;
- Quality measures of Acceptance;
- Material accept/reject provisions;
- Contractor testing – Quality Control;
- Verification; and
- Independent Assurance.

Acceptance programs are based on quality assurance principles to provide "…confidence that a product or facility will perform satisfactorily in service." QA programs have been described as consisting of three primary functions: Quality Control, Acceptance, and Independent Assurance; Figure 2.1 in the literature review chapter provides this generic view of the quality assurance process and the primary functions incorporated into a typical State Highway Agency (SHA) QA program. ODOT utilizes a slight variant of this generic view, preferring instead to use *Verification*, rather than *Acceptance*, for the second function of *Quality Assurance*. The resulting diagram used by ODOT is shown in Figure 3.1. Definitions of the functions indicated in Figure 3.1 are as further described in this section. The procedures and guidelines for the Oregon Department of Transportation QA Program take into account the items listed above, including specific material testing procedure guidelines outlined in its Manual of Field Test Procedures (MFTP).

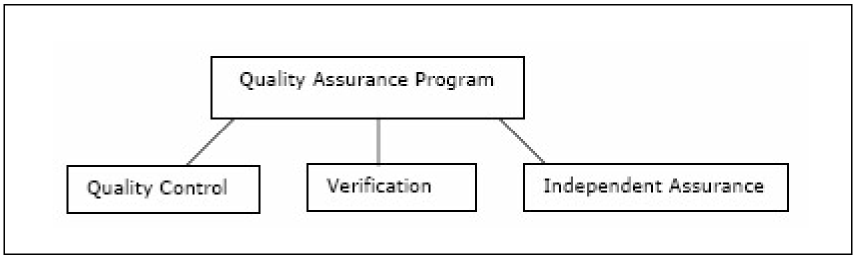


Figure 3.1: ODOT representation of QA functions

3.1.2 ODOT quality assurance program

There are several methods used for meeting the principal guidelines of a quality assurance (QA) program. Each of these functions is accomplished through various methods among state DOTs in order to develop specific guidelines for material quality assurance. Such choices include agency-furnished versus contractor-furnished quality control, agency versus contractor sampling for Independent Assurance and Verification sampling, and the use of D2S or similar evaluations versus t-test and F-test methods for comparing between agency and contractor test results. ODOT's choices are described below, and are further assessed regarding their effectiveness and regarding their compliance with FHWA guidance.

It should be noted that the discussion here relates to those products that are tested for properties that may be acceptable at varying levels, those for which an Independent Assurance process has been established, and especially those for which pay adjustments are generated using statistical data. Some materials are not considered in this analysis due to the nature of the testing and the immediate decision to accept or reject or correct the material. Such materials are tested in the field and 1) are subject to immediate rejection (or immediate correction) due to the failure of the material to meet a threshold (for example, Portland cement concrete slump); or 2) are tested for project-level product compliance (for example, batched materials such as aggregates or concrete mixes). To gain a better understanding of the quality assurance and sampling process, a description of some key quality assurance testing requirements for these materials is provided below.

The Oregon Department of Transportation has produced and continuously updates a manual to outline the specific procedures for collecting, inspecting, and testing materials for construction projects. This document is referred to as the ODOT Manual of Field Test Procedures (ODOT 2006) and will be referenced continuously throughout this section as the "MFTP." Included in this manual is direction for the QA guidelines that includes Quality Control, Verification, and Independent Assurance procedures. ODOT specifies QC as mandatory and denotes the appropriate QC procedures and standard specifications contractors must abide by on all state highway and bridge contracts.

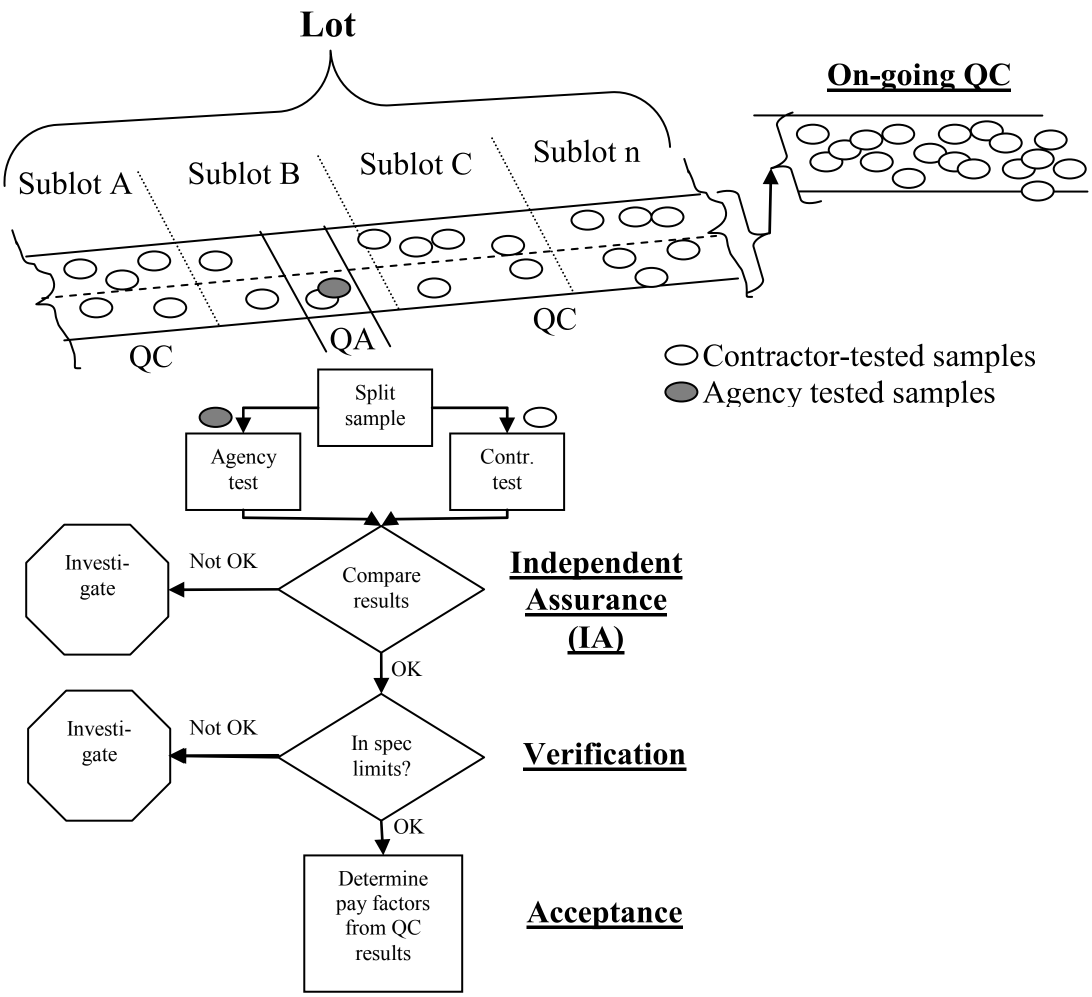


Figure 3.2: ODOT QA/QC process

Figure 3.2 illustrates the general process for QA at ODOT; a brief description of the key elements is as follows:

- **Quality Control (QC)** is the responsibility of the contractor. In this category, the contractor provides both the sampling and the testing for the material. The contractor's QC Certified Technician must observe and perform testing operations, properly document test results, and sign the documentation. The Quality Control tests performed by the contractor will be used if verified by ODOT's Quality Assurance process.

- **Independent Assurance (IA)** is the responsibility of the Agency. ODOT tests a selected number of split samples independently obtained with contractor assistance; note that both Independent Assurance and Verification in practice use results from the same sample sets (see Figure 3.2). The purpose of the testing is to determine that the contractor-provided testing is accurate. The method of choosing the frequency and quantity of the split samples varies and is outlined in the MFTP. In general, the split sample is randomly selected from one day's production or other specified quantity of a specific material. The sample is tested by ODOT at a certified laboratory and compared against contractor test results for its portion of the split. A comparison of the test statistic is made between the contractor-based and the ODOT-based test results to determine whether the two results are acceptably close (see Table 3.1). This method is heretofore called the "one-to-one tolerance" evaluation.

- **Verification** is the responsibility of the Agency. ODOT's process includes splitting a contractor-obtained sample (the same sample used for Independent Assurance) and testing its portion of the split for material compliance. The contractor independently tests its portion of the split. The results of ODOT's and the contractor's tests are compared against specification requirements as an indication of material compliance and validation of the contractor's QC results.

The focus of this research is to evaluate whether an alternate method of comparison, the t-test and the F-test for example, may be helpful in improving the effectiveness of the QA process.

3.1.2.1 Random sample procedure

A fundamental basis of the ODOT Quality Assurance Program is the hypothetical assumption of sample equality and the application of statistical tools to test that assumption (ODOT 2006). For the assumptions of these tools to be met, and in order to ensure that the scope of inference is sufficiently broad, Independent Assurance (IA) and Verification samples are collected at random. In order for a sample to be considered random, all material must have an equal chance of appearing in a sample, and all sampling patterns that present bias are removed from the sample population. For density test locations, random samples may be generated according to ODOT TM 400, Determining Random Sampling and Testing Locations, or an alternative approved method; for other samplings, randomness is achieved on a more subjective basis—the timing and location of the sampling is based on the judgment of the ODOT field engineer.

3.1.2.2 Sampling parameters and procedure

To effectively ensure that contractor testing processes are reliable, ODOT compares test results between the contractor results and the agency test results. According to the *Manual of Field Test Procedures*, allowable differences between these values are predetermined and may vary depending on the material. A summary of the IA tolerances and the allowable differences are provided in Table 3.1.

Table 3.1: Independent Assurance tolerances

Material	Specification	Maximum Allowable Difference
Gradation (Sieve Sizes)		
	Larger than 2.36 mm (No. 8)	5.00%
	2.36 mm (No. 8)	4.00%
	2.00 mm (No.10)	4.00%
	Larger than 75 υm (No. 200) and smaller than 2.00 mm (No. 10) 2%	2.00%
	75 υm (No. 200) and smaller	1.00%
Asphalt Content		0.40%
Fracture		5.00%
Wood Particles		0.05%
Elongated Pieces		2.00 - 4.00%
Sand Equivalent		8 points
Moisture Content (Plant Mix Aggregate Base)		0.50%
Soil Curves - Maximum Density - Df		
	Density	50 kg/m3 (3.0 lbs/ft^3)
	Moisture 2%	2.00%
Aggregate Base - Maximum Density - Df		
	Density	50 kg/m3 (3.0 lbs/ft^3)
	Moisture	2.00%
Maximum Specific Gravity (Rice T-209)		0.020
Bulk Specific Gravity (Lab fabricated specimens T-I66)		0.020
Maximum Specific Gravity (T-85)		0.032
Air Content of Concrete (T-152)		0.50%
Slump of Concrete (T-119)		20 mm or ¾”
Temperature of Concrete (T-309)		2º C (3° F)
Unit Weight of Concrete (T-121)		50 kg/m3 (3.0 lbs/ft^3)

As a general rule, the agency provides IA and/or Verification testing for each category of contractor-performed tests that gauges quality parameters for which a range may be acceptable. Sample quantity depends on project size; for many ODOT projects and materials, only one sample per material lot is required by the Agency; in many cases this may be one sample for the entire project. Each sample is evenly split between the contractor and the agency and tested by each; the result of ODOT’s test and the contractor’s test are then compared to each other, and must fall within IA tolerances. If the difference between the contractor’s test results and the agency’s fall within these tolerances, the contractor’s QC results are considered verified and used for Acceptance and pay factor evaluation.

If the Contractor and ODOT test results compare favorably within the allowable tolerances specified in Table 3.1, then Independent Assurance is established and the Verification process begins. For Verification, the results from both split sample tests are compared against the specified requirements for the material. If the both material tests fall within the specified bounds for the material, then ODOT accepts the results as an indication that the contractor’s QC results are reliable (Verification). If that is not the case, then ODOT investigates and performs appropriate actions to resolve the discrepancy.

Several state DOTs implement similar simple comparisons. Such simple comparison methods are not recommended in the literature as a statistically-valid procedure for product acceptance (CFR 1995); however, as will be discussed further, investigation of ODOT's sampling methods show that the ODOT methods provide insight into material acceptability that are helpful in protecting product quality. For reference, Burati, et al. (2004) provide an overview of a variety of QA comparison methods including methods no longer recommended by the Federal Highway Administration (FHWA).

3.2 ODOT MATERIAL TEST PROCEDURES AND GUIDELINES

Four material groups are further considered here to determine whether a different statistical procedure is useful for the Verification process. These include:

- Hot Mix Asphalt Concrete;
- Portland Cement Concrete (structural);
- Aggregates; and
- Earthwork.

The primary focus of this section is to research and document the general practice of sampling, testing, and verification for quality assurance purposes. A summary of each of these materials is examined, including information for each material regarding the practice of Quality Control, Verification, and Independent Assurance. Tables are also provided to indicate the types of tests that are performed and at what frequencies.

3.2.1 Hot Mixed Asphalt Concrete (HMAC)

3.2.1.1 Quality Control

To comply with test procedure requirements, the contractor performs random sampling and testing of its field-placed materials and tracks the results as part of its ongoing Quality Control (QC) process. The contractor's CAT I (Certified Asphalt Technician I) performs the tests and records the results. ODOT requires HMAC to be tested for both a variety of mixture and compaction statistics, as indicated in Table 3.2 per the MFTP (ODOT 2006). If QC testing fails to meet the specifications, the contractor is to report to the ODOT project manager to determine product acceptance or rejection.

3.2.1.2 Independent Assurance (IA) and Verification

ODOT's Independent Assurance (IA) Program requires laboratory certification, technician certification, proficiency samples, and split samples. For mix evaluation, the contractor takes a random material sample at the batching plant under strict ODOT direction; a split of the sample is given to ODOT, and both conduct tests for asphalt mixture statistics. The contractor's test results are compared to ODOT's test results for compliance with allowable differences (see Table 3.1) for IA compliance and to determine that the results fall within specified limits. Compaction testing occurs in the field; there are no specific IA comparison checks made, except for occasional ad hoc, side-by-side checks to ensure that the calibrated testing machines produce similar results. Verification of compaction testing occurs through a procedure by which ODOT randomly selects a location within a sublot to test, and verifies that its results comply with specification requirements. Table 3.2 provides a further look into the QA elements and sampling frequencies required, based on material characteristics.

Table 3.2: Testing frequencies and requirements for HMAC

Material and Operation	Test Description	Test Method	Minimum Test Frequency	
			Contractor QC	IA/Verification
Aggregate Production				
	Sampling	AASHTO T 2	1/Sublot	10% of QC
	Reducing	AASHTO T 248	1/Sublot	10% of QC
	Sieve Analysis	AASHTO T 27/T 11	1/Sublot	10% of QC
	Sand Equivalent	AASHTO T 176	1/Sublot	10% of QC
	Elongated Pieces	ODOT TM 229	1/5 Sublots*	10% of QC
	Fracture	AASHTO TP 61	1/5 Sublots*	10% of QC
	Wood Particles	ODOT TM 225	1/5 Sublots*	10% of QC
Gradation				
Ignition Method	Sampling	AASHTO T 168	1/Sublot	10% of QC
	Reducing	WAQTC TM 5	1/Sublot	10% of QC
	Sieve Analysis	AASHTO T 30	1/Sublot	10% of QC
Asphalt Content				
Meter Method		ODOT TM 321	1/Sublot; Min 1/day	10% of QC
		ODOT TM 322	1/Sublot; Min 1/day	10% of QC
	Cold Feed Moisture	AASHTO T 255/265	1/Sublot	10% of QC
	Sampling	AASHTO T 168	1/Sublot; Min 1/day	10% of QC
	Reducing	WAQTC TM 5	1/Sublot; Min 1/day	10% of QC
	Asphalt Content	AASHTO T 308	1/Sublot; Min 1/day	10% of QC
Mix Design Verification Testing				
Fabrication	Gyratory Specimen	ODOT TM 326	1/Sublot	10% of QC
Maximum Density Test	Max Specific Gravity	AASHTO T 209	1/Sublot	10% of QC
Determination of Gmb	Bulk Specific Gravity	AASHTO T 166	1/Sublot	10% of QC
Compaction	Nuclear Density	WAQTC TM 8	5/Sublot	10% of QC
Asphalt Cement	Compliance	AASHTO T 40	1/Sublot	10% of QC
Lime or Latex Treatment of Aggregate	% Hydrated Lime	ODOT TM 321	1/Sublot	10% of QC
		ODOT TM 322	1/Sublot	10% of QC

* Testing also required at start of production

Note: 1 sublot = 1000 tons for HMAC

If the Contractor's HMAC QC test results are out of specification, the IA test results are not within the IA parameters, or the ODOT Verification tests are out of specification, the ODOT Project Manager investigates and takes appropriate action to determine whether the contractor's QC results will be acceptable for calculation of the Pay Factors. A summary of testing frequency requirements is provided in Table 3.2.

3.2.2 Aggregates – base, subbase, and shoulders

3.2.2.1 Quality Control

The contractor performs Quality Control testing as specified in the Manual of Field Test Procedures for all the aggregates, including base, subbase and shoulder aggregate. The contractor's QC technician identifies the maximum densities and optimum moisture content for each unique aggregate mixture type included in the project, performs testing on a continual basis during production, tracks the results, and reports the results to ODOT. For aggregate placement and compaction of subbase, base, and shoulders the contractor QC technicians perform random nuclear gauge density testing at a frequency of five tests per sub lot. Quality Control test frequencies are summarized in Table 3.3.

3.2.2.2 Independent Assurance (IA) and Verification

Aggregate materials production is tested by ODOT for Independent Assurance and Verification using split sample methods. Tests are categorized based on the specific use of the material itself – subbase, base, and shoulder aggregates. ODOT's Independent Assurance (IA) Program requires laboratory certification, technician certification, proficiency samples, and split samples for the IA process. The contractor takes a random sample; a split of the sample is given to ODOT; and both conduct tests for aggregate mixture statistics. The contractor's test results are compared to ODOT's test results for compliance with allowable differences (see Table 3.1). If the results are within the parameters of Table 3.1, the results are then evaluated for compliance with the specified limits.

Compaction testing occurs in the field; there are no specific IA comparison checks made, except for occasional ad hoc, side-by-side checks to ensure that the calibrated testing machines produce similar results. Verification of compaction testing occurs through a procedure by which ODOT randomly selects a location within a sublot to test and verifies that its results comply with specification requirements. If the density test fails, the Contractor identifies the limits of failing compaction, takes corrective action, and notifies the ODOT Project Manager. Subsequently, the ODOT Project Manager schedules a new Verification test. Additional lifts are not allowed until the Verification test confirms that the specified densities exist.

Contractor Quality Control test results, whose reliability is inferred through ODOT's Verification tests, provide the basis for the project manager's decision for acceptance, partial acceptance, or rejection of materials and products. Quality Control, Verification and Independent Assurance testing frequencies for base, subbase, and shoulder aggregates are provided in Table 3.3.

Table 3.3: Testing frequencies and requirements for aggregate

Material and Operation	Test Description	Test Method	Minimum Test Frequency Contractor QC	 IA/Verification
Aggregate Subbase				
	Sampling	AASHTO T 2	1/Project	Visual
	Reducing	AASHTO T 248	1/Project	Visual
	Sieve Analysis	AASHTO T 27	1/Project	Visual
	Sand Equivalent	AASHTO T 176	1/Project	Visual
Aggregate Base and Shoulders				
	Sampling	AASHTO T 2	1/Sublot	10% of QC
	Reducing	AASHTO T 248	1/Sublot	10% of QC
	Sieve Analysis	AASHTO T 27	1/Sublot	10% of QC
	Sand Equivalent	AASHTO T 176	1/Sublot	10% of QC
	Fracture	ASSHTO TP 61	1/5 Sublots	NR
PLACEMENT				
Aggregate Base only				
Plant mix aggregate only	Sampling	AASHTO T 2	1/Sublot	10% of QC
	Reducing	AASHTO T 248	1/Sublot	10% of QC
	Moisture	AASHTO T 255/265	1/Sublot	10% of QC
Compaction	Deflection Testing	ODOT TM 158	5/Sublot	10% of QC
	Nuclear Gauge	AASHTO T 310	5/Sublot	10% of QC

Note: 1 Sublot = 2000 tons for aggregate

3.2.3 Portland Cement Concrete (PCC):

3.2.3.1 Quality Control

The contractor is required to perform Quality Control (QC) sampling and testing at frequencies outlined in the ODOT Field Testing Manual. Testing requirements for Portland Cement Concrete are summarized in Table 3.4. The contractor must perform tests using specified cure methods and other testing procedures, including the required size and number of cylinders to be cast for each sample set.

3.2.3.2 Independent Assurance (IA) and Verification

For plastic property verification, the contractor is required to test the same load and portion of load from which Verification samples are taken. The agency representative will verify that the sample results are within IA parameters. If results are not within compliance, the Quality Assurance Technician (QAT) resolves the issue in the field. For strength testing, the contractor QC technician casts cylinders equal to the cylinders required in the QC plan. ODOT independently makes and tests a separate set of cylinders, and differences between ODOT's and the contractor's results are resolved by the ODOT project manager.

Table 3.4: Testing frequencies and requirements for PCC (structural)

Material and Operation	Test Description	Test Method	Minimum Test Frequency	
			Contractor QC	IA/Verification
Section 00540 Structural Concrete				
Aggregate Production				
	Sampling	AASHTO T 2	1/Sublot	10% of QC
	Reducing	AASHTO T 248	1/Sublot	10% of QC
	Sieve Analysis	AASHTO T 27/T 11	1/Sublot	10% of QC
	Fineness Modulus	AASHTO T 27/T 11	1/Sublot	10% of QC
	Wood Particles	ODOT TM 225	1/Sublot	10% of QC
	Sand Equivalent	AASHTO T 176	1/Sublot	10% of QC
Portland Cement Concrete				
	Sampling	WAQTC TM 2	1 set/100 CY; Minimum 1/Day: 1 Set/200 CY after reaching 600 CY in one day	Projects < 100 CY; 1/Project representing all PCC classes: Projects > 100 CY; 1/500 CY
	Air Content	AASHTO T 152		
	Slump	AASHTO T 119		
	Temperature	AASHTO T 309		
	Density	AASHTO T 121		
	Yield	AASHTO T 122		
	Water/Cement Ratio	AASHTO T 123		
	Strength	AASHTO T 22/23		

Note: 1 Sublot = 2000 Tons for aggregate
1 cylinder set = 3 cylinders averaged at 28 day cure

3.2.4 Earthwork

3.2.4.1 Quality Control

The Contractor is required to perform Quality Control (QC) testing according to the ODOT Manual of Field Test Procedures (MFTP) for all the earth work material. The Contractor's QC technician identifies the maximum densities and optimum moisture content for each unique soil type and soil/aggregate mixture used on the project. For soil compaction, the contractor’s technician performs random sampling to ensure that materials and compaction are in accordance with the project specifications. Quality Control test frequencies are summarized in Table 3.5.

3.2.4.2 Independent Assurance (IA) and Verification

For Independent Assurance and Verification, compaction testing occurs in the field; there are no specific IA comparison checks made, except for occasional ad hoc, side-by-side checks to ensure that the calibrated testing machines produce similar results. Verification of compaction testing occurs through a procedure by which ODOT randomly selects a location within a sublot to test, and if its results comply with specification requirements, then the contractor’s QC data is inferred to be reliable. In the case of compaction, ODOT takes random samples at a minimum rate of 10% of the required QC tests. The 10% testing may be increased when deemed necessary by the project engineer.

If the soil compaction does not match the established curves, the contractor is required to establish a new curve from the soil at the test location prior to installation of additional lifts on top of the material in question. If the density test fails, the contractor identifies the limit of failing compaction, performs corrective measures, and reports to ODOT, who verifies corrective measures were effective with additional tests. New lifts cannot be added until the Verification test results show the densities meet contract specifications.

Table 3.5: Testing frequencies and requirements for embankment

Material and Operation	Test Description	Test Method	Minimum Test Frequency	
			Contractor QC	IA/Verification
Earthwork				
Establishing Maximum Density	Density Curve	AASHTO T 99	1/Soil type	1/Project
	Bulk Specific Gravity	AASHTO T 85	1/Soil type	1/Project
		AASHTO T 272	NR	NR
Compaction	Deflection Testing	ODOT TM 158	1/Yard in depth	
	Nuclear Gauge	AASHTO T 310	Projects < 3500 CY; 1/500 CY: Projects > 3500 CY; 1/3000 CY	10% of QC
	Coarse Particle Correction	AASHTO T 224		
	Deflection Testing	ODOT TM 158		

3.2.5 Very small project quantity guidelines

ODOT does not require a full QC analysis on all projects. According to the Manual of Field Test Procedures (ODOT 2006), projects with material quantities less than indicated values are exempt from the QA/QC process. The boundaries for these material quantities that are exempt from typical testing frequencies are provided in Section 4(B) of the MFTP. This information has been reproduced in Table 3.6.

Table 3.6: Very small project quantity table

Section	Type of Material	Approximate Quantity
00330	Earthwork (Embankment)	500 m^3 (yd^3)
00330	Earthwork (Excavation)	500 m^2 (yd^2)
00345 & 00346	Lime & Cement Treated Subgrade	2000 m^2 (yd^2)
00390 & 00395	RipRap & Rock Gabions	100 m^3 (yd^3)
00405	Ditch & Trench Excavation, Bedding and Backfill	50 m^3 (yd^3)
00440	Commercial Grade Concrete	50 m^3 (yd^3)
00641 & 00642	Aggregate Sub-base, Base & Shoulders	2000 Mg (Ton)
00680	Stockpiled Aggregate	2000 m^3 (yd^3)
00730	Asphalt Tack Coat	50 Mg (Ton)
00735	Emulsified Asphalt Concrete Pavement (includes asphalt cement)	2500 Mg (Ton)
00744 & 00745	Hot Mix Asphalt Concrete (HMAC)	2500 Mg (Ton)

If a project involves a quantity of material that is equal to or less than the quantities listed in Table 3.6, the contractor is not required to perform QC testing if at least one of the following conditions are met:

- Similar material from the same source has been accepted for use on ODOT projects within the past two years, and was found to be satisfactory under the Department's QA program;
- A Quality Compliance Certificate is provided, verifying that the material conforms to the contract requirements;
- Other information is provided, indicating the method or workmanship that the Contractor will utilize to assure that all the contract requirements will be met; or
- For section 00330 (Earthwork) a minimum of one Deflection test per area is provided, performed by an ODOT Certified Density Technician (CDT). The contractor's written request must identify the distinct work areas that small quantity acceptance is required.

3.2.6 Assessment of ODOT Procedures for Independent Assurance and Verification

ODOT provides procedures for both Independent Assurance and Verification testing of the work. The work is generally broken down into Lots, and often a project may contain only one Lot of a given material. These Lots are additionally broken down into Sublots, which are then each assessed as to what degree they meet product specifications based on test results produced by the contractor. If pay factors apply, the quality of each Sublot is used to establish the payment for the material provided.

Independent Assurance testing (identified in Figure 3.2), i.e., testing of a split sample to ensure that the contractor's testing methods produce accurate results, begins with ODOT's identification of a sublot from which a prescribed sample is set aside, which is then split into two equal parts ("splits"). One split is tested by the Agency and one split is tested by the contractor. The results from the tests are then compared. If the difference between them is found to be within allowable tolerances, then the contractor's testing method is considered to be validated, and the Agency then uses the test results to compare against specified values in its Verification process. In the case of compaction, there are no specific IA comparison checks made, except for occasional ad hoc, side-by-side checks to ensure that the calibrated testing machines produce similar results.

Verification of the contractor's Quality Control (QC in Figure 3.2) results (i.e., whether there is a consistent bias toward higher or lower test results) is accomplished using the same results as obtained in the Independent Assurance testing. In this procedure, the results of the Agency's and the contractor's split sample tests are compared against the specified parameters for the given material. If the result falls within acceptable limits—and if the Independent Assurance test was acceptable—then the Contractor's QC results are considered verified and are subsequently used to establish the appropriate pay factors. If the result does not fall within acceptable limits, then additional testing, further investigation, and potential rejection of the work may ensue.

3.2.7 Sampling quantities

In order to examine the efficacy of alternative statistical methods, the number of samples and test results for a variety of project situations were determined. A breakdown of three project sizes (small, medium, and large) was established; each size was given a range of material quantity for the four focal materials of HMAC, Aggregate Base, Soil fill, and Structural PCC. The resulting project models for each material are shown in Table 3.7; note that the low range for the "small" projects is the upper limit for the very small projects exempted from QA, as noted earlier in Table 3.6. Recommendations and policy decisions may be evaluated based on project size and sampling quantity, showing how statistical procedures may benefit different sizes of projects. The size categories allow the analysis to be broken down based on margin of error and statistical characteristics. It should be noted that the quantities established in these models are solely for the purpose of this study and are not intended to indicate any formal or policy-based distinction among projects; further, the upper bound on the large project size is established only for modeling purposes—actual projects may well exceed the limit shown.

Table 3.7: Project size parameters

		Small Project				Medium Project				Large Project			
		Low range		High range		Low range		High range		Low range		High range	
Description	Unit	IA	QC	IA	QC	IA	QC	IA	QC	IA	QC	IA	QC
HMAC (mixture)	Tons	2,500		6,000		6,000		30,000		30,000		200,000	
HMAC (compaction)	Tons	2,500		6,000		6,000		30,000		30,000		200,000	
Aggregate (mixture)	Tons	2,000		3,000		3,000		25,000		25,000		100,000	
Aggregate (compaction)	Tons	2,000		3,000		3,000		25,000		25,000		100,000	
Embankment (compaction)	CY	500		5,000		5,000		40,000		40,000		200,000	
Portland Cement Concrete	CY	50		500		500		2,500		2,500		50,000	

Based on the minimum testing requirements in the ODOT MFTP and standard specifications, the total number of tests for the material can be found by dividing the project size parameters by the number of required tests per quantity of material. Data from Table 3.2 through Table 3.5 were summarized to produce the simplified table of testing frequencies (Table 3.8), indicating the required number of tests per quantity of material. The quantity of test samples is based on the testing frequency requirements for the contractor's Quality Control program, the random IA sampling performed by ODOT, and the IA testing requirements. Note that some material applications differ, and testing may be done at varying frequencies based on contract and project conditions.

Table 3.8: Testing frequencies (per MFTP guidelines, Section 4(D))

Description	QC	QA	Minimum Test Frequency:
HMAC (mixture)	1/Sublot	10% of QC	1 sublot = 1000 Tons
HMAC (compaction)	5/Sublot	10% of QC	1 sublot = 1000 Tons
Aggregate (mixture)	1/Sublot	10% of QC	1 sublot = 2000 Tons
Aggregate (compaction)	5/Sublot	10% of QC	1 sublot = 2000 Tons
Embankment (compaction)	1/500 CY	10% of QC	QC: Projects ≤ 3500 CY - 1/500 CY; Projects ≥ 3500 CY - 1/3000 CY
Portland Cement Concrete	1/100 CY	1/500 CY	QA: Projects ≤ 100 CY - 1/Class; Projects ≥ 100 CY - 1/500 CY QC: First 600 CY - 1/100 CY; Above 600 CY - 1/200 CY

Using this methodology, Table 3.9 summarizes the minimum number of test samples required for each project category. The sampling amounts in Table 3.9 were derived from a listing of the material and test frequency requirements, as shown in Table 3.8. Table 3.9 provides the number of samples available for statistical analysis based on the size of the project. The QC column refers to the number of test results required for the contractor's Quality Control program, while the value in the IA column refers to the number of samples required based on the testing requirements for Independent Assurance and Verification. These testing requirements are established in Section 4(D) of the ODOT MFTP. Test sample requirements compiled for this report are derived from the minimum frequencies listed in Table 3.2 through Table 3.5.

Table 3.9: Minimum number of QA samples, based on project size

		Small Project				Medium Project				Large Project			
		Low range		High range		Low range		High range		Low range		High range	
Description	Freq.	QA	QC	QA	QC	QA	QC	QA	QC	QA	QC	QA	QC
HMAC (mixture)	**1000**	1	3	1	6	1	6	3	30	3	30	20	200
HMAC (compaction)	**1000**	1	3	1	6	1	6	3	30	3	30	20	200
Aggregate (mixture)	**2000**	1	1	1	2	1	2	2	13	2	13	5	50
Aggregate (compaction)	**2000**	1	1	1	2	1	2	2	13	2	13	5	50
Embankment (compaction)	**3000**	1	1	1	2	1	2	2	14	2	14	7	67
Portland Cement Concrete	**500**	1	1	1	1	1	1	1	5	1	5	10	100

The data above represent the main types of tests frequently encountered on a project. Other tests may be required per contract requirements, although many of the tests are conducted from the same samples and therefore represent the same number of samples as the tests listed in Table 3.9.

3.3 PROCESS EVALUATION

3.3.1 NCHRP Study Results

NCHRP published a study in 2006 (Parker and Turochy), the purpose of which was to compare the effectiveness of using contractor-performed tests in quality assurance, especially in the Verification/Acceptance decision. This study obtained historic HMAC test results from six states—Georgia, Florida, North Carolina, Kansas, California, and New Mexico—and evaluated those results, comparing agency test results against contractor QC test results. Three of these states (Georgia, Florida, and N. Carolina) verified contractor results using a one-to-one tolerance comparison, similar to the method used at ODOT. Two states used F&t-tests as the verification method (Kansas and New Mexico), and California used both. Statistical comparisons between agency and contractor tests were made at the 1% level (Alpha=0.01) of significance, even though the 5% level (Alpha=0.05) is the more commonly used measure. The authors chose the 1% level, since it is a more strict determinant of differences.

The study found that consistently, "contractors' and state DOT test results for HMAC are statistically different (Alpha=0.01)", and that this difference existed regardless of the details of the Verification process. In other words, consistent bias between contractor QC results and agency Verification results was found in the study. The comparisons further indicated that the variance led to more favorable contractor results in the Acceptance decision, over the Acceptance outcome if DOT results had been used.

The conclusion of the study's authors was that DOTs should consider moving from contractor QC test-based Acceptance decisions toward using agency-tests for the Acceptance decision.

3.3.2 ODOT Case Studies

The NCHRP study referenced above indicated a significant trend among contractor-based tests versus agency-based tests—that contractor-based results tend to be biased toward favorable acceptance decisions. The NCHRP conclusions were based on a large study of many years' worth of longitudinal data. To test whether these conclusions may apply to ODOT, a limited analysis of two instances of QA project data was undertaken in this study. The first instance was data from a project for which ODOT personnel did not suspect significant differences between QC and Verification results; the second case was the converse—ODOT personnel suspected that a latent significant difference did, in fact, exist. However, using current one-to-one tolerance verification methods, in both cases any observed contractor/agency differences were dealt with as the project moved forward, and contractor QC data was the basis for Acceptance and pay factor determination. Clearly, these two cases represent anecdotal and opportunistic information that does not in itself indicate a statistically-strong conclusion; but the purpose of investigating these two instances was to provide some insight into the value of use of F&t-test methods.

3.3.2.1 Oregon 18: Oregon Coast Highway to Oldsville Road Section

The first project examined was on Oregon Highway 18, running from the Oregon Coast Highway (Hwy 101) to Oldsville Road, near McMinnville, OR, approximately 40 miles in length. The project consisted of structures and paving, performed by Mainline Paving, LLC, under the ODOT direction of project manager Shane Ottosen. The total project value was approximately $15.8 million, and was completed in October 2008.

This preservation paving project included improvements to the guardrail, signing, and other safety improvements. In addition, the contractor replaced many of the bridge decks along this route and updated the bridge railing where necessary.

Interviews with ODOT personnel indicated the following:

> The first season of paving on the Oregon Coast Hwy-Oldsville Rd project showed more variability on the percent passing the gravel screens than the second season paving. Further, a problem with the running average of four on the volumetric tests showed up at the beginning of the first season's paving. During the second summer's paving, the variability of the material in the coarser aggregate screens ceased. Speculatively, this change may have resulted from a change the contractor made between the first and second season. In the first season, Oregon Mainline used a parallel flow drum plant. In the second season, the contractor brought in a counter flow drum plant. Those working at the batch plants were excited when the parallel drum plant left and the counter flow drum plant arrived; it appeared that the counter flow drum plant could possibly produce a more consistent mixture.

> An additional factor noted on the project was related to the use of recycled asphalt pavement (RAP). The material that returned to the mixture in RAP consisted of dense graded mix, open graded mix, and a fine mix, which ODOT maintenance often uses to extend pavement life. This could have caused more of a problem for the parallel flow plant then the counter flow plant.

ODOT personnel suspected that undiscovered quality problems existed in this project.

QA data resulting from that project is attached in Appendix E.

3.3.2.2 Oregon 18: Fort Hill to Wallace Bridge Section

The second project examined was also on Oregon Highway 18, running from Fort Hill to the Wallace Bridge, approximately 2.7 miles in length. This project widened the highway, including adding a concrete median barrier and a local access road. The total project value was approximately $15 million, and was completed in November 2009.

This project included grading, drainage, bridge construction, paving, signing, illumination, and roadside development work, including a new interchange with a bridge over Oregon 18 and access to Fort Hill and South Yamhill River Roads.

Interviews with ODOT personnel indicated the following:

> This project was conducted by the same contractor as in the first case study, Oregon Mainline. Paving for the first case study spanned two seasons, while this project began in the second season, benefitting from the plant changes made as a result of the first season experience on the first case study.

This project exhibited fewer quality control issues than the first project, from the perspective of ODOT construction personnel.

QA data resulting from that project is attached in Appendix E.

3.3.2.3 Case study evaluations

Each set of project data was evaluated using existing Independent Assurance and Verification methods (see Figure 3.2) and using F-test and t-test methods. The purpose of this evaluation was to gauge whether the latter, statistically-based methods may provide differing views of the project test data, especially related to the Independent Assurance, Verification, and Acceptance decisions.

The case study analysis focused on HMAC, and on those constituents which comprised pay factors and for which split samples were taken. Further, for simplification, three constituents that are heavily weighted in the pay factor calculation—#200 gradation, asphalt content, and compaction, for a total weight of 76%—are presented here in the most detail. See Table 3.10 for constituent pay factor weights.

Table 3.10: Case study constituent pay factor weights

Constituent	Weight	Constituent	Weight
¾" gradation	1%	#200 gradation	10%
½" gradation	1%	Asphalt content	26%
#4 gradation	5%	Moisture	8%
#8 gradation	6%	Compaction	40%
#30 gradation	3%		

Initially, control charts of test results for the quality control samples and split samples (both contractor and ODOT results) were plotted. The control charts are shown in Figure 3.3.

The charts in Figure 3.3 contain five types of data:

- Upper and Lower specification limits, shown as the bold horizontal lines in each chart;
- IA/Verification data, shown as two coincident points—the ODOT and contractor test results; ODOT results are shown as triangles (labeled IAO), and contractor results are squares (labeled IAC); and
- Contractor QC data, shown as the varying line (labeled QC).

The horizontal axis on the charts indicates successive test results in chronological order, and the vertical axis represents the measure of the test.

By observation, the #200 gradation control charts indicate that results of QC and all split sample tests fall within the upper (USL) and lower (LSL) specification limits, with the exception of one QC test late in the Ft. Hill/Wallace Br project. For the Coast Hwy/Oldsville project, the range of variability of the QC results is smaller than that for the Ft Hill/Wallace BR project. For split sample results, there is a strong tendency for ODOT's test results (the "IAO" triangles) to be higher than the contractor's split sample test results (shown as "IAC" squares). For the Coast Hwy/Oldsville project, ODOT's split sample results appear to be generally higher than the QC results, while this pattern is not evident for the Ft Hill/Wallace Br project.

The asphalt content control charts indicate that the results for QC and split sample tests fall within the specification limits (USL and LSL), save for one exception for each project; in both cases the exception was an ODOT split sample result that fell below the lower specification limit. The range of variability of QC results appears to be roughly equivalent in both cases. Except in one case (Coast Hwy/Oldsville), ODOT's split sample results are lower than those of the contractor's test results.

The compaction control charts indicate QC results that fall entirely within specification limits, albeit generally closer to the lower specification limit (LSL) than to the upper specification limit (USL), especially in the Ft. Hill/Wallace Br case. Split samples are not taken in the case of compaction, so only ODOT test results are shown. By observation, the ODOT test results generally fall within the range of the QC results. No pattern of concern for ODOT versus QC test results is noted.

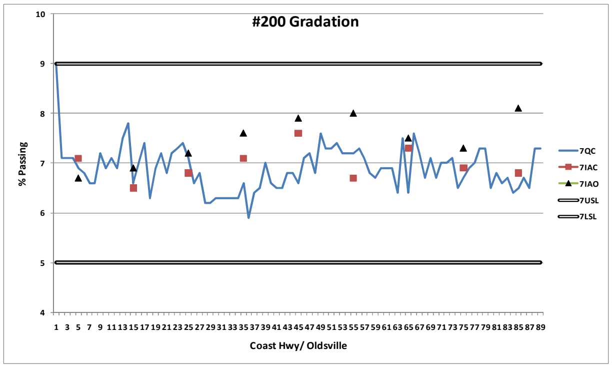
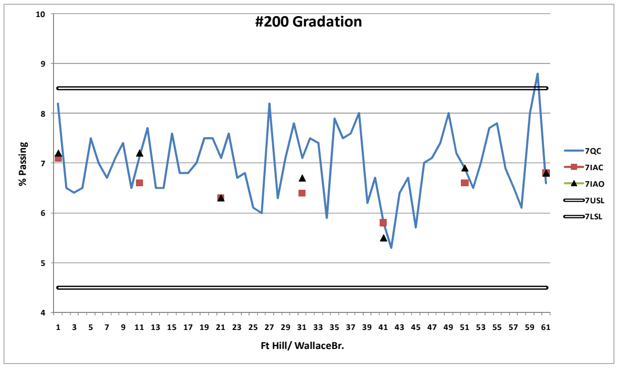
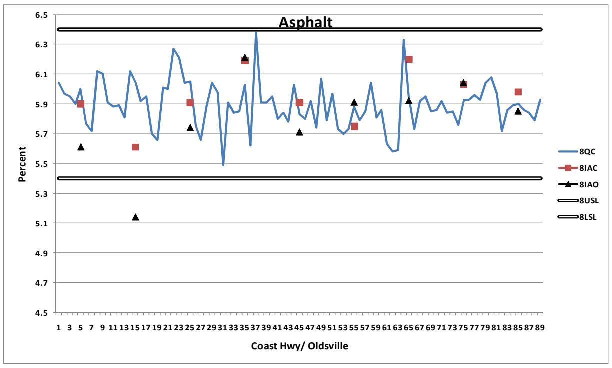
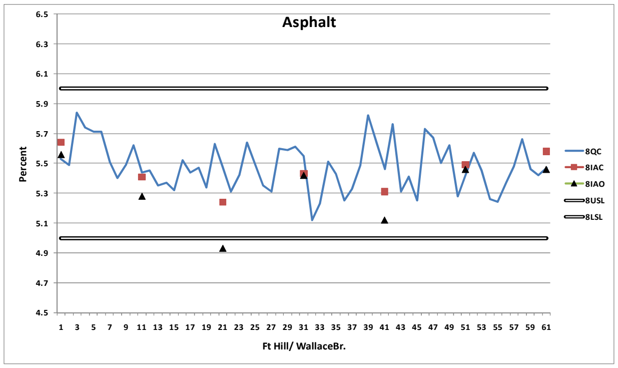
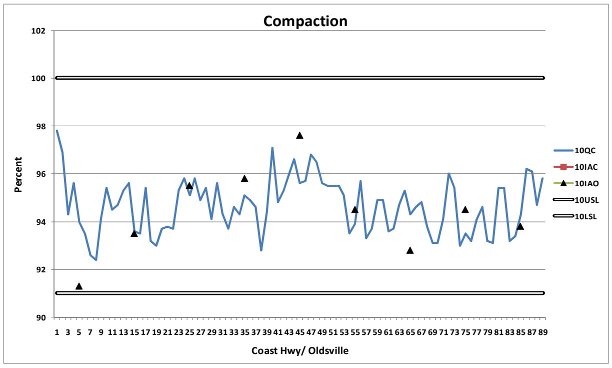
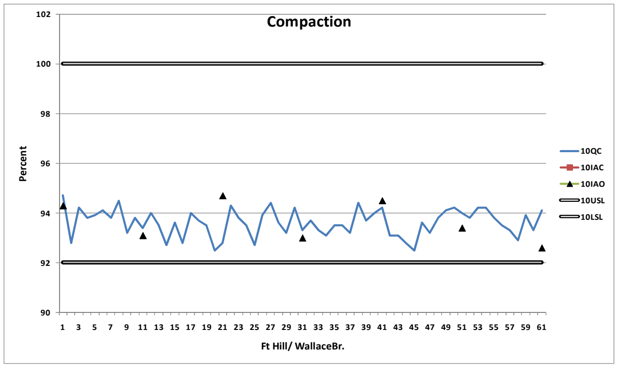
OR18: Oregon Coast Hwy/Oldsville – a paving project that reported significant issues relating to non-specification or quality of the mix.	**OR18: Ft Hill/Wallace Bridge** – a paving project that reported no significant issues relating to non-specification or quality of the mix.
	
	
	

Figure 3.3: Case study key constituents control charts

The #200 gradation results and asphalt content results appear to show patterns of difference among the QC, ODOT split sample, and contractor split sample results that bears further investigation; however, no such observed patterns are seen in the compaction control chart. Since compaction is measured and corrected in real time as a result of independent ODOT field testing, the consistency shown in the compaction control charts is not unexpected. The remainder of this case study analysis will, therefore, focus on the #200 gradation and asphalt content measures.

ODOT's process for Independent Assurance and Verification is shown in Figure 3.2. This process includes two primary evaluations. First, a comparison (IA) occurs between split sample test results. The difference between these test results should be within the tolerance that is established by ODOT (see Table 3.1). For #200 gradation the acceptable difference is within +/- 1%; and for asphalt content the acceptable difference is +/- 0.4%. The second evaluation (Verification) is to compare the split sample test results against the upper and lower specification limits. The split sample results are expected to fall within the lower specification limit and the upper specification limit.

Data for IA and Verification evaluations from the two case study projects was developed in tabular form for the two focal constituents noted above. The results are shown in Table 3.11 and Table 3.12.

The Coast Hwy/Oldsville IA and Verification evaluations indicate mixed results (Table 3.11). For IA—comparison between ODOT's and the contractor's split sample result—largely acceptable results among the various constituents is seen, ranging from 0% to 22% failure. Some failure (among all split samples) was indicated for the two focus constituents of #200 gradation (11% failure) and asphalt content (22% failure). It's worthy of note that in both of the latter two cases, the variance was largely one-sided (lower contractor values for #200 gradation and higher contractor values for asphalt content).

For Verification—comparison of split sample results to the specification limits—contractor results were regularly within specification limits (except for three instances for ½" gradation and one instance for #4 gradation). More frequently, the ODOT split sample results fell outside of specification limits (two instances for ½" gradation, four for #4 gradation, two for #8 gradation, and one for asphalt content).

Table 3.11: Coast Hwy/Oldsville IA and Verification results

OR18: Oregon Coast Hwy/ Oldsville									
			Split sample components						
			Variable Pay Factor Components						
			3/4"	1/2"	#4	#8	#30	#200	Asph
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*fail*	*fail*	*fail*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	fail	pass	pass	pass	fail
Is ODOT result within spec?>>		**Verification**	*pass*	*fail*	*fail*	*fail*	*pass*	*pass*	*fail*
Is contractor result within spec?>>			pass	fail	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	fail	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*fail*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*fail*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	fail	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	fail	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	fail	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		IA	pass	pass	pass	pass	pass	fail	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
		n(samples)=	9	9	9	9	9	9	9
Independent Assurance (IA)>>		n(fail)=	0	0	1	0	0	2	1
		n(fail)%=	0%	0%	11%	0%	0%	22%	11%
Verification>>	*ODOT>>*	*n(fail)=*	*0*	*2*	*4*	*2*	*0*	*0*	*1*
		n(fail)%=	*0%*	*22%*	*44%*	*22%*	*0%*	*0%*	*11%*
	Contractor>>	n(fail)=	0	3	1	0	0	0	0
		n(fail)%=	0%	33%	11%	0%	0%	0%	0%

The Ft Hill/Wallace Br independent assurance and verification evaluations also indicate mixed results (Table 3.12). For IA, comparison between ODOT's and the contractor's split sample result, no failing results were found.

For Verification, comparison of split sample results to the specification limits, contractor results were regularly within specification limits (except for one instance each for #4 and #8 gradations). As with the prior case, the ODOT split sample results more frequently fell outside of specification limits (one instance for #4 gradation, two for #8 gradation, one for #30 gradation, and one for asphalt content).

Table 3.12: Ft Hill/Wallace Br IA and Verification results

OR18: Ft hill-Wallace Bridge									
			Split Sample Components						
			Variable Pay Factor Components						
		Constituent	3/4"	1/2"	#4	#8	#30	#200	Asph
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*fail*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*fail*	*fail*	*fail*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	fail	fail	pass	pass	pass
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
ODOT result close to Contractor result?>>		**IA**	pass	pass	pass	pass	pass	pass	pass
Is ODOT result within spec?>>		**Verification**	*pass*	*pass*	*pass*	*fail*	*pass*	*pass*	*pass*
Is contractor result within spec?>>			pass	pass	pass	pass	pass	pass	pass
		n(samples)=	7	7	7	7	7	7	7
Independent Assurance (IA)>>		n(fail)=	0	0	0	0	0	0	0
		n(fail)%=	0%	0%	0%	0%	0%	0%	0%
Verification>>	*ODOT>>*	*n(fail)=*	*0*	*0*	*1*	*2*	*1*	*0*	*1*
		n(fail)%=	*0%*	*0%*	*14%*	*29%*	*14%*	*0%*	*14%*
	Contractor>>	n(fail)=	0	0	1	1	0	0	0
		n(fail)%=	0%	0%	14%	14%	0%	0%	0%

The IA and Verification evaluations performed in these case studies found a handful of instances where management action was indicated to determine reasons for variances from specified tolerances. Anecdotally, however, there was a lingering concern that undetected quality issues may have been present, especially in the case of the Coast Highway/Oldsville Road project.

3.3.2.4 IA evaluation: Paired t-test

The ODOT split sample one-to-one tolerance evaluation is designed to provide Independent Assurance, on a split sample-by-split sample basis, that the contractor's testing results are reliable. These differences are easily discovered as shown above, leading ODOT personnel to investigate the cause of any out of tolerance differences. An additional evaluation is available to test whether such differences form a pattern across a number of split sample tests—perhaps indicating a consistent bias even if, as a group, they fall within the tolerances. The evaluation used to detect this bias is the paired t-test.

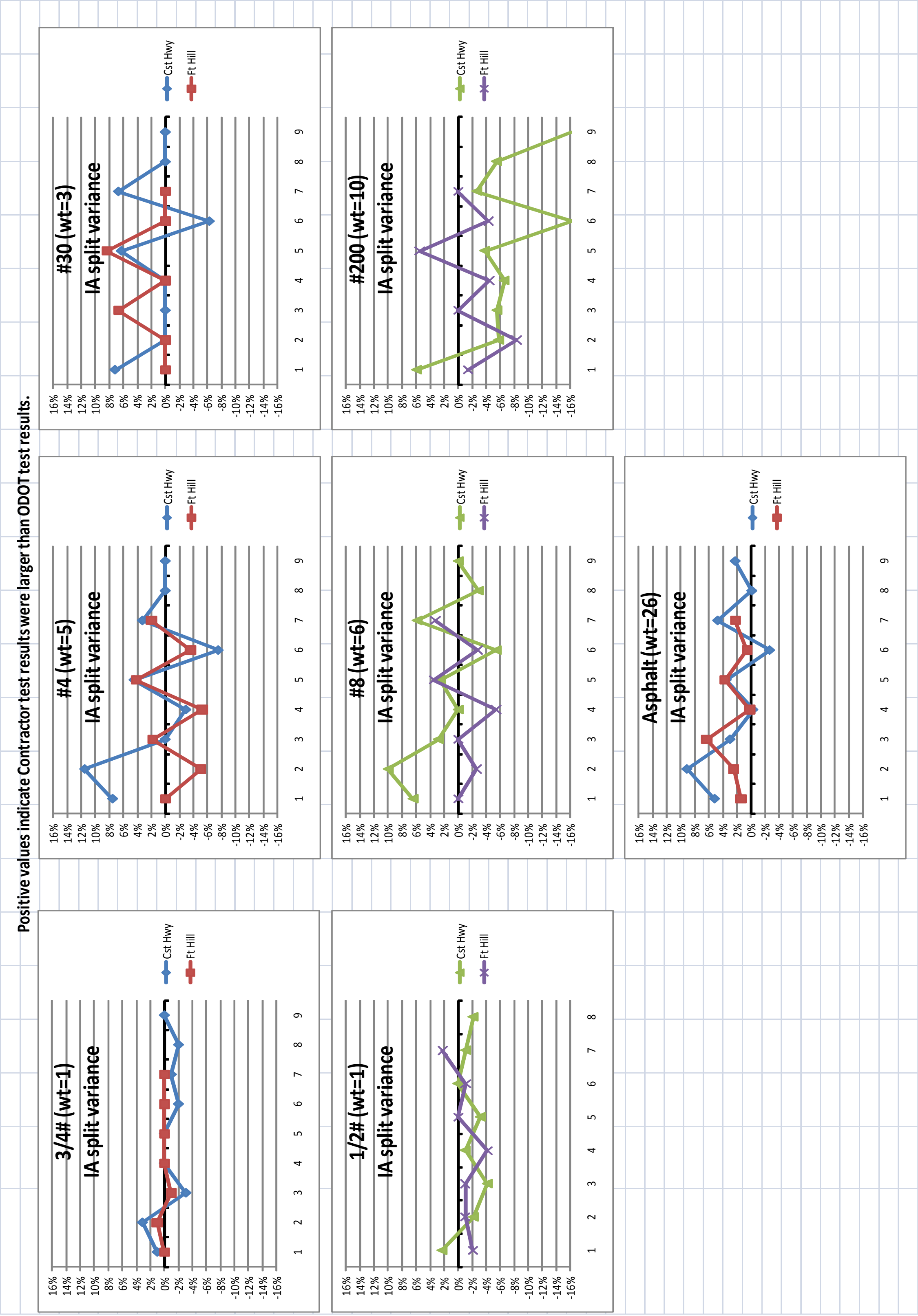


Figure 3.4: Split sample results variances

The paired t-test was conducted on the data for the two case study projects. To help visualize the comparison among the various split sample tests, Figure 3.4 was developed. This figure plots the percentage difference between ODOT versus contractor split sample results for the various pay factor constituents that had split sample results. Ideally, any such differences should be close to zero and should be evenly distributed above and below zero when not zero. Variances that are consistently above or below zero indicate a bias; these patterns of difference are shown in Figure 3.4.

The paired t-test was applied to the split sample results. The purpose of this test is to determine if there is a pattern of differences that would not be consistent with random instances of differences; if so, a systematic bias may be indicated. This test may be developed at varying levels of confidence; Table 3.13 indicates the test conducted at the levels of Alpha=0.01, 0.05, and 0.10—indicating the likelihood (1%, 5%, 10%) that a false indication of a failure may result. For example, at an Alpha of 0.05, there is only a 5% likelihood that a conclusion of bias may be incorrect, or conversely a 95% likelihood that the conclusion of systematic bias is correct.

Table 3.13: Case study IA paired t-test results

Coast Highway/Oldsville							
	3/4"	1/2"	#4	#8	#30	#200	Asph
t stat=	0.736	0.730	0.809	1.265	1.000	2.766	2.352
D.F.=n-1=	8						
tcrit (alpha=0.01)=	3.355						
OK?	pass	pass	pass	pass	pass	pass	pass
tcrit (alpha=0.05)=	2.306						
OK?	pass	pass	pass	pass	pass	fail	fail
tcrit (alpha=0.10)=	1.86						
OK?	pass	pass	pass	pass	pass	fail	fail

Ft Hill/Wallace Br.							
	3/4"	1/2"	#4	#8	#30	#200	Asph
t stat=	0.00	1.45	0.73	0.68	1.55	1.31	3.21
D.F.=n-1=	6						
tcrit (alpha=0.01)=	3.707						
OK?	pass	pass	pass	pass	pass	pass	pass
tcrit (alpha=0.05)=	2.447						
OK?	pass	pass	pass	pass	pass	pass	fail
tcrit (alpha=0.10)=	1.943						
OK?	pass	pass	pass	pass	pass	pass	fail

In Table 3.13 the case study split sample results were evaluated using the procedures indicated earlier in the section entitled "Paired t-test." A failure is indicated when the t-statistic ("t stat") for the individual constituent is larger than the t-critical value ("tcrit") for the Alpha level.

The results of the paired t-test analysis on the case study data indicate that a statistically consistent bias was discovered with two constituents that represent a large amount of pay factor weight – #200 gradation and asphalt content. In Table 3.13, three levels of confidence are represented for each project/constituent combination. Where the indication is "pass" (t stat<t crit), the interpretation is that, within the degree of confidence, randomness may be the reason for observed variations.

Where "fail" is indicated, there is strong likelihood of non-random, or systematic, bias. "Fail" indications are shown for #200 gradation for the Coast Hwy/Oldsville project, at the level that there is only a five percent likelihood (Alpha=0.05) that this bias may be incorrectly concluded. Further, both projects exhibit results that indicate that asphalt content is consistently biased higher in the contractor results, at the 95% confidence level (Alpha=0.05).

The biases under analysis with the paired t-test are those that indicate the contractor's testing methods regularly provide results that vary one way or the other from ODOT's results. While such biases may be suspected by observation of graphics, such as those shown in Figure 3.4, the paired t-test provides an objective means for evaluating the degree to which such biases may or may not be random in nature.

3.3.2.5 Verification evaluation: F-test and t-test

Pay factors are calculated from the contractor's quality control (QC) test results. These pay factors depend on the contractor's delivery of the constituent at the specification level, within the upper and lower specification levels. QC testing results that show a statistical pattern of difference from ODOT's testing results may be an indication of a bias that inaccurately represents installed quality and perhaps favors a higher pay factor. If no such pattern is found, then this comparison between the contractor's QC results and ODOT's test results provides a strengthened verification that the QC results reasonably accurately reflect the quality of the installed material.

ODOT's current process of Verification is done on the basis of comparison as to whether both the contractor's split test results and ODOT's split test results fall within the upper and lower specification limits. This comparison will indicate, on a split sample-by-split sample basis, whether there is a problem that should be investigated. An additional series of tests, the t-test and the F-test are designed to indicate whether there is bias across a group of QC and agency split sample results. This bias is tested by examining the average of a group of test results; the t-test provides the comparison between the averages. The manner in which the t-test is applied depends on whether the two sets of data are similar in statistical variance; the F-test provides that initial check.

The two case studies were evaluated using the F-test and t-test. The results of this evaluation for the pay factor constituents that are split-tested are shown in Table 3.14.

Table 3.14: Case study Verification t-test results

Coast Highway/Oldsville							
	3/4"	1/2"	#4	#8	#30	#200	Asph
DF pooled=	96						
Var pooled=	1.037	5.599	9.353	3.121	0.645	0.201	0.030
t stat=	0.631	0.729	1.028	1.251	0.965	3.629	1.676
t crit (alpha=0.01)=	2.686						
OK?	pass	pass	pass	pass	pass	fail	pass
t crit (alpha=0.05)=	2.012						
OK?	pass	pass	pass	pass	pass	fail	pass
t crit (alpha=0.10)=	1.679						
OK?	pass	pass	pass	pass	pass	fail	pass
Ft Hill/Wallace Br							
DF pooled=	66						
Var pooled=	0.347	1.979	4.991	3.462	1.387	0.475	0.027
t stat=	0.179	2.140	0.602	1.312	2.452	1.289	2.468
t crit (alpha=0.01)=	2.664						
OK?	pass	pass	pass	pass	pass	pass	pass
t crit (alpha=0.05)=	2.002						
OK?	pass	fail	pass	pass	fail	pass	fail
t crit (alpha=0.10)=	1.659						
OK?	pass	fail	pass	pass	fail	pass	fail

In Table 3.14, values from the contractor's QC samples were evaluated against the State's split samples using the procedures indicated earlier in the section entitled "T-Test." A failure is indicated when the t-statistic ("t stat") for the individual constituent is larger than the t-critical value ("tcrit" for the Alpha level).

Examination of the Verification t-test results (Table 3.14) indicates a "fail" reading for two key constituents. For the Coast Highway/Oldsville project, the probabilistic testing indicates that there is a 99% (Alpha=0.01) confidence level that the contractor's average #200 gradation QC test results are biased lower than ODOT's test results (contractor mean=6.90%; ODOT mean=7.47%). This finding is complemented and perhaps related to the paired t-test finding. For the Ft Hill/Wallace Br project, the t-test in Table 3.14 indicates that there is a 95% (Alpha=0.05) confidence level that the contractor's average QC asphalt content results are biased higher than ODOT's independent test results (contractor mean=5.48%; ODOT mean=5.32%). The paired t-test also indicated a likely problem with the Ft Hill/Wallace Br asphalt content test results. The reader is referred to Figure 3.3 for a visual confirmation of these results.

3.3.2.6 Case study assessment

While the results of two cases are relatively anecdotal and cannot be widely generalized, the analysis of these two case studies indicate that the use of the paired t-test and the F-test/t-test combinations may provide an objective means to detect and evaluate bias across a group of contractor test results. Note that similar results are also shown for the ½" and #30 gradation constituents. This sort of bias is similar to that reported in the NCHRP study (Parker and Turochy 2006) that evaluated historical trends among contractor QC test results versus agency test results. It is worthy of note that the analyses provided here did not require additional physical testing, although some additional time was involved in structuring the data for the analyses.

3.4 SUMMARY

The Code of Federal Regulations and the FHWA require specific procedures to allow State Highway Agencies (SHAs) to acquire federal aid for state transportation projects. To comply with these requirements, SHA's keep well structured and organized QA programs that have evolved into complex guidelines for material compliance testing and monitoring programs. The procedures followed at ODOT include a structured set of specifications, standards, and QA procedures with which contractors must comply during the design and construction of bridge and highway projects. The system works well in many areas; however QA programs have evolved, and FHWA guidelines now recommend procedures that may add to ODOT's current practices.

HMAC, aggregates, PCC, and earthwork materials have been described in terms of QA requirements. All of these materials exhibit parameters of quality that may vary in their final construction; therefore they are subject to Quality Control evaluation and Acceptance requirements. For this reason, material-specific testing procedures, frequencies, test parameters, and verification guidelines are established in the ODOT QA program documents.

ODOT's procedure has in practice been useful in discovering QC results that differ from IA tests and in discovering testing process problems. Yet, according to the FHWA testimonials, ODOT has not yet adopted a statistical QA procedure for these construction materials. As discussed in the literature review section of this report, a significant number of data points must comprise both the contractor sample and the agency sample in order for an adequate comparison to be made. Therefore, in order to use these FHWA-recommended statistical techniques, ODOT's process of evaluating one-to-one comparisons would need to expand to the analysis of multiple samples for Quality Assurance.

A review of two selected case studies indicates that the practice of using the split-sample one-to-one comparisons expose important differences that cause investigation when contractor results differ significantly from ODOT test results, and when either results deviate from acceptable specification limits. From testimony, these differences caused ODOT to undertake corrective action in the cases illustrated here. An additional evaluation made in this study—application of statistical techniques to compare longitudinal data—showed a systematic and potentially troublesome bias that was not otherwise obvious in the one-to-one split sample comparisons. Statistical analysis provided additional information that in these cases may have been cause for potential added ODOT management action.

While the case studies revealed previously undiscovered systematic biases, it is unclear whether those biases are isolated to these projects or are indicators of a larger bias. Such a bias may occur on a statewide basis or among a grouping of projects. The 2006 NCHRP study indicated that broad statewide biases may exist. If such statewide biases exist within the ODOT system, management action would likely be very different than if testing biases are restricted to isolated projects or contractors. It appears important, therefore, to determine what may be the scope of any bias before instituting policies for reacting to failed statistical tests on a project-by-project basis.

Determining whether larger system bias exists could be accomplished by examining a large set of project data— perhaps data collected over the span of a construction season, using both the split-sample t-test and the F-test/t-test methods,. This data would necessarily be collected from projects large enough to provide a minimum number of split samples for more robust statistical values. The data could be either historical data or data collected over the next available working season. In addition to collecting the test result data, such a study should additionally collect information about potential factors of bias, including such factors as contractor, project type, project size, procurement type, etc., so that correlations may be made between biased versus unbiased projects. The results would be valuable in establishing the scope and causes of bias for establishment of potential management actions.

4.0 PROCESS RECOMMENDATIONS

ODOT's QA process, as discussed earlier, consists of three parts—Quality Control, Independent Assurance, and Verification:

- Quality Control (QC) is the contractor's efforts at process control and other activities to fulfill the contract requirements. Quality Control testing is used for Acceptance and payment if verified by the agency. No major changes to the Quality Control and Acceptance procedures are recommended here.

- Independent Assurance (IA) is testing, using split samples, to confirm that the contractor's technicians are performing test procedures correctly and that all equipment is calibrated correctly. Some change is recommended to this procedure.

- Verification is independent testing conducted by ODOT to validate the contractor's QC test results. This includes sampling and testing of split samples, as well as comparison against specification parameters to provide an indication of the reliability of the contractor's QC results; verified QC results are then used to determine Acceptance payment adjustments. Some change is also recommended to this procedure.

The recommended changes are discussed below and include further study to evaluate the degree of changes which ODOT wishes to adopt. Such changes may include altering the process of project-based QA, as illustrated in Figure 4.1. Utilizing these additional statistically-based analyses, ODOT may increase its confidence that the QC results are valid indicators of the actual quality of the constructed product. This confidence may further increase with increasing numbers of samples.

4.1 SAMPLING

It is important for the integrity of the process that Independent Assurance samples be acquired in such a manner that they are independent of any potential bias that may also be within the contractor's samples. This includes bias that results from the actual method of sampling as well as bias resulting from the physical location of the sampling; this importance is emphasized in 23 CFR 637 and FHWA Technical Advisory 6120.3.

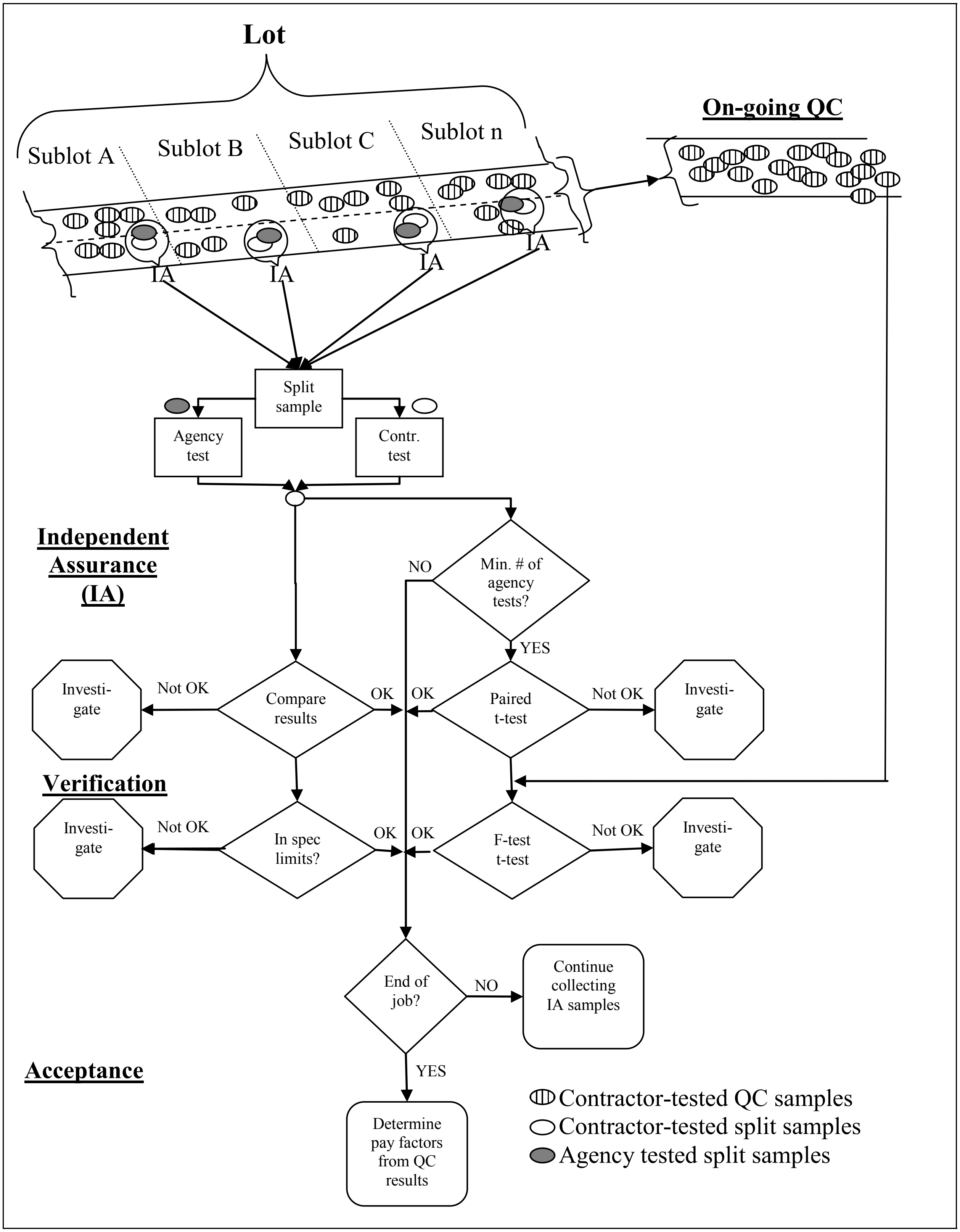


Figure 4.1: Recommended ODOT QA/QC process

In a July 2005 audit of ODOT's Quality Assurance Program, FHWA suggests that "the state needs to increase the use of independent samples that are taken above the number indicated in the current procedure (10%) and provide an independent evaluation of the test data". This independent sampling procedure would involve physical sampling and handling by ODOT or its agent without any contractor involvement, even without any contractor assistance—which is currently the process used at ODOT. This rigorous procedure would increase cost, yet would increase confidence by removing a potential source of bias and would also satisfy FHWA's suggestion on the use of independent samples.

The FHWA guidelines recognize that there may be circumstances, however, where it is impractical, unsafe, etc. to obtain the samples separately from contractor involvement; in those cases the agency may ask for assistance from the contractor in obtaining the samples. In this circumstance, the samples may be used for IA; and Verification may occur from the agency's portion of the same split samples. ODOT indicates that its IA split sample process for HMAC is managed in accordance with the following steps, which comply with the guidelines presented in FHWA Technical Advisory T6120.3 (CFR 1995):

- The IA and Verification sample location or time has been randomly selected by ODOT and is only given to the contractor immediately prior to sampling;
- The contractor's personnel are used only to provide labor to assist in physically obtaining the IA/Verification sample of the material;
- ODOT is present to witness the taking of the IA/Verification sample;
- Both ODOT's technician and the contractor labor are qualified sampling personnel;
- ODOT's technician controls the sampling process by choosing the location or timing and directs the taking of the IA/Verification sample; and
- ODOT's technician immediately takes possession of the IA/Verification sample.

When the split samples are taken under the conditions noted above, then these samples are considered to be independent, and ODOT's split sample results may be used for the F-test and t-test Verification procedures as indicated in Figure 4.1.

FHWA recommends that this contractor-assisted process be conducted on a temporary basis only; this recommendation is likely due to a risk of relaxing control over the long term, causing procedural failure, and potentially leading to contractor bias. Due to this risk, it is the recommendation of the authors of this report that ODOT regularly and systematically ensure that its current contractor-assisted split sampling process continue to comply with the FHWA's guidance. In that circumstance, the ODOT test results, when sufficient samples have accrued, may be used to develop F-test and t-test statistics.

4.2 INDEPENDENT ASSURANCE

ODOT's current IA process was indicated earlier in Figure 3.2, which illustrates the ODOT method for obtaining split samples and using those samples by comparing ODOT versus contractor results on a one-to-one tolerance evaluation basis for Independent Assurance. This method appears to be well-grounded and is well-established within the Agency. Split-sampling methods work well to isolate testing process differences between contractor and Agency testing, as indicated practice and in the literature review. The authors recommend that ODOT consider two options for continuation of its Independent Assurance program:

1) Continuation of the current, project-centric process; or

2) Establish Independent Assurance by using an alternate "system" (time-based) basis, as provided in 23 CFR 207.

Further, if the current project-centric IA process is retained, then for those projects where a minimum number of IA (split) samples have been collected, the authors recommend that a Paired t-test process be used to evaluate the group of ODOT test results against the group of contractor split sample test results as an additional measure of confidence in similarity of results. This procedure will provide insight into systematic testing process variability and may resolve whether individual test conflicts are the result of singular issues or are part of a larger pattern of significant bias; or it may even discover bias when no or few individual test failures are seen.

4.3 VERIFICATION

Verification, as currently practiced by ODOT, is comparison of the agency's and contractor's single split sample test result against the limits specified for the material under consideration. This process has historically been useful in detecting product quality issues as they reveal themselves, allowing for immediate corrective action. As discussed previously, the confidence in validating the contractor's ongoing QC results could be enhanced by utilizing a statistically-based method that compares a group of the agency's test results directly against the contractor's QC test results. Consequently, this section suggests enhancements to ODOT's Verification procedure, adding statistical comparison to ODOT's current Verification process, where sufficient IA samples are available to do so. The intent of these suggested changes is to provide improved confidence that the contractor's QC results are accurate, before application of ODOT's well-established Acceptance procedures. The remaining discussion in this section focuses on the statistical comparisons for Verification, which were adapted largely from *Optimal Acceptance Standards for Statistical Construction Specifications* (Burati, et al. 2003).

4.3.1 Statistical verification procedures

As discussed earlier in this report, it is widely recommended by industry researchers and a policy of the FHWA, that DOTs adopt a program of independent sampling and statistical analysis for verification of QC results that are subsequently used for Acceptance. The authors of this report recommend that split sampling continue to be conducted in a manner that allows for independence and that statistical comparison between the two groups of samples (ODOT's IA results and the contractor's QC results), using the common F-test and t-test methods, be added to the current Verification process. Under the current guidelines for sampling frequency, some projects would have sufficient samples for the F&t-test methods, while others would not. As additional experience is gained with the statistical evaluations, ODOT may consider increased sampling frequency to expand the statistical evaluations to more projects. Such experience would provide sufficient information to evaluate the cost to benefit between quality risk and the cost of additional sampling.

To promote efficiency, ODOT has established a program that utilizes the results from its Independent Assurance tests for its Verification processes. The small sample sizes obtained using the frequencies noted in Table 3.2 through Table 3.5 often include only one Verification sample per project (see Table 3.9, presented earlier). As a practical matter, there are some limitations to this low frequency:

- A singular test result may naturally fall within the extreme upper or lower bounds of the range of the actual material. In such a case, there may be a false indication that the material has failed or is acceptable; and
- Regardless of the result of the test, a single value does not provide any information about variance, which is necessary to describe its statistical properties and to detect systematic test biases. This may mask an important indicator when inferring whether the agency sample may validate the contactor's QC results.

The issue of increasing confidence for the Verification process is largely a matter of independence and sample size. Importantly, as sample sizes (the "*n*" value in the equations shown in the Literature Review section) grow larger, then the likelihood of false rejection and/or false Acceptance decisions becomes rapidly less concerning. Therefore, independent sampling in sufficient quantity for statistical analysis is a means toward improving the confidence of the Verification process.

There are two statistics that are evaluated by these tests—comparison of the mean and comparison of the variance of the respective samples. As discussed earlier in detail in the literature review (Chapter 2) of this report, the t-test compares the means, whereas the F-test compares the variances. In order to develop these statistics, multiple samples must be gathered and tested.

Each test is designed to answer a fundamental question—do ODOT's independent test results, as a group, confirm the contractor's test results? If not, then it is appropriate for investigation to occur as to why the two sets of results differ.

The statistic of considerable interest is derived from the t-test, since it determines whether ODOT's <u>mean</u> test result for a group of samples confirms the contractor's <u>mean</u> QC results—which is then subsequently used for establishing Verification and Acceptance. <u>Variance</u> is evaluated with the F-test, so the F-test is the natural precedent to the t-test. If variance differences are discovered, an alternate t-test calculation must be used. If sample sizes are equal or nearly equal, then variance differences as large as 3-fold do not appear to have substantial effect on the t-test. However, in many situations here, especially on larger projects, there is a significant difference in sample sizes (often 1:10); so to obtain a trustworthy t-test score, the F-test should be used to determine which t-test is appropriate.

There are two questions to be answered regarding sampling size and frequency: 1) At what minimum number of sample sizes should the statistical evaluations be considered effective; and 2) Should split sampling frequency be increased to include additional, smaller projects?

4.3.2 Establishing minimum sample quantities for statistical evaluation

The decision surrounding minimum sample sizes for effectively using statistical evaluations is a subjective management decision that involves comfort level with degrees of certainty. There are two primary levels of uncertainty that may be evaluated, involving potential error in the Verification process. The first of these is termed a Type I error (commonly termed α), and is described as the likelihood of falsely rejecting a valid contractor sample; this is known as the contractor's risk. At an Alpha value of 0.01, this likelihood is characterized as 1%; in other words, there is a 1% chance that acceptable material will be falsely rejected. It is difficult to recommend a specific Alpha value, as no universally accepted value was discovered in the literature review. However, to help put the decision in perspective, the following is a brief description of Alpha values discovered through the literature review.

Burati reports that guidance is provided by AASHTO R-9, as indicated in Table 4.1. Burati notes that: "*While alpha values of 0.10, 0.05, and 0.01 are common, many agencies select a value of 0.01 to minimize the likelihood of incorrectly concluding that the results are different when they actually came from the same population.*" Other DOTs have differing approaches for establishing Alpha values. For example, the Colorado DOT sets Alpha values that vary from 0.001 through 0.050 (specified by project); South Carolina DOT sets an Alpha value of 0.05 for the F-test and 0.01 for the t-test.

Table 4.1: AASHTO R-9 α recommendations

Criticality	Recommended α
Critical	0.050
Major	0.010
Minor	0.005
Contractual	0.001

Critical: essential to preservation of life
Major: necessary for the prevention of substantial financial loss
Minor: does not materially affect performance
Contractual: established only to provide uniform standards for bidding

The choice of an Alpha value is not independent; it is related to a different type of error, called the Beta error. The Beta error is often called the agency risk—it is the likelihood that the agency may accept material that is actually unacceptable. As Burati notes: "*… it should be recognized that selecting a low alpha value reduces the chance of detecting a real difference when one actually exists.*" It was recommended earlier, based on the case study results, that ODOT evaluate a large set of project data to detect whether systematic bias exists for any group of projects. It is additionally recommended here that ODOT develop this explorative testing using Alpha=0.05 for the F-test and using three levels of Alpha for the t-test (Alpha=0.10, 0.05, and 0.01). This is easily accomplished in a simple spreadsheet template using IA and QC data that is already routinely collected by ODOT. It is anticipated that the result of these evaluations will provide guidance to establish an appropriate value of Alpha for institutionalization in formal IA, Verification, and Acceptance procedures.

The discussion below chooses the South Carolina example as a basis for evaluating sample sizes in this report, setting Alpha equal to 0.01 for the t-test and to 0.05 for the F-test, which is particularly subject to low levels of Beta confidence at small sample sizes. For the purposes of the remaining discussion in this chapter, these Alpha values are assumed. Note that in evaluation of the two cases studied in the previous chapter, systematic bias was most evident at an Alpha value of 0.05 for the t-test; therefore, it is recommended that any policy adopted by ODOT should consider a construction season's worth of data, which should make apparent the implications of choice of Alpha values.

Armed with an appropriate Alpha value, it is possible to calculate the confidence of the F-test and t-test (the Beta values) in detecting important differences between quality levels indicated by the QC tests and actual quality levels of the in-place material. These likelihoods are developed using what are termed "Operating Characteristic" (or OC) curves. The number of samples (both ODOT and QC) significantly affect this confidence level—the higher the number of samples, the greater the confidence in the result. To provide guidance, Table 4.2 and Table 4.3 have been prepared to indicate some selected combinations of sample sizes versus Verification confidence.

Table 4.2: Sample size versus likelihood of detecting a three-fold difference in standard deviations (F-test)

Ratio of variances, λ = 3; α = 0.05				Number of Contractor Samples								
Number of Agency Samples	3	4	5	7	10	15	20	30	40	50	75	100
3	19%	19%	20%	20%	20%	21%	22%	23%	24%	24%	24%	24%
4	31%	34%	35%	37%	38%	40%	41%	42%	42%	42%	42%	42%
5	40%	44%	48%	52%	55%	58%	59%	61%	62%	62%	63%	63%
6	45%	52%	56%	62%	67%	70%	71%	71%	71%	71%	72%	72%
7	49%	57%	62%	70%	76%	81%	83%	84%	85%	85%	86%	86%
8	52%	61%	67%	75%	81%	87%	89%	91%	92%	92%	93%	93%
9	54%	63%	70%	78%	85%	91%	93%	95%	96%	96%	97%	97%
10	55%	65%	72%	80%	88%	93%	95%	97%	98%	98%	99%	99%
20	-	-	81%	88%	96%	99%	100%	100%	100%	100%	100%	100%

Source: Burati, Weed, et al.; Optimal procedures for Quality Assurance Specifications, Tables 37 & 39

Table 4.2 refers to the likelihood of detecting, when using the F-test method, an actual three-fold difference in standard deviations between the contractor's QC test result and the Verification test result (this ratio is referred to as λ). This likelihood is a measure of confidence of the F-test itself; smaller levels of confidence result when smaller sample sizes exist. As may be seen from Table 4.2, confidence levels increase significantly as sample sizes increase. For example, at a combination of 6 agency samples and 50 contractor samples, the confidence would be approximately 71%.

Table 4.3: Sample size versus likelihood of detecting a two-standard deviation difference in means (t-test)

SD diff. in means, α = 0.01			d =	2.0	Number of Contractor Samples							
Number of Agency Samples	3	4	5	7	10	15	20	30	40	50	75	100
1	5%	8%	10%	14%	18%	25%	25%	25%	25%	27%	31%	31%
2	15%	17%	21%	29%	35%	51%	51%	51%	53%	58%	62%	68%
3	17%	24%	30%	44%	57%	62%	69%	73%	80%	80%	90%	90%
4	24%	35%	38%	54%	71%	80%	82%	86%	87%	89%	91%	95%
5	32%	40%	52%	63%	79%	88%	90%	93%	94%	95%	96%	98%
6	35%	51%	68%	76%	82%	91%	95%	97%	96%	99%	100%	100%
7	45%	55%	70%	80%	87%	95%	97%	99%	100%	100%	100%	100%
8	48%	58%	68%	86%	90%	95%	100%	100%	100%	100%	100%	100%
9	54%	66%	76%	84%	93%	97%	100%	100%	100%	100%	100%	100%
10	62%	72%	85%	89%	96%	99%	100%	100%	100%	100%	100%	100%
20	68%	86%	90%	99%	100%	100%	100%	100%	100%	100%	100%	100%

Note: values indicated are interpolated using straight-line methods, and are therefore approximate only.

Source: Burati, Weed, et al.; Optimal procedures for Quality Assurance Specifications, Figure 52

Table 4.3 exposes the possibility of detecting a difference in the mean test statistic (estimated by the mean of the sample) of two standard deviations, when using the t-test. For example, if there are four agency samples and ten contractor samples, then there is a 71% chance that a difference between the QC "population" mean and the reported test results will be detected by the t-test. Similar to the F-test, the higher the number of samples, the stronger the confidence that the t-test can identify an actual, material difference. Continuing the earlier example, for a combination of 6 agency samples and 50 contractor samples, the confidence would be approximately 99%.

These tables are provided to assist with the decision regarding numbers of Verification samples to be obtained for a given project. The t-test identifies real differences in the test statistic that forms the basis of the payment decision; thus high certainty about this statistic is desirable. Therefore, it is recommended that selection of minimum samples be guided by consideration of Table 4.3. For the proposed testing program to evaluate the breadth of systematic bias, a conclusive (95% confidence) minimum number of samples appears to be in the range of five or more agency (IA) samples, corresponding to fifty or more contractor QC samples at the current 10% Verification sampling rate. The level of confidence is, of course, ultimately a management risk tolerance decision.

4.3.3 Sampling frequency

Sampling frequency is a policy decision; currently, the default sampling frequency is 1 split sample per every 10 QC samples. It is necessarily a compromise between available resources and the statistical reliability of the comparison. The factor in question is the number of samples that are used for the comparisons—the larger the number of samples, the greater the ability of the comparison to identify substantive differences between the Verification results and the QC results. It is the quantity of split samples that represents the smallest number for the statistical evaluations; therefore increased split sample frequency would result in larger sample sizes and better statistical inference for a typical project and may allow smaller projects to reach minimum sample sizes for effective use of statistical evaluations.

ODOT's current standard 10% rate of sampling produces sample quantities as indicated in Table 3.9, a review of that table shows that until projects reach the "Large Project" category, the number of agency samples generally remains at very low numbers, often at only one sample. The number of samples taken for IA and Verification is a function of two factors: the overall number of QC samples, and the ratio of Verification samples to QC samples. As may be seen from Table 3.9, the variance of material quantity between small, medium and large projects creates a wide variance in the number of both QC and Verification samples. If one is interested in the confidence of the t-test to detect an actual difference of, say, two standard deviations (SDs) between the QC mean results and the Verification mean results, and given the sampling quantities of Table 3.9, then the resulting range of confidence in the t-test is summarized in Table 4.4.

Table 4.4: Current scheme: 10% Verification sampling--likelihood of detecting a two-standard deviation difference in means

SD diff. in means, d=2.0; α = 0.01	<<<<<Small Project>>>>>		<<<<Medium Project>>>>		<<<<<Large Project>>>>>	
Description	Low range	High range	Low range	High range	Low range	High range
HMAC (mixture)	5%	11%	11%	73%	73%	100%
HMAC (compaction)	5%	11%	11%	73%	73%	100%
Aggregate (mixture)	Very low	Very low	Very low	45%	45%	95%
Aggregate (compaction)	Very low	Very low	Very low	45%	45%	95%
Embankment (compaction)	Very low	Very low	Very low	48%	49%	100%
Portland Cement Concrete	Very low	Very low	Very low	10%	10%	100%

Table 4.4 illustrates that the sample quantities at a 10% Verification sampling rate do not produce high levels of t-test confidence for small and many medium-level projects. For example, there is a 45% likelihood (less than a 50/50 chance) that the t-test would detect an actual two standard deviation difference in mean values for aggregate compaction for the high range of a medium-sized project. Darker shading indicates increasing levels of confidence.

Alternate schemes for numbers of Verification samples may be explored to provide some sense of the effect of increases in Verification samples (and in some cases, QC samples). Two alternate schemes are presented here, primarily for the purpose of providing insight into the effect of requirement of additional Verification tests.

The first scheme (Scheme A) increases the minimum number of Verification tests to three, and imposes the same minimum on the contractor's QC tests. The result is shown in Table 4.5; the underlined, italicized cells indicate those values that have been changed from the values presented previously in TableTable 3.9.

Table 4.5: Scheme A: Minimum of 3 Verification samples, except for very small projects

	Small Project				Medium Project				Large Project			
	Low		High		Low		High		Low		High	
Description	QA	QC	QA	QC	QA	QC	QA	QC	QA	QC	QA	QC
HMAC (mixture)	*3*	3	*3*	6	*3*	6	3	30	3	30	20	200
HMAC (compaction)	*3*	3	*3*	6	*3*	6	3	30	3	30	20	200
Aggregate (mixture)	*3*	*3*	*3*	*3*	*3*	*3*	*3*	13	*3*	13	5	50
Aggregate (compaction)	*3*	*3*	*3*	*3*	*3*	*3*	*3*	13	*3*	13	5	50
Embankment (compaction)	*3*	*3*	*3*	*3*	*3*	*3*	*3*	14	*3*	14	7	67
Portland Cement Concrete	*3*	*3*	*3*	*3*	*3*	*3*	*3*	5	*3*	5	10	100

When adjusted to the values of Scheme A (see Table 4.5), the likelihood of the t-test detecting an actual two SD difference increases appreciably from the values shown in Table 4.4. In this scenario, the likelihood of detection increases but does remain below 50% for most measures in small and many medium-sized projects. The underlined, italicized cells in Table 4.6 indicate changes from Table 4.4; darker shading indicates increasing levels of confidence.

Table 4.6: Scheme A: Likelihood of detecting a two-standard deviation difference in means

SD diff. in means, d=2.0; α = 0.01	<<<<<Small Project>>>>>		<<<<Medium Project>>>>		<<<<<Large Project>>>>>	
Description	Low range	High range	Low range	High range	Low range	High range
HMAC (mixture)	*17%*	*36%*	*36%*	73%	73%	100%
HMAC (compaction)	*17%*	*36%*	*36%*	73%	73%	100%
Aggregate (mixture)	*17%*	*17%*	*17%*	*60%*	*60%*	95%
Aggregate (compaction)	*17%*	*17%*	*17%*	*60%*	*60%*	95%
Embankment (compaction)	*17%*	*17%*	*17%*	*61%*	*61%*	100%
Portland Cement Concrete	*17%*	*17%*	*17%*	*30%*	*30%*	100%

The second scheme (Scheme B) further increases the number of Verification tests to a minimum of six and imposes the same minimum on the contractor's QC tests. The result is shown in Table 4.7; the italicized, underlined cells indicate those values that have been changed from the values presented previously in Table 3.9.

Table 4.7: Scheme B: Minimum of 6 Verification samples, except for very small projects

	Small Project				Medium Project				Large Project			
	Low		High		Low		High		Low		High	
Description	QA	QC	QA	QC	QA	QC	QA	QC	QA	QC	QA	QC
HMAC (mixture)	*6*	*6*	*6*	6	*6*	6	*6*	30	*6*	30	20	200
HMAC (compaction)	*6*	*6*	*6*	6	*6*	6	*6*	30	*6*	30	20	200
Aggregate (mixture)	*6*	*6*	*6*	*6*	*6*	*6*	*6*	13	*6*	13	*6*	50
Aggregate (compaction)	*6*	*6*	*6*	*6*	*6*	*6*	*6*	13	*6*	13	*6*	50
Embankment (compaction)	*6*	*6*	*6*	*6*	*6*	*6*	*6*	14	*6*	14	7	67
Portland Cement Concrete	*6*	*6*	*6*	*6*	*6*	*6*	*6*	*6*	*6*	*6*	10	100

As shown in Table 4.8, when adjusted to the values of Scheme B (Table 4.7), the likelihood of the t-test detecting an actual two SD difference increases even further from the values shown in Table 4.4. The underlined, italicized cells in Table 4.8 indicate changes from Table 4.4; darker shading indicates increasing levels of confidence. In this scenario, the likelihood of detection is increased above 50% for all projects.

Table 4.8: Scheme B: Likelihood of detecting a two standard deviation difference in means

SD diff. in means, d=2.0; α = 0.01	<<<<<Small Project>>>>>		<<<<Medium Project>>>>		<<<<<Large Project>>>>>	
Description	**Low range**	**High range**	**Low range**	**High range**	**Low range**	**High range**
HMAC (mixture)	*72%*	*72%*	*72%*	*97%*	*97%*	100%
HMAC (compaction)	*72%*	*72%*	*72%*	*97%*	*97%*	100%
Aggregate (mixture)	*72%*	*72%*	*72%*	*85%*	*85%*	*99%*
Aggregate (compaction)	*72%*	*72%*	*72%*	*85%*	*85%*	*99%*
Embankment (compaction)	*72%*	*72%*	*72%*	*88%*	*88%*	100%
Portland Cement Concrete	*72%*	*72%*	*72%*	*72%*	*72%*	100%

As mentioned earlier, the decision to change either the Verification or QC sampling frequency is a matter of management decision; important factors include balancing between agency resources and risk tolerance. For large projects, a general rule of Verification testing at the level of 10% of QC sampling may provide a sufficiently low level of risk for the agency. For small- and medium-level projects, however, it may be worthwhile to establish a minimum number of Verification and QC tests to provide sufficiently low levels of risk, as suggested by the schemes and the tables above. For the very smallest projects a table similar to Table 3.6: Very small project quantity table – may be established for which alternate means of verification are applicable, as is the current policy.

The case studies performed earlier indicate the presence of systematic bias in the projects studied, but it is unclear whether that systematic bias would be found to exist across only certain projects, or across a group of projects, or across most projects within the state, etc; and if bias is found, to what level. Therefore, before adding the cost of increasing sample sizes for smaller projects (except for the very small projects noted in Table 3.6: Very small project quantity table), it appears prudent to determine the scope of systematic bias. The minimum number of samples used for statistical comparisons should be no less than three (the minimum required for establishing variance), but it likely need not be higher than six. As an example, California adopts a minimum value of three IA/Verification tests before statistical analysis is triggered (see Appendix F). In review of the tables above and to develop strong levels of confidence, it is recommended that a policy be established that projects acquire at least five split samples before the statistical comparison process begins, and that new statistical results are calculated, reported, and tracked against the ongoing QC results to establish trends across the duration of the project as additional IA/Verification results are obtained. Such trends would provide valuable insight into the stability of the QC process and the reliability of its data; and tracking the statistical results as the project unfolds would provide a basis for course correction before project completion.

4.4 VERIFICATION TESTING SPECIFICATIONS

No universally accepted model or standard specification for Verification testing was discovered in this project. It is anticipated that such a procedure should discuss the fundamental purpose of such testing, the basic principles behind development of independent verification sampling, the steps necessary for the statistical F-test and t-test, and the actions to take when such tests do not confirm the contractor's QC test results. For information purposes, Appendix G and Appendix H are examples of procedures utilized by the South Carolina and Colorado DOTs, respectively. These models may be easily adapted into the ODOT MFTP when decisions regarding testing levels have been finalized.

4.5 ELECTRONIC METHODS OF QA DATA ANALYSIS

As is discussed in this report, it is important for increased levels of reliability to use statistical procedures for comparison between agency and contractor test results. The most straightforward techniques for doing so involve the use of the paired t-test for Independent Assurance and the F-test (to test variances) and the t-test (to test means) for Verification. The statistical procedures of the F-test and t-tests appear complicated and may be intimidating if they are developed by hand. Institution of a policy to use these tests, then, should be accompanied by the provision of appropriate software that eases the calculations.

4.5.1 Overview

The FHWA has produced a report, "Quality Assurance Software for the Personal Computer" (Weed 1981). That report provides the reader with a comprehensive overview of statistical procedures for QA/QC and describes a software system for PCs that can be used to analyze all types of QA/QC data. The report outlines a variety of software that is used for simple and sophisticated QA/QC management, including statistical analyses and determination of pay factors for various material standards. Although somewhat dated, it may be useful in reviewing the options available for electronic QA/QC analysis.

West Virginia, Colorado, and Florida State Departments of Transportation were identified with well-developed QA/QC processes and data analysis software. It appears, however, that much of the software was developed primarily to support the Acceptance decision, rather than the Verification process.

Although some states have developed custom programs for the F-test and t-test, modern spreadsheet programs such as Microsoft Excel include built-in functions to easily calculate the results. In 2007 a survey of DOTs was conducted to determine what types of QA/QC software was in use (see Appendix I). Review of the survey results indicates that much of the software is focused on the Acceptance decision or on process control (QC). Further, many of the programs noted were developed in-house.

4.5.2 Electronic QA analysis at ODOT

ODOT has developed a time-proven Excel-based QA/QC software product, STATSPEC, which is used for determining pay factors for HMAC materials based on contractor QC test results. This package models the formulas and factors contained within ODOT's Manual of Field Test Procedures, Standard Specifications, and other guidance documents. The authors of this report recommend that an expansion of the STATSPEC product be developed for modeling the Verification statistics methodology proposed in this study.

Inputs required for the statistical comparisons include split sample test results (both contractor and ODOT results) and QC test results. QC results are already collected in STATSPEC; the program would require only minimal changes to additionally record the split sample test results. These results may be easily analyzed for the paired t-test and the F-test/t-test using the procedures outlined in this report, or using built-in Excel functions.

4.6 MATERIALS TO CONSIDER FOR IMPROVED VERIFICATION PROCEDURES

This study has explored the QA process for four primary materials, and these are simplistically shown in Table 4.9. All of these materials explored here are important structural components of the work; and all are subject to immediate adjustment when a test result indicates that the material is out of specification compliance. Some materials, however, are also subject to varying degrees of allowable specification compliance; and further, this variant compliance may additionally be used to adjust payment for the in-place material. Table 4.9 indicates which among these materials are subject to variable pay adjustments. The test results for those materials that are subject to variable pay adjustments are especially important in regards to this study, as the ongoing QC results form the basis for the pay adjustments; therefore considerable economic pressure is placed on both the contractor and ODOT to ensure that the QC results are reliable. Developing a sound, objective basis for the payment decision will help reduce bias in the process.

It is the recommendation of the authors of this report that the Verification procedures recommended herein should initially be primarily focused on HMAC mixture and compaction, and that implementation could then move forward to aggregate base and subbase. Calculation of the F-test and t-test statistics may uncover a systematic bias in the contractor's sampling program that may otherwise go undetected. Soil compaction and PCC strength are tests that result in threshold (go/no go) determinations of adequacy; these are subject to immediate rejection and re-work if found unacceptable; and their results are not averaged over many sublots. As this process is implemented, and as experience with the program grows, it is recommended that other appropriate materials may be included.

Table 4.9: Material category compliance characteristics

Material	Threshold (T) or Variable (V) compliance?	Subject to pay factors?	Results judged on Average of sublots (A) or by Individual sublot (I)?
HMAC (mixture)	V	Y	A
HMAC (Compaction)	V	Y	A
Aggregate (mixture)	V	Y	A
Aggregate (compaction)	T	N	I
Embankment (compaction)	T	N	I
Portland Cement Concrete	T	N	I

4.7 IMPLEMENTATION STEPS

It is suggested that implementation of this process involve the following steps:

1. Determine for which materials/tests and project sizes the changed QA process will apply.
2. Evaluate a significant number of projects (perhaps a construction season's worth) to determine at what level systematic bias may occur. Determine actions if large-scale bias is found.
3. Establish the method for Independent Assurance (project quantity-based or system-wide time-based).
4. Expand policy and specification for Independent Assurance testing to include statistical evaluations.
5. Expand policy and specification for Verification sampling and testing to include statistical evaluations.
6. Expand STATSPEC software for accumulation of IA sample results and statistical calculation of data.
7. Identify a project or projects for testing the new methodology (in parallel with the existing methods).
8. Evaluate results and process; modify draft procedures and specifications.
9. Finalize procedures and specifications and software.
10. Develop training materials.
11. Provide training and publish finalized materials.
12. Implement changes.

4.8 SUMMARY

ODOT's current process of Verification is directly linked to its process of Independent Assurance (IA). These results are used to infer reliability of the contractor's ongoing Quality Control (QC) testing. However, the linkage between the IA testing and Verification introduces several challenges to this inference, including small sample sizes (introducing high likelihood of accepting rejectable material) and lack of evaluation of groups of samples (introducing the likelihood of not detecting systematic sampling bias). Further, FHWA policy indicates that Independent Assurance testing should be separate from Verification testing, except in special circumstances.

The current ODOT procedure is a balance of perceived benefit versus the cost of QA/QC sampling, testing, and reporting. To more fully evaluate this balance under a change in procedure, both additional cost and additional benefit must be more definitively established. There are two levels at which such an evaluation may be made. First is to simply continue to accumulate data (results from split samples and contractor QC tests) under the current process, but to add the statistical tests discussed in this report. The additional work to accomplish this would include a one-time expansion of the STATSPEC worksheet; project-by-project entry of split sample results into the revised worksheet; and evaluation of the results for each project. This cost is expected to be minimal compared to the cost of materials under evaluation; the benefit would be possible detection of issues that could easily return more benefit than the cost. Second, if a study of data across a large group of projects indicates a systematic or frequent bias (which may result in higher pay factors) that would have resulted in payment savings, then these statistics could be used to more definitively calculate benefits. These benefits could be compared against the cost of increased sampling to determine whether such changes are justified.

Several recommendations to reinforce the current process are provided as a result of this study. These include the following:

- Determine the breadth of systematic bias in agency versus QC results through examination of a large grouping of project data, perhaps a construction season’s worth.
- Using insight from the above analysis, fine-tune the recommendations in this report for parameters such as Alpha levels, minimum sample sizes, and sampling frequency for project-level evaluations.
- If the current quantity/project-based approach system of Independent Assurance testing is retained, consider periodically ensuring that the contractor-assisted sampling process is within control, and consider adding a Paired-t testing process to better detect systematic differences among test results.
- A statistically-based Verification evaluation process, using F&t-tests, could be added for those projects with a minimum number of agency samples (perhaps five) also to better detect systematic differences among test results.
- Increased sampling frequency, above the levels currently used for Independent Assurance, should be considered to improve the confidence in the Verification testing process for lower-quantity projects. The minimum required for statistical evaluation would be 3 samples, although the confidence level at that sample size is low.
- These changes should strongly be considered for HMAC paving, and potentially for aggregate base and subbase courses. Other materials may be considered as program experience develops.
- The STATSPEC program is easily altered to acquire split sample test results and to automatically calculate the Paired t-test for Independent Assurance and the F&t-tests for Verification.

- Draft specifications and procedures for a revised process should be tested on a few trial projects, and the process revised as appropriate.
- A training program should be developed for internal personnel prior to institutional implementation of the full system.

It is the opinion of the authors of this report that any of the above steps will strengthen ODOT’s Independent Assurance and Verification processes; certainly, the more of these recommendations implemented, the stronger the program will become.

5.0 CONCLUSION

This study has investigated the national literature regarding suggested and implemented processes for improved certainty in Independent Assurance and Verification testing for the national highways. It was found that there is a substantial and well-considered body of work that explains probabilistically-based analytical tools for evaluating conformance with specified construction standards. This literature indicates that several principles are important foundations in establishing this improved certainty. These principles include separation of the processes of validating contractor testing (Independent Assurance) from the process of validating the contractor's Quality Control results (Verification) and independent acquisition of samples. Simple and well-understood calculations for the statistical F-test and t-test have become accepted tools for the probabilistic calculations. Many states have adopted these techniques.

ODOT currently has a combined process for both Independent Assurance (IA) and Verification. This process is a very efficient process, since it uses the same test results for both Independent Assurance and Verification, and it utilizes the services of the contractor as an aid in the agency sampling process. Based on the result of ODOT's testing, inference is made regarding the validity of the contractor's ongoing Quality Control (QC) test results. Finally, if ODOT is confident in the validity of the QC results, payment is made based on those QC results.

A number of recommendations are the result of this study. Chief among these are the following:

- Ensure that the contractor-assisted sampling process is periodically evaluated to ensure sampling independence.
- Review a large sample of QA results to establish at what level systematic bias may occur within ODOT's projects and as a guide for fine-tuning the statistical evaluation parameters suggested in this report.
- Use existing time-proven models from other states to establish specifications and procedures for enhanced project-based IA and Verification.
- Incorporate statistical evaluations into ODOT's existing STATSPEC program, and use the results for evaluating projects for which a minimum number of agency samples (suggested at five) have been collected.
- Develop this program initially for HMAC and expand as appropriate to other materials.
- Identify a few projects for a trial run of the new procedures.
- Provide training of ODOT staff prior to institution-wide implementation of the new processes.

The authors are confident that the suggested changes will enhance ODOT's current Quality Assurance program.

It should be noted that during this study, a few items were unable to be definitively defined, and will require ODOT to strike a balance between efficiency and certainty and between ODOT risk and contractor risk. Specifically, the decision regarding minimum sampling quantities involves the cost of additional sampling versus the benefit of improved quality. While the cost of added sampling may be simply calculated, the benefit of improved quality is not apparent but may be established by evaluating whether new methods may improve quality delivery, through evaluation of either long-term operating history or historic projects. Recommendations of minimum sampling have been suggested in this report; those recommendations are largely based on anecdotal observations of sampling frequencies adopted by other states, as well as recognition of the mathematical condition that increased sampling increases confidence in the estimate of QC validity. The suggested values are starting points; it is expected that with experience and further study, ODOT will develop a better sense of this trade-off of risk and will make adjustments that fit the risk tolerance of the agency and the contracting community.

6.0 REFERENCES

AASHTO. *Acceptance Sampling Plans for Highway Construction.* Part I: Specifications. American Association of State Highway and Transportation Officials. 1984.

AASHTO. *AASHTO Implementation Manual for Quality Assurance*. American Association of State Highway and Transportation Officials. 1996.

Benson, P. E. Comparison of End Result and Method Specifications for Managing Quality. *Transportation Research Record: Journal of the Transportation Research Board*, 3(10). 1995.

Burati, J. L., R. M. Weed, C. S. Hughes, and H. S. Hill. *Optimal Procedures for Quality Assurance Specifications*. Publication No. FHWA-RD-02-095. Federal Highway Administration. Office of Research, Development, and Technology. 2003.

Burati, J. L., R. M. Weed, C. S. Hughes, and H. S. Hill. *Evaluation of Procedures for Quality Assurance Specifications*. Publication No. FHWA-HRT-04-046. Federal Highway Administration. Office of Research, Development, and Technology. 2004.

Caltrans. *Quality Control and Quality Assurance Manual*. State of California. Department of Transportation. 2002.

Dixon, K. *Basic Quality Assurance and Quality Control*. Project No. E-20-H72. Georgia Institute of Technology. School of Civil and Environmental Engineering. 2003.

CFR. Quality Assurance Procedures for Construction. 23 CFR Part 637. Chapter I - Federal Highway Administration, Department Of Transportation, Part 637_Construction Inspection And Approval. Code of Federal Regulations. U.S. Government Printing Office. 1995.

Gentry, C. and W. A. Yrjanson. Specifications for Quality Control: A Case Study. *Transportation Research Record: Journal of the Transportation Research Board*, 11(26), 37-41. 1987.

Gibria, I. N. Recent Developments in Control Chart Techniques. *Journal of Quality Technology*, 7(4), 183-192. 1975.

Hughes, C. S. *State Construction Quality Assurance Programs*. NCHRP Synthesis 346. National Cooperative Highway Research Program. Transportation Research Board. Washington, DC. 2005.

NCHRP. *Statistically Oriented End-Result Specifications*. NCHRP Synthesis 38. National Cooperative Highway Research Program. Transportation Research Board. Washington, DC. 1976.

ODOT. *Manual of Field Test Procedures*. Oregon Department of Transportation. Highway Division, Construction Section. Salem, OR. 2006.

Parker, F. and R. E. Turochy. *Using the Results of Contractor-Performed Tests in Quality Assurance*. NCHRP Web-Only Document 115. National Cooperative Highway Research Program. Transportation Research Board. Washington, DC. 2006.

Ramsey, F. and D. Schafer. *The Statistical Slueth*. Duxbury. Pacific Grove. 2002.

TRB. *Glossary of Highway Quality Assurance Terms*. Transportation Research Circular, Number E-C074. Transportation Research Board. Washington, DC. 2005.

Weed, R. M. Quality Assurance Software for Personal Computer. *Transportation Research Record: Journal of the Transportation Research Board*, 15(44), 116-124. 1981.

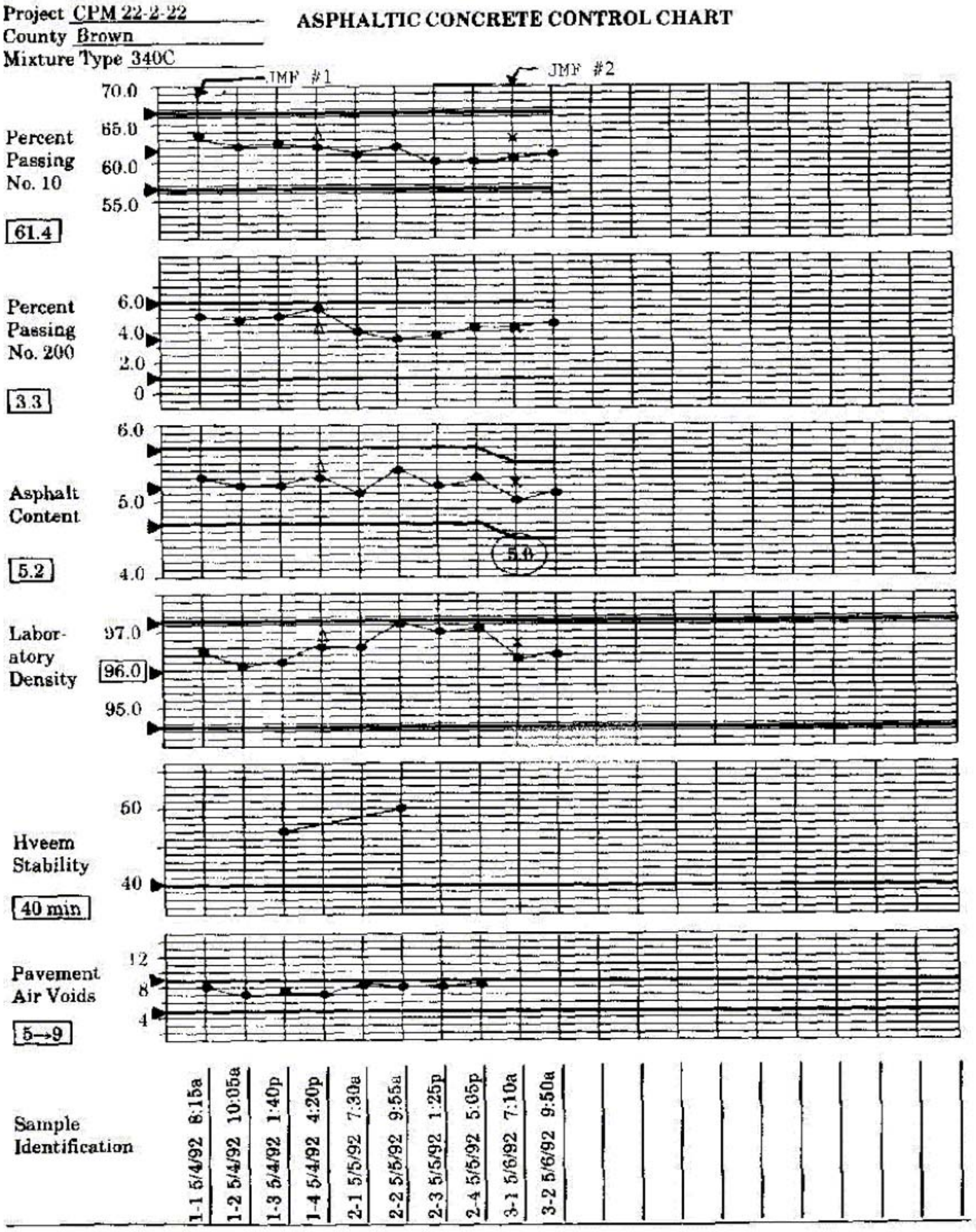
Weed, R. M. Development of Multicharacteristic Acceptance Procedures for Rigid Pavement. *Transportation Research Record: Journal of the Transportation Research Board*, 885, 25-36. 1984.

Weed, R. M. Revision of a Flawed Standard. *Transportation Research Record: Journal of the Transportation Research Board*, 1056, 21-35. 1986.

APPENDICES

APPENDIX A: SAMPLE CONTROL CHART

Control Chart Example (TXDOT 1999)



APPENDIX B: FEDERAL REGISTER QUALITY ASSURANCE PROCEDURES FOR CONSTRUCTION

23 CFR 637
[Code of Federal Regulations]
[Title 23, Volume 1]
[Revised as of April 1, 2001]
From the U.S. Government Printing Office via GPO Access
[CITE: 23CFR637.201]

TITLE 23--HIGHWAYS

CHAPTER I--FEDERAL HIGHWAY ADMINISTRATION, DEPARTMENT OF TRANSPORTATION

PART 637--CONSTRUCTION INSPECTION AND APPROVAL--Table of Contents

Subpart B--Quality Assurance Procedures for Construction

Sec. 637.201 Purpose.

To prescribe policies, procedures, and guidelines to assure the quality of materials and construction in all Federal-aid highway projects on the National Highway System.

Sec. 637.203 Definitions.

Acceptance program. All factors that comprise the State highway agency's (SHA) determination of the quality of the product as specified in the contract requirements. These factors include verification sampling, testing, and inspection and may include results of quality control sampling and testing.

Independent assurance program. Activities that are an unbiased and independent evaluation of all the sampling and testing procedures used in the acceptance program. Test procedures used in the acceptance program which are performed in the SHA's central laboratory would not be covered by an independent assurance program.

Proficiency samples. Homogeneous samples that are distributed and tested by two or more laboratories. The test results are compared to assure that the laboratories are obtaining the same results.

Qualified laboratories. Laboratories that are capable as defined by appropriate programs established by each SHA. As a minimum, the qualification program shall include provisions for checking test equipment and the laboratory shall keep records of calibration checks.

Qualified sampling and testing personnel. Personnel who are capable as defined by appropriate programs established by each SHA.

Quality assurance. All those planned and systematic actions necessary to provide confidence that a product or service will satisfy given requirements for quality.

Quality control. All contractor/vendor operational techniques and activities that are performed or conducted to fulfill the contract requirements.

Random sample. A sample drawn from a lot in which each increment in the lot has an equal probability of being chosen.

Vendor. A supplier of project-produced material that is not the contractor.

Verification sampling and testing. Sampling and testing performed to validate the quality of the product.

Sec. 637.205 Policy.

(a) Quality assurance program. Each SHA shall develop a quality assurance program which will assure that the materials and workmanship incorporated into each Federal-aid highway construction project on the NHS are in conformity with the requirements of the approved plans and specifications, including approved changes. The program must meet the criteria in Sec. 637.207 and be approved by the FHWA.

(b) SHA capabilities. The SHA shall maintain an adequate, qualified staff to administer its quality assurance program. The State shall also maintain a central laboratory. The State's central laboratory shall meet the requirements in Sec. 637.209(a)(2).

(c) Independent assurance program. Independent assurance samples and tests or other procedures shall be performed by qualified sampling and testing personnel employed by the SHA or its designated agent.

(d) Verification sampling and testing. The verification sampling and testing are to be performed by qualified testing personnel employed by the SHA or its designated agent, excluding the contractor and vendor.

(e) Random samples. All samples used for quality control and verification sampling and testing shall be random samples.

Sec. 637.207 Quality assurance program.

(a) Each SHA's quality assurance program shall provide for an acceptance program and an independent assurance (IA) program consisting of the following:

(1) Acceptance program.

(i) Each SHA's acceptance program shall consist of the following:

(A) Frequency guide schedules for verification sampling and testing which will give general guidance to personnel responsible for the program and allow adaptation to specific project conditions and needs.

(B) Identification of the specific location in the construction or production operation at which verification sampling and testing is to be accomplished.

(C) Identification of the specific attributes to be inspected which reflect the quality of the finished product.

(ii) Quality control sampling and testing results may be used as part of the acceptance decision provided that:

(A) The sampling and testing has been performed by qualified laboratories and qualified sampling and testing personnel.

(B) The quality of the material has been validated by the verification sampling and testing. The verification testing shall be performed on samples that are taken independently of the quality control samples.

(C) The quality control sampling and testing is evaluated by an IA program.

(iii) If the results from the quality control sampling and testing are used in the acceptance program, the SHA shall establish a dispute resolution system. The dispute resolution system shall address the resolution of discrepancies occurring between the verification sampling and testing and the quality control sampling and testing. The dispute resolution system may be administered entirely within the SHA.

(2) The IA program shall evaluate the qualified sampling and testing personnel and the testing equipment. The program shall cover sampling procedures, testing procedures, and testing equipment. Each IA program shall include a schedule of frequency for IA evaluation. The schedule may be established based on either a project basis or a system basis. The frequency can be based on either a unit of production or on a unit of time.

(i) The testing equipment shall be evaluated by using one or more of the following: Calibration checks, split samples, or proficiency samples.

(ii) Testing personnel shall be evaluated by observations and split samples or proficiency samples.

(iii) A prompt comparison and documentation shall be made of test results obtained by the tester being evaluated and the IA tester. The SHA shall develop guidelines including tolerance limits for the comparison of test results.

(iv) If the SHA uses the system approach to the IA program, the SHA shall provide an annual report to the FHWA summarizing the results of the IA program.

(3) The preparation of a materials certification, conforming in substance to Appendix A of this subpart, shall be submitted to the FHWA Division Administrator for each construction project which is subject to FHWA construction oversight activities.

(b) [Reserved]

Sec. 637.209 Laboratory and sampling and testing personnel qualifications.

(a) Laboratories.

(1) After June 29, 2000, all contractor, vendor, and SHA testing used in the acceptance decision shall be performed by qualified laboratories.

(2) After June 30, 1997, each SHA shall have its central laboratory accredited by the AASHTO Accreditation Program or a comparable laboratory accreditation program approved by the FHWA.

(3) After June 29, 2000, any non-SHA designated laboratory which performs IA sampling and testing shall be accredited in the testing to be performed by the AASHTO Accreditation Program or a comparable laboratory accreditation program approved by the FHWA.

(4) After June 29, 2000, any non-SHA laboratory that is used in dispute resolution sampling and testing shall be accredited in the testing to be performed by the AASHTO Accreditation Program or a comparable laboratory accreditation program approved by the FHWA.

(b) Sampling and testing personnel. After June 29, 2000, all sampling and testing data to be used in the acceptance decision or the IA program shall be executed by qualified sampling and testing personnel.

(c) Conflict of interest. In order to avoid an appearance of a conflict of interest, any qualified non-SHA laboratory shall perform only one of the following types of testing on the same project: Verification testing, quality control testing, IA testing, or dispute resolution testing.

Appendix A to Subpart B--Guide Letter of Certification by State Engineer

Date____
Project No.____

This is to certify that:

The results of the tests used in the acceptance program indicate that the materials incorporated in the construction work, and the construction operations controlled by sampling and testing, were in conformity with the approved plans and specifications. (The following sentence should be added if the IA testing frequencies are based on project quantities. All independent assurance samples and tests are within tolerance limits of the samples and tests that are used in the acceptance program.)

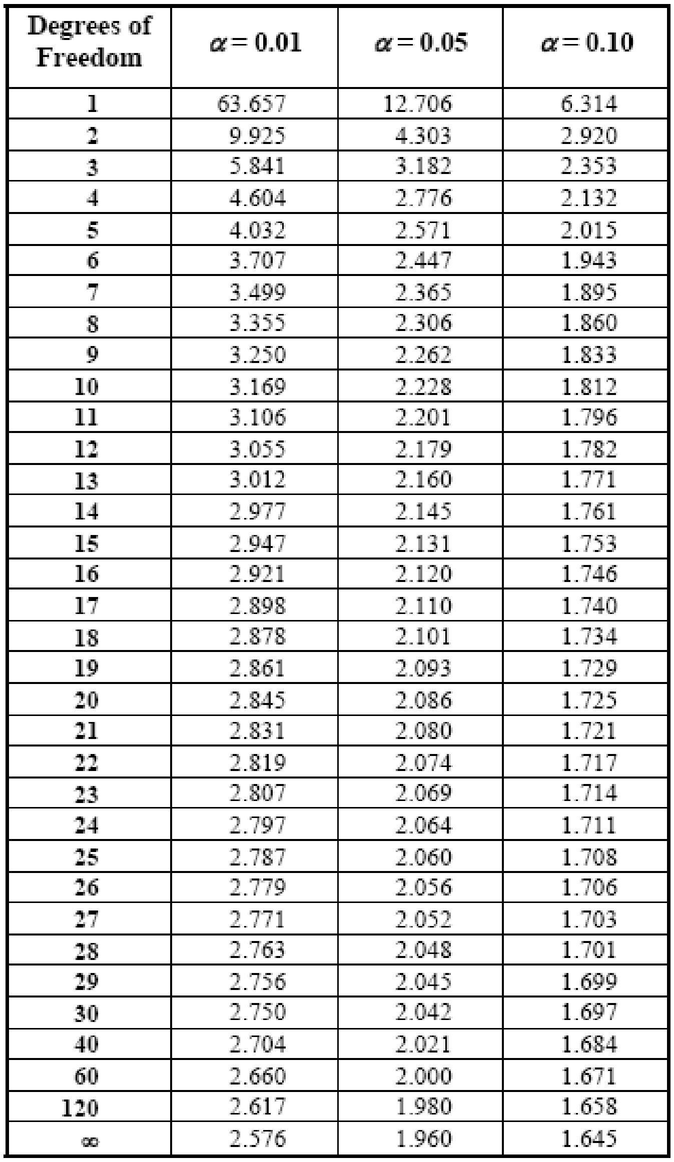
Exceptions to the plans and specifications are explained on the back hereof (or on attached sheet).

____

____
Director of SHA Laboratory or other appropriate SHA Official.

APPENDIX C: CRITICAL VALUES FOR THE *T*-TEST

Critical Values, *tcrit*, for the *t*–test*
*This Table may be interpolated



Degrees of Freedom	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	63.657	12.706	6.314
2	9.925	4.303	2.920
3	5.841	3.182	2.353
4	4.604	2.776	2.132
5	4.032	2.571	2.015
6	3.707	2.447	1.943
7	3.499	2.365	1.895
8	3.355	2.306	1.860
9	3.250	2.262	1.833
10	3.169	2.228	1.812
11	3.106	2.201	1.796
12	3.055	2.179	1.782
13	3.012	2.160	1.771
14	2.977	2.145	1.761
15	2.947	2.131	1.753
16	2.921	2.120	1.746
17	2.898	2.110	1.740
18	2.878	2.101	1.734
19	2.861	2.093	1.729
20	2.845	2.086	1.725
21	2.831	2.080	1.721
22	2.819	2.074	1.717
23	2.807	2.069	1.714
24	2.797	2.064	1.711
25	2.787	2.060	1.708
26	2.779	2.056	1.706
27	2.771	2.052	1.703
28	2.763	2.048	1.701
29	2.756	2.045	1.699
30	2.750	2.042	1.697
40	2.704	2.021	1.684
60	2.660	2.000	1.671
120	2.617	1.980	1.658
∞	2.576	1.960	1.645

[1] **NOTE:** This is for a two-tailed test with the null and alternate hypotheses shown below:

$$H_0\colon \overline{X}_c = \overline{X}_a$$
$$H_a\colon \overline{X}_c \neq \overline{X}_a$$

APPENDIX D: CRITICAL VALUES FOR *F*-TEST

Critical Values, *Fcrit* , for the *F*–test for a Level of Significance, α = 0.05*

*The following tables are for a two-tailed test; these tables may be interpolated

degrees of freedom for numerator

degrees of freedom for denominator

	1	2	3	4	5	6	7	8	9	10	11	12
1	648	799	864	900	922	937	948	957	963	969	973	977
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.4	14.3
4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10	2.05

degrees of freedom for numerator

degrees of freedom for denominator

	15	20	24	30	40	50	60	100	120	200	500
1	648	799	864	900	922	937	948	957	963	969	973
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.4
4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10

APPENDIX E: CASE STUDY DATA

OR18: Oregon Coast Hwy/ Oldsville											
		Constituent	3/4"	1/2"	#4	#8	#30	#200	Asph	Moist	Comp
		Max IA Diff	5%	5%	5%	4%	2%	1%	0.4%	1%	
		Weight	1	1	5	6	3	10	26	8	40
		Spec	100	96	61	36	16	7	5.9		
		USL	100	100	66	40	20	9	6.4	0.8	100
		LSL	95	90	56	32	12	5	5.4	0	91
Type	Sequence	**Constituent**	**1**	**2**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
QC	1	**Sublot 001**	99	94	65	41	19	9	6.04	0.48	97.8
QC	2	**Sublot 002**	100	92	61	36	16	7.1	5.97	0.42	96.9
QC	3	**Sublot 003**	99	92	60	35	15	7.1	5.95	0.41	94.3
QC	4	**Sublot 004**	100	94	62	36	15	7.1	5.9	0.43	95.6
QC	5	**Sublot 005**	100	92	59	34	15	6.9	6	0.38	94
IAC	5.5	**Sublot 092**	100	91	57	33	15	7.1	5.9		
IAO	5.5	**Sublot 093**	99	89	53	31	14	6.7	5.61		91.3
QC	6	**Sublot 006**	100	93	61	35	15	6.8	5.77	0.36	93.5
QC	7	**Sublot 007**	99	90	59	34	15	6.6	5.72	0.37	92.6
QC	8	**Sublot 008**	99	91	60	34	14	6.6	6.12	0.47	92.4
QC	9	**Sublot 009**	98	91	65	37	16	7.2	6.1	0.37	94.1
QC	10	**Sublot 010**	100	92	63	36	15	6.9	5.91	0.38	95.4
QC	11	**Sublot 011**	100	93	62	35	15	7.1	5.88	0.43	94.5
QC	12	**Sublot 012**	100	92	59	35	15	6.9	5.89	0.38	94.7
QC	13	**Sublot 013**	99	91	64	38	17	7.5	5.81	0.39	95.3
QC	14	**Sublot 014**	99	95	62	36	16	7.8	6.12	0.33	95.6
QC	15	**Sublot 015**	99	93	57	34	16	6.6	6.04	0.39	93.6
IAC	15.5	**Sublot 095**	99	88	58	33	14	6.5	5.61		
IAO	15.5	**Sublot 096**	96	83	52	30	14	6.9	5.14		93.5
QC	16	**Sublot 016**	99	91	62	36	15	7	5.92	0.26	93.5
QC	17	**Sublot 017**	99	90	56	33	16	7.4	5.95	0.24	95.4
QC	18	**Sublot 018**	98	89	56	31	14	6.3	5.7	0.35	93.2
QC	19	**Sublot 019**	100	90	63	35	15	6.9	5.66	0.38	93
QC	20	**Sublot 020**	99	90	69	37	16	7.2	6.01	0.29	93.7
QC	21	**Sublot 021**	100	91	68	37	15	6.8	6	0.32	93.8
QC	22	**Sublot 022**	97	90	65	36	16	7.2	6.27	0.24	93.7
QC	23	**Sublot 023**	96	87	67	38	16	7.3	6.21	0.32	95.3
QC	24	**Sublot 024**	100	95	69	38	17	7.4	6.04	0.32	95.8
QC	25	**Sublot 025**	99	91	64	37	16	7.1	6.05	0.28	95.1
IAC	25.5	**Sublot 098**	95	88	63	36	15	6.8	5.91		
IAO	25.5	**Sublot 099**	98	90	63	35	15	7.2	5.74		95.5
QC	26	**Sublot 026**	99	87	61	33	14	6.6	5.75	0.28	95.8

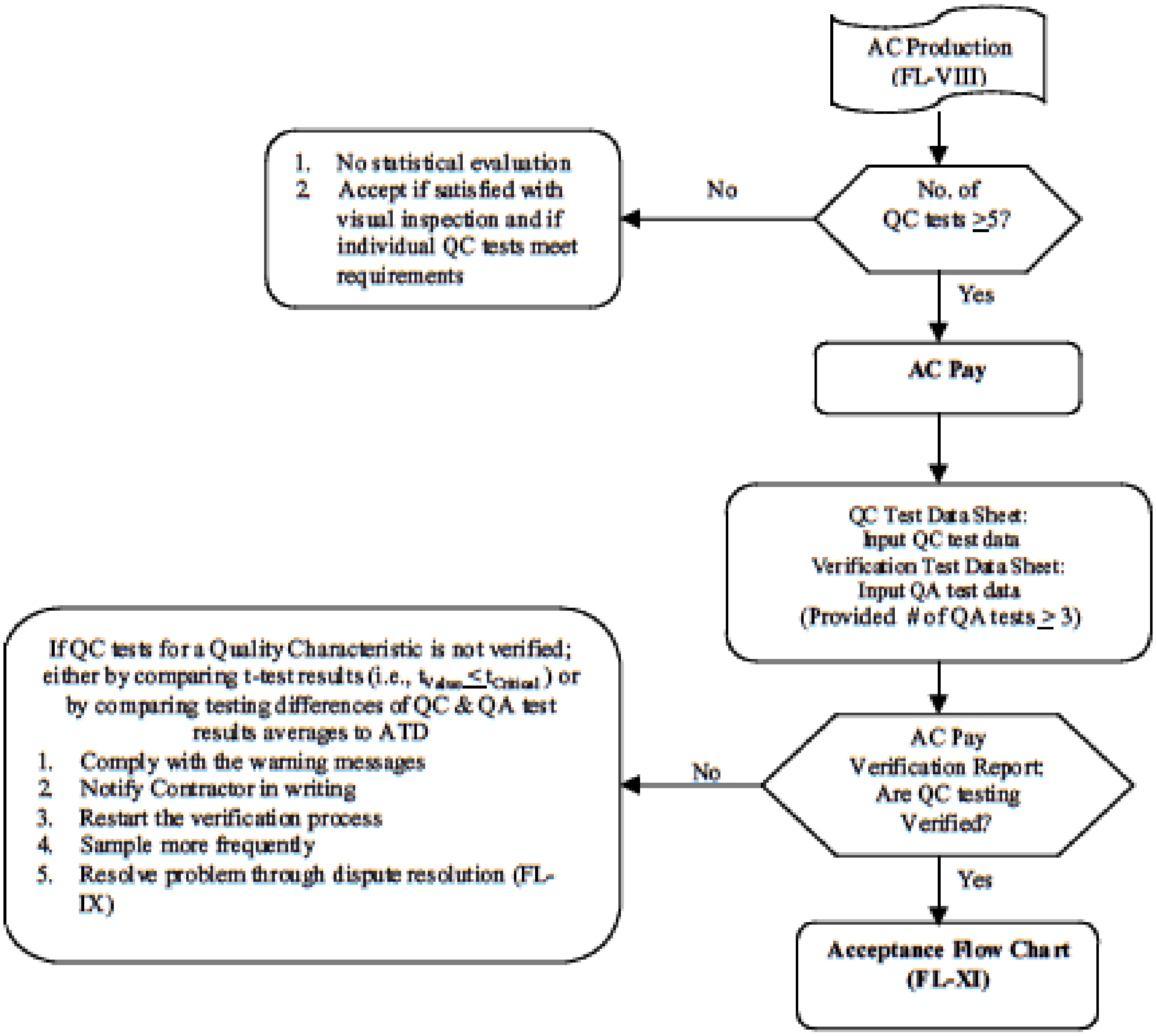
QC	27	**Sublot 027**	97	90	61	34	15	6.8	5.66	0.37	94.9
QC	28	**Sublot 028**	99	86	59	35	15	6.2	5.88	0.16	95.4
QC	29	**Sublot 029**	98	89	63	35	16	6.2	6.04	0.26	94.1
QC	30	**Sublot 030**	99	87	60	36	16	6.3	5.98	0.28	95.6
QC	31	**Sublot 031**	96	90	63	35	15	6.3	5.49	0.28	94.3
QC	32	**Sublot 032**	97	89	63	35	14	6.3	5.91	0.13	93.7
QC	33	**Sublot 033**	97	89	63	36	15	6.3	5.84	0.26	94.6
QC	34	**Sublot 034**	98	85	60	35	15	6.3	5.85	0.16	94.3
QC	35	**Sublot 035**	97	87	61	34	15	6.6	6.03	0.16	95.1
IAC	35.5	**Sublot 101**	99	91	66	38	16	7.1	6.19		
IAO	35.5	**Sublot 102**	99	95	68	38	16	7.6	6.21		95.8
QC	36	**Sublot 036**	96	89	62	33	14	5.9	5.62	0.21	94.9
QC	37	**Sublot 037**	99	95	68	37	15	6.4	6.38	0.16	94.6
QC	38	**Sublot 038**	97	85	59	35	15	6.5	5.91	0.26	92.8
QC	39	**Sublot 039**	98	91	63	37	16	7	5.91	0.16	94.4
QC	40	**Sublot 040**	98	91	61	36	16	6.6	5.95	0.17	97.1
QC	41	**Sublot 041**	100	94	65	35	15	6.5	5.8	0.19	94.8
QC	42	**Sublot 042**	99	92	64	35	16	6.5	5.84	0.18	95.3
QC	43	**Sublot 043**	97	91	64	37	15	6.8	5.78	0.15	96
QC	44	**Sublot 044**	99	91	65	37	15	6.8	6.03	0.34	96.6
QC	45	**Sublot 045**	98	87	61	36	16	6.6	5.83	0.32	95.6
IAC	45.5	**Sublot 104**	100	95	70	40	17	7.6	5.91		
IAO	45.5	**Sublot 105**	100	96	67	39	16	7.9	5.71		97.6
QC	46	**Sublot 046**	100	93	66	37	16	7.1	5.8	0.22	95.7
QC	47	**Sublot 047**	99	90	62	36	16	7.2	5.92	0.22	96.8
QC	48	**Sublot 048**	97	90	63	36	16	6.8	5.74	0.21	96.5
QC	49	**Sublot 049**	99	94	66	39	17	7.6	6.07	0.3	95.6
QC	50	**Sublot 050**	99	92	64	38	16	7.3	5.79	0.3	95.5
QC	51	**Sublot 051**	99	91	64	38	16	7.3	5.97	0.36	95.5
QC	52	**Sublot 052**	99	90	62	35	16	7.4	5.73	0.32	95.5
QC	53	**Sublot 053**	100	93	66	39	16	7.2	5.7	0.4	95.1
QC	54	**Sublot 054**	100	93	66	39	16	7.2	5.73	0.33	93.5
QC	55	**Sublot 055**	99	92	63	37	16	7.2	5.88	0.36	93.9
IAC	55.5	**Sublot 107**	98	91	61	35	15	6.7	5.75		
IAO	55.5	**Sublot 108**	100	94	66	37	16	8	5.91		94.5
QC	56	**Sublot 056**	99	92	65	37	16	7.3	5.79	0.25	95.7
QC	57	**Sublot 057**	99	90	64	36	15	7.1	5.85	0.3	93.3
QC	58	**Sublot 058**	98	90	62	36	15	6.8	6.04	0.22	93.7
QC	59	**Sublot 059**	99	91	61	34	15	6.7	5.81	0.31	94.9
QC	60	**Sublot 060**	99	91	61	34	15	6.9	5.86	0.3	94.9
QC	61	**Sublot 061**	99	91	61	36	16	6.9	5.63	0.32	93.6
QC	62	**Sublot 062**	99	92	62	34	15	6.9	5.58	0.17	93.7
QC	63	**Sublot 063**	99	92	62	36	15	6.4	5.59	0.19	94.7

QC	64	**Sublot 064**	99	92	63	37	16	7.5	6.33	0.25	95.3
QC	65	**Sublot 065**	98	87	60	35	15	6.4	5.95	0.23	94.3
IAC	65.5	**Sublot 110**	99	92	63	36	16	7.3	6.2		
IAO	65.5	**Sublot 111**	100	92	61	34	15	7.5	5.92		92.8
QC	66	**Sublot 066**	99	93	63	35	16	7.6	5.73	0.27	94.6
QC	67	**Sublot 067**	99	89	62	36	16	7.2	5.92	0.26	94.8
QC	68	**Sublot 068**	99	89	62	37	15	6.7	5.95	0.29	93.8
QC	69	**Sublot 069**	99	93	60	34	15	7.1	5.85	0.35	93.1
QC	70	**Sublot 070**	99	90	60	35	16	6.7	5.86	0.32	93.1
QC	71	**Sublot 071**	98	90	60	35	15	7	5.92	0.31	94.1
QC	72	**Sublot 072**	99	90	60	35	15	7	5.84	0.31	96
QC	73	**Sublot 073**	99	90	61	36	15	7.1	5.85	0.31	95.4
QC	74	**Sublot 074**	99	88	58	33	14	6.5	5.76	0.35	93
QC	75	**Sublot 075**	98	89	60	35	15	6.7	5.93	0.3	93.5
IAC	75.5	**Sublot 113**	97	89	59	34	15	6.9	6.03		
IAO	75.5	**Sublot 114**	99	90	59	35	15	7.3	6.04		94.5
QC	76	**Sublot 076**	98	89	60	35	15	6.9	5.93	0.27	93.2
QC	77	**Sublot 077**	99	90	61	35	15	7	5.96	0.26	94.1
QC	78	**Sublot 078**	98	89	60	34	16	7.3	5.93	0.33	94.6
QC	79	**Sublot 079**	99	89	62	36	16	7.3	6.04	0.33	93.2
QC	80	**Sublot 080**	100	90	59	34	14	6.5	6.08	0.27	93.1
QC	81	**Sublot 081**	99	89	60	34	15	6.8	5.97	0.32	95.4
QC	82	**Sublot 082**	99	93	63	34	14	6.6	5.72	0.32	95.4
QC	83	**Sublot 083**	100	94	60	34	15	6.7	5.86	0.22	93.2
QC	84	**Sublot 084**	99	92	61	34	14	6.4	5.89	0.35	93.4
QC	85	**Sublot 085**	99	91	60	34	15	6.5	5.9	0.33	94.3
IAC	85.5	**Sublot 116**	100	91	59	34	15	6.8	5.98		
IAO	85.5	**Sublot 117**	100	93	59	34	15	8.1	5.85		93.8
QC	86	**Sublot 086**	99	94	59	34	15	6.7	5.86	0.25	96.2
QC	87	**Sublot 087**	99	92	60	35	15	6.5	5.84	0.36	96.1
QC	88	**Sublot 088**	100	91	57	34	15	7.3	5.79	0.42	94.7
QC	89	**Sublot 089**	100	91	62	37	16	7.3	5.93	0.35	95.8

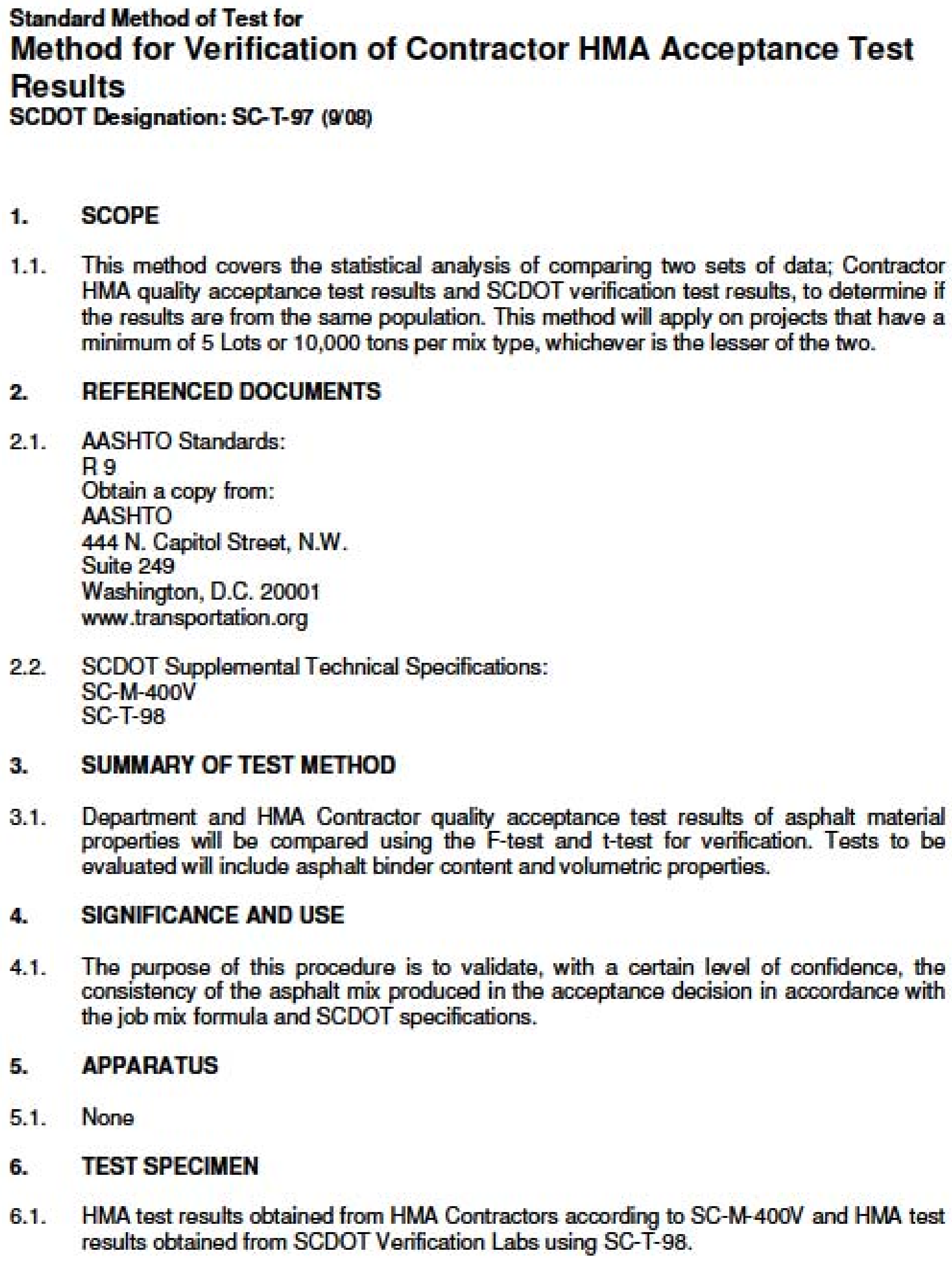
OR18: Ft Hill-Wallace Bridge											
		Constituent	3/4"	1/2"	#4	#8	#30	#200	Asph	Moist	Comp
		Max IA Diff	5%	5%	5%	4%	2%	1%	0.4%	1%	
		Weight	1	1	5	6	3	10	26	8	40
		Spec	100	96	61	36	16	7	5.9		
		USL	100	100	61	40	21	8.5	6	0.8	100
		LSL	95	90	51	32	13	4.5	5	0	92
Type	Sequence	**Constituent #**	**1**	**2**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
QC	1	**Sublot 001**	99	93	60	40	19	8.2	5.53	0.35	94.7
IAC	1.5	**Sublot 064**	99	94	58	37	16	7.1	5.64		
IAO	1.5	**Sublot 065**	99	96	58	37	16	7.2	5.56		94.3
QC	2	**Sublot 002**	100	94	58	36	15	6.5	5.49	0.36	92.8
QC	3	**Sublot 003**	100	93	58	37	16	6.4	5.84	0.46	94.2
QC	4	**Sublot 004**	100	93	56	36	15	6.5	5.74	0.43	93.8
QC	5	**Sublot 005**	98	93	58	38	17	7.5	5.71	0.41	93.9
QC	6	**Sublot 006**	100	93	57	36	16	7	5.71	0.38	94.1
QC	7	**Sublot 007**	99	91	55	34	15	6.7	5.51	0.44	93.8
QC	8	**Sublot 008**	100	94	59	39	17	7.1	5.4	0.29	94.5
QC	9	**Sublot 009**	100	93	59	39	18	7.4	5.49	0.25	93.2
QC	10	**Sublot 010**	100	92	57	37	17	6.5	5.62	0.29	93.8
QC	11	**Sublot 011**	100	92	57	37	17	7.1	5.44	0.22	93.4
IAC	11.5	**Sublot 067**	100	94	56	37	17	6.6	5.41		
IAO	11.5	**Sublot 068**	99	95	59	38	17	7.2	5.28		93.1
QC	12	**Sublot 012**	99	92	55	35	17	7.7	5.45	0.27	94
QC	13	**Sublot 013**	99	91	53	35	16	6.5	5.35	0.3	93.5
QC	14	**Sublot 014**	100	94	57	37	16	6.5	5.37	0.36	92.7
QC	15	**Sublot 015**	100	91	55	36	17	7.6	5.32	0.34	93.6
QC	16	**Sublot 016**	100	93	56	37	17	6.8	5.52	0.35	92.8
QC	17	**Sublot 017**	99	92	55	35	16	6.8	5.44	0.27	94
QC	18	**Sublot 018**	100	93	57	37	17	7	5.47	0.33	93.7
QC	19	**Sublot 019**	99	91	56	36	17	7.5	5.34	0.33	93.5
QC	20	**Sublot 020**	99	93	58	37	18	7.5	5.63	0.33	92.5
QC	21	**Sublot 021**	98	92	56	37	17	7.1	5.48	0.41	92.8
IAC	21.5	**Sublot 070**	99	94	56	35	16	6.3	5.24		
IAO	21.5	**Sublot 071**	100	95	55	35	15	6.3	4.93		94.7
QC	22	**Sublot 022**	100	92	56	37	17	7.6	5.31	0.31	94.3
QC	23	**Sublot 023**	100	94	55	35	17	6.7	5.42	0.34	93.8
QC	24	**Sublot 024**	100	94	58	38	17	6.8	5.64	0.27	93.5
QC	25	**Sublot 025**	100	93	56	36	16	6.1	5.49	0.26	92.7
QC	26	**Sublot 026**	100	94	55	35	15	6	5.35	0.31	93.9
QC	27	**Sublot 027**	99	93	56	37	18	8.2	5.31	0.31	94.4

QC	28	**Sublot 028**	100	93	53	34	15	6.3	5.6	0.34	93.6
QC	29	**Sublot 029**	100	94	58	38	18	7.1	5.59	0.4	93.2
QC	30	**Sublot 030**	100	94	59	37	18	7.8	5.61	0.4	94.2
QC	31	**Sublot 031**	100	95	56	37	17	7.1	5.55	0.32	93.3
IAC	31.5	**Sublot 073**	100	91	54	35	16	6.4	5.43		
IAO	31.5	**Sublot 074**	100	95	57	37	16	6.7	5.42		93
QC	32	**Sublot 032**	98	90	53	35	17	7.5	5.12	0.32	93.7
QC	33	**Sublot 033**	100	91	55	36	17	7.4	5.23	0.28	93.3
QC	34	**Sublot 034**	100	92	52	32	14	5.9	5.51	0.32	93.1
QC	35	**Sublot 035**	100	94	55	36	17	7.9	5.43	0.21	93.5
QC	36	**Sublot 036**	99	91	53	34	16	7.5	5.25	0.23	93.5
QC	37	**Sublot 037**	100	92	53	34	16	7.6	5.33	0.3	93.2
QC	38	**Sublot 038**	100	95	55	36	17	8	5.49	0.32	94.4
QC	39	**Sublot 039**	100	93	55	34	15	6.2	5.82	0.26	93.7
QC	40	**Sublot 040**	100	90	52	33	15	6.7	5.64	0.34	94
QC	41	**Sublot 041**	98	90	53	34	15	5.8	5.46	0.54	94.2
IAC	41.5	**Sublot 076**	100	91	50	30	13	5.8	5.31		
IAO	41.5	**Sublot 077**	100	91	48	29	12	5.5	5.12		94.5
QC	42	**Sublot 042**	100	93	56	36	14	5.3	5.76	0.39	93.1
QC	43	**Sublot 043**	100	95	58	37	16	6.4	5.31	0.39	93.1
QC	44	**Sublot 044**	100	93	53	35	16	6.7	5.41	0.32	92.8
QC	45	**Sublot 045**	100	92	54	34	15	5.7	5.25	0.28	92.5
QC	46	**Sublot 046**	99	94	53	34	15	7	5.73	0.46	93.6
QC	47	**Sublot 047**	100	93	53	34	16	7.1	5.67	0.41	93.2
QC	48	**Sublot 048**	100	95	56	35	16	7.4	5.5	0.44	93.8
QC	49	**Sublot 049**	99	91	54	34	17	8	5.62	0.35	94.1
QC	50	**Sublot 050**	99	92	55	35	16	7.2	5.28	0.39	94.2
QC	51	**Sublot 051**	100	91	54	34	16	6.9	5.42	0.45	94
IAC	51.5	**Sublot 079**	100	93	54	35	16	6.6	5.49		
IAO	51.5	**Sublot 080**	100	94	56	36	16	6.9	5.46		93.4
QC	52	**Sublot 052**	100	96	55	33	15	6.5	5.57	0.33	93.8
QC	53	**Sublot 053**	100	93	53	34	15	7	5.45	0.48	94.2
QC	54	**Sublot 054**	100	92	54	35	16	7.7	5.26	0.44	94.2
QC	55	**Sublot 055**	100	92	52	34	16	7.8	5.24	0.32	93.8
QC	56	**Sublot 056**	100	92	56	36	16	6.9	5.37	0.34	93.5
QC	57	**Sublot 057**	100	91	53	34	15	6.5	5.48	0.28	93.3
QC	58	**Sublot 058**	100	93	55	35	15	6.1	5.66	0.37	92.9
QC	59	**Sublot 059**	100	93	55	36	17	8	5.46	0.28	93.9
QC	60	**Sublot 060**	100	92	56	38	19	8.8	5.42	0.4	93.3
QC	61	**Sublot 061**	100	92	54	35	16	6.6	5.47	0.38	94.1
IAC	61.5	**Sublot 082**	100	93	52	32	14	6.8	5.58		
IAO	61.5	**Sublot 083**	100	91	51	31	14	6.8	5.46		92.6

APPENDIX F: CALIFORNIA VERIFICATION FLOWCHART



APPENDIX G: SOUTH CAROLINA METHOD FOR VERIFICATION OF CONTRACTOR HMA ACCEPTANCE TEST RESULTS



Standard Method of Test for

Method for Verification of Contractor HMA Acceptance Test Results

SCDOT Designation: SC-T-97 (9/08)

1. SCOPE

1.1. This method covers the statistical analysis of comparing two sets of data; Contractor HMA quality acceptance test results and SCDOT verification test results, to determine if the results are from the same population. This method will apply on projects that have a minimum of 5 Lots or 10,000 tons per mix type, whichever is the lesser of the two.

2. REFERENCED DOCUMENTS

2.1. AASHTO Standards:
R 9
Obtain a copy from:
AASHTO
444 N. Capitol Street, N.W.
Suite 249
Washington, D.C. 20001
www.transportation.org

2.2. SCDOT Supplemental Technical Specifications:
SC-M-400V
SC-T-98

3. SUMMARY OF TEST METHOD

3.1. Department and HMA Contractor quality acceptance test results of asphalt material properties will be compared using the F-test and t-test for verification. Tests to be evaluated will include asphalt binder content and volumetric properties.

4. SIGNIFICANCE AND USE

4.1. The purpose of this procedure is to validate, with a certain level of confidence, the consistency of the asphalt mix produced in the acceptance decision in accordance with the job mix formula and SCDOT specifications.

5. APPARATUS

5.1. None

6. TEST SPECIMEN

6.1. HMA test results obtained from HMA Contractors according to SC-M-400V and HMA test results obtained from SCDOT Verification Labs using SC-T-98.

7. PROCEDURE

7.1. Obtain Contractor HMA quality acceptance test results and SCDOT verification test results for the asphalt mix properties for binder content, gradation, Maximum Specific Gravity, Bulk Specific Gravity, % Air Voids, % VMA and % VFA, and forward to the HMA Verification Manager.

7.1.1. The average number of SCDOT verification tests conducted will be at a minimum of 1 test per LOT.

7.2. Input test data into the computerized statistical program spreadsheet (verified by SCDOT Materials and Research Engineer). Ensure that the test data is correlated into the spreadsheet by LOT number and date.

7.2.1. The data set to be evaluated will be test results in increment of 5 LOTS. Contractor HMA quality acceptance test results and SCDOT verification test results from LOT 1 thru LOT 5 will be statistically analyzed and a decision to accept the Contractor HMA quality acceptance test results will be based on whether the data set is believed equal and therefore, have come from the same population. If the analysis of the data set proves a non-comparison of the test results, then the SCDOT verification test results will be used for acceptance. The second data set will comprise test results from LOT 6 thru LOT 10, statistically analyzed and a decision will be made whether to accept or reject the Contractor HMA quality acceptance test results. The third data set will comprise test results from LOT 11 thru LOT 15, statistically analyzed and a decision will be made whether to accept or reject the Contractor HMA quality acceptance test results. This process continues until production is completed. If the last data set is less than 5 LOTS, then go back to the previous LOTS far enough to yield the 5 LOTS needed in the data set.

7.3. Following AASHTO R9, conduct an F-Test for the sample variances for binder content, % Air Voids and % VMA.

7.3.1. Use a level of significance of $\alpha = 0.05$ for the test.

7.3.2. Compute the mean for the Contractor HMA quality acceptance tests ($\overline{X}_c$) and the SCDOT verification tests ($\overline{X}_v$) and standard deviation for the Contractor HMA quality acceptance tests (s_c) and the SCDOT verification tests (s_v).

7.3.3. Compute the variance for the Contractor HMA quality acceptance tests (s_c^2) and the SCDOT verification tests (s_v^2). (Variance is the square of the standard deviation).

7.3.4. Compute F-statistic (F), using the largest variance (s^2) value in the numerator.

7.3.5. Determine the critical F value (F_{crit}) from the F-distribution table making sure to use the correct degrees of freedom (n-1) associated with each set of tests results. (See Table 1 in the appendix for Critical Values, F_{crit}, for the F-test for a Level of Significance $\alpha = 0.05$).

7.3.6. Evaluate if $F \geq F_{crit}$, or $F < F_{crit}$.

7.3.7. If $F \geq F_{crit,}$ then conclude that the two data sets of tests have significantly different variabilities. If $F < F_{crit,}$ then conclude that there is no reason to believe that the variabilities are significantly different.

7.4. Following AASHTO R9, conduct t-test for the sample means for binder content, % Air Voids and % VMA.

Note: Two approaches for the t-test are necessary. If the sample variances are found to be equal from the F- Test, then the t-test is conducted based on the two samples using a pooled estimate for the variance and the pooled degrees of freedom. If the sample variances are found to be different from the F-Test, then the t-test is conducted using the individual sample variances, the individual sample sizes, and the effective degrees of freedom.

7.4.1. Use a level of significance of $\alpha = 0.01$ for the test.

7.4.2. Compute the mean for the Contractor HMA quality acceptance tests $(\overline{X}_c)$ and the SCDOT verification tests $(\overline{X}_v)$.

7.4.3. Compute the t-statistic (t), using either the pooled variance (s_p^2) equation (equal variances) or the equation for unequal variances as appropriate.

7.4.4. Determine the critical t value (t_{crit}) using the level of significance (α) the pooled degrees of freedom ($n_c + n_a - 2$) or the effective degrees of freedom (f') as appropriate. (See Table 2 in the appendix for Critical Values, t_{crit}, for the t-test).

7.4.5. Evaluate if $t \geq t_{crit,}$ or $t < t_{crit}$.

7.4.6. If $t \geq t_{crit,}$ then conclude that the two data sets of tests have significantly different means. If $t < t_{crit,}$ then conclude that there is no reason to believe that the means are significantly different.

7.5. Determine if the two data sets are statistically equal based on statistical hypothesis two-tailed tests.

7.6. If the results of either the F- or t - test is deemed not statistically equal (the null hypothesis is rejected and either the means or variances are significantly different), contact the District Asphalt Manager and Asphalt Materials Manager and investigate to determine if any discrepancies or issues in the production, sampling or testing of the HMA can be identified.

8. CALCULATION

8.1. All calculations for validation of the HMA volumetric properties will be conducted using the statistical F-test and t-test method for computing two sets of data as indicated in AASHTO R 9.

Example 1 – Percent Air Voids in Hot Mix Asphalt: Sample Variances Assumed to be Equal

A contractor has run 12 quality acceptance tests and SCDOT has run 5 verification tests over the same period of time for the percent air voids (%AV). Is it likely that the test came from the same population or lot?

Contractor QA Test Results (n_c)			SCDOT Verification Test Results (n_v)
3.50	3.77	3.79	5.05
3.56	3.05	2.77	2.65
3.06	3.78		3.78
3.12	4.48		3.18
4.00	3.34		4.51

Step 1. Compute the mean and standard deviation for each set of data:

<u>QA test results</u>	<u>SCDOT Verification test results</u>
$\overline{X}_c = 3.5183$	$\overline{X}_v = 3.834$
$s_c = 0.4809$	$s_v = 0.9706$

Step 2. Compute the variance, s^2, for each set of tests:
$s_c^2 = 0.2313$ $s_v^2 = 0.9421$

Step 3. Compute F, using the largest variance (s^2) in the numerator:

$F = s_v^2 \div s_c^2 = 0.9421 \div 0.2313 = 4.07$

Step 4. Determine F_{crit} from the F-distribution table making sure to use the correct degrees of freedom for the numerator ($n_v - 1 = 5 - 1 = 4$) and the denominator ($n_c - 1 = 12 - 1 = 11$).

Using $\alpha = 5$ % and the degrees of freedom (n-1) $F_{crit} = 4.28$

Conclusion: Since $F < F_{crit}$ (ie. $4.07 < 4.28$), there is no reason to believe that the two data sets of tests have different variabilities. That is, they could have come from the same population. Since we can assume that the variances are equal, we can use *the pooled variance* to calculate the t-test statistic, and *the pooled degrees of freedom* to determine the critical t value, t_{crit}.

Step 5. Compute the pooled variance, s_p^2, using the pooled sample variance and the pooled degrees of freedom from above.

$$s_p^2 = \frac{s_c^2(n_c - 1) + s_v^2(n_v - 1)}{n_c + n_v - 2}$$

$$s_p^2 = \frac{(0.2313)(11) + (0.9421)(4)}{12 + 5 - 2} = 0.4208$$

Step 6. Compute the t-test statistic, t, using the pooled sample variance.

$$t = \frac{\left|\overline{X}_c - \overline{X}_v\right|}{\sqrt{\frac{s_p^2}{n_c} + \frac{s_p^2}{n_v}}} \qquad t = \frac{\left|3.5183 - 3.834\right|}{\sqrt{\frac{0.4208}{12} + \frac{0.4208}{5}}} = \frac{0.3157}{\sqrt{0.1192}} = 0.914$$

Step 7. Determine the critical t value t_{crit}, for the pooled degrees of freedom.

Degrees of freedom = $(n_c + n_v - 2) = (12 + 5 - 2) = 15$

Using α = 1 % and the degrees of freedom = 15 (pooled variance will always be an integer for degrees of freedom) $t_{crit} = 2.947$

Conclusion: Since $t < t_{crit}$ (ie. $0.914 < 2.947$), there is no reason to assume that the sample means are not equal. It is, therefore, reasonable to assume that the sets of test results came from the same population (or LOT). Therefore, the contractor's test results can be used for acceptance.

Example 2 – Percent Air Voids in Hot Mix Asphalt: Sample Variances Assumed to be Different

A contractor has run 10 quality acceptance tests and SCDOT has run 5 verification tests over the same period of time for the percent air voids (%AV). Is it likely that the test came from the same population or lot?

Contractor QA Test Results (n_c)		SCDOT Verification Test Results (n_v)
6.42	7.98	7.52
7.18	6.32	11.38
5.04	6.08	9.20
4.56	5.92	5.32
7.12	5.78	3.18

Step 1. Compute the mean and standard deviation for each set of data:

QA test results	SCDOT Verification test results
$\overline{X}_c = 6.24$	$\overline{X}_v = 7.32$
$s_c = 1.036$	$s_v = 10.299$

Step 2. Compute the variance, s^2, for each set of tests:

$s_c^2 = 1.036$ $s_v^2 = 10.299$

Step 3. Compute F, using the largest variance (s^2) in the numerator:

$F = s_v^2 \div s_c^2 = 10.299 \div 1.036 = 9.94$

Step 4. Determine F_{crit} from the F-distribution table making sure to use the correct degrees of freedom for the numerator ($n_v - 1 = 5\text{-}1 = 4$) and the denominator ($n_c - 1 = 10\text{-}1 = 9$).

Using $\alpha = 5$ % and the degrees of freedom (n-1) $F_{crit} = 4.72$

Conclusion: Since $F > F_{crit}$ (ie. $9.94 > 4.72$), it is unlikely that the two data sets came from the same population. Therefore, conclude that the contractor and SCDOT results came from different populations.

Step 6. Compute the t-test statistic, t, using the equation for unequal variances.

$$t = \frac{\left|\overline{X}_c - \overline{X}_v\right|}{\sqrt{\dfrac{s_c^2}{n_c} + \dfrac{s_v^2}{n_v}}} \qquad t = \frac{\left|6.24 - 7.32\right|}{\sqrt{\dfrac{1.036}{10} + \dfrac{10.299}{5}}} = \frac{1.08}{\sqrt{2.1634}} = 0.734$$

Step 7. Determine the critical t value t_{crit}, for the effective degrees of freedom, f.

$$f = \frac{\left(\dfrac{s_c^2}{n_c} + \dfrac{s_v^2}{n_v}\right)^2}{\left[\dfrac{\left(\dfrac{s_c^2}{n_c}\right)^2}{n_c + 1} + \dfrac{\left(\dfrac{s_v^2}{n_v}\right)^2}{n_v + 1}\right]} - 2 = \frac{\left(\dfrac{1.036}{10} + \dfrac{10.299}{5}\right)^2}{\left[\dfrac{\left(\dfrac{1.036}{10}\right)^2}{10 + 1} + \dfrac{\left(\dfrac{10.299}{5}\right)^2}{5 + 1}\right]} - 2 = 4.61 = 4$$

Using $\alpha = 1$ % and the effective degrees of freedom = 4
(the calculated value for effective degrees of freedom is truncated to the lower integer)

$t_{crit} = 4.604$

Conclusion: Since $t < t_{crit}$ (ie. $0.734 < 4.604$), there is no reason to assume that the sample means are not equal. It is, therefore, reasonable to assume that the sets of test results came from the same population (or LOT). However, since the results of the F-test were different the SCDOT should investigate the cause for the differences in variances.

9. REPORT

9.1. Report – F-test and t-test results and a statement as to whether or not the Contractor's HMA quality acceptance test results compare to the SCDOT verification results and can be used for acceptance will be forwarded to the District Asphalt Manager for completing the project payment functions.

Table 1 **Critical Values, F_{crit} for the F-test for a Level of Significance, $\alpha = 5\%$**

DEGREES OF FREEDOM FOR NUMERATOR

DEGREES OF FREEDOM FOR DENOMINATOR

	1	2	3	4	5	6	7	8	9	10	11	12
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	973.03	976.71
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.41
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.37	14.34
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10	2.05

NOTE : This is for a ***two-tailed test*** with the null and alternate hypotheses shown below:

$H_o: \ s_o^2 = s_v^2$

$H_a: \ s_o^2 \neq s_v^2$

Table 1 Critical Values, F_{crit} for the F-test for a Level of Significance, $\alpha = 5\%$ (contd.)

DEGREES OF FREEDOM FOR NUMERATOR

DEGREES OF FREEDOM FOR DENOMINATOR

	13	14	15	20	24	30	40	50	60	100	120
1	979.84	982.53	984.87	993.10	997.25	1001.41	1005.60	1008.12	1009.80	1013.17	1014.02
2	39.42	39.43	39.43	39.45	39.46	39.46	39.47	39.48	39.48	39.49	39.49
3	14.30	14.28	14.25	14.17	14.12	14.08	14.04	14.01	13.99	13.96	13.95
4	8.71	8.68	8.66	8.56	8.51	8.46	8.41	8.38	8.36	8.32	8.31
5	6.49	6.46	6.43	6.33	6.28	6.23	6.18	6.14	6.12	6.08	6.07
6	5.33	5.30	5.27	5.17	5.12	5.07	5.01	4.98	4.96	4.92	4.90
7	4.63	4.60	4.57	4.47	4.41	4.36	4.31	4.28	4.25	4.21	4.20
8	4.16	4.13	4.10	4.00	3.95	3.89	3.84	3.81	3.78	3.74	3.73
9	3.83	3.80	3.77	3.67	3.61	3.56	3.51	3.47	3.45	3.40	3.39
10	3.58	3.55	3.52	3.42	3.37	3.31	3.26	3.22	3.20	3.15	3.14
11	3.39	3.36	3.33	3.23	3.17	3.12	3.06	3.03	3.00	2.96	2.94
12	3.24	3.21	3.18	3.07	3.02	2.96	2.91	2.87	2.85	2.80	2.79
13	3.12	3.08	3.05	2.95	2.89	2.84	2.78	2.74	2.72	2.67	2.66
14	3.01	2.98	2.95	2.84	2.79	2.73	2.67	2.64	2.61	2.56	2.55
15	2.92	2.89	2.86	2.76	2.70	2.64	2.59	2.55	2.52	2.47	2.46
20	2.64	2.60	2.57	2.46	2.41	2.35	2.29	2.25	2.22	2.17	2.16
24	2.50	2.47	2.44	2.33	2.27	2.21	2.15	2.11	2.08	2.02	2.01
30	2.37	2.34	2.31	2.20	2.14	2.07	2.01	1.97	1.94	1.88	1.87
40	2.25	2.21	2.18	2.07	2.01	1.94	1.88	1.83	1.80	1.74	1.72
60	2.18	2.14	2.11	1.99	1.93	1.87	1.80	1.75	1.72	1.66	1.64
120	2.13	2.09	2.06	1.94	1.88	1.82	1.74	1.70	1.67	1.60	1.58

NOTE : This is for a ***two-tailed test*** with the null and alternate hypotheses shown below:

$H_o: \; s_c^2 = s_v^2$

$H_a: \; s_c^2 \neq s_v^2$

Table 2 **Critical Values, t_{crit} for the t-test**

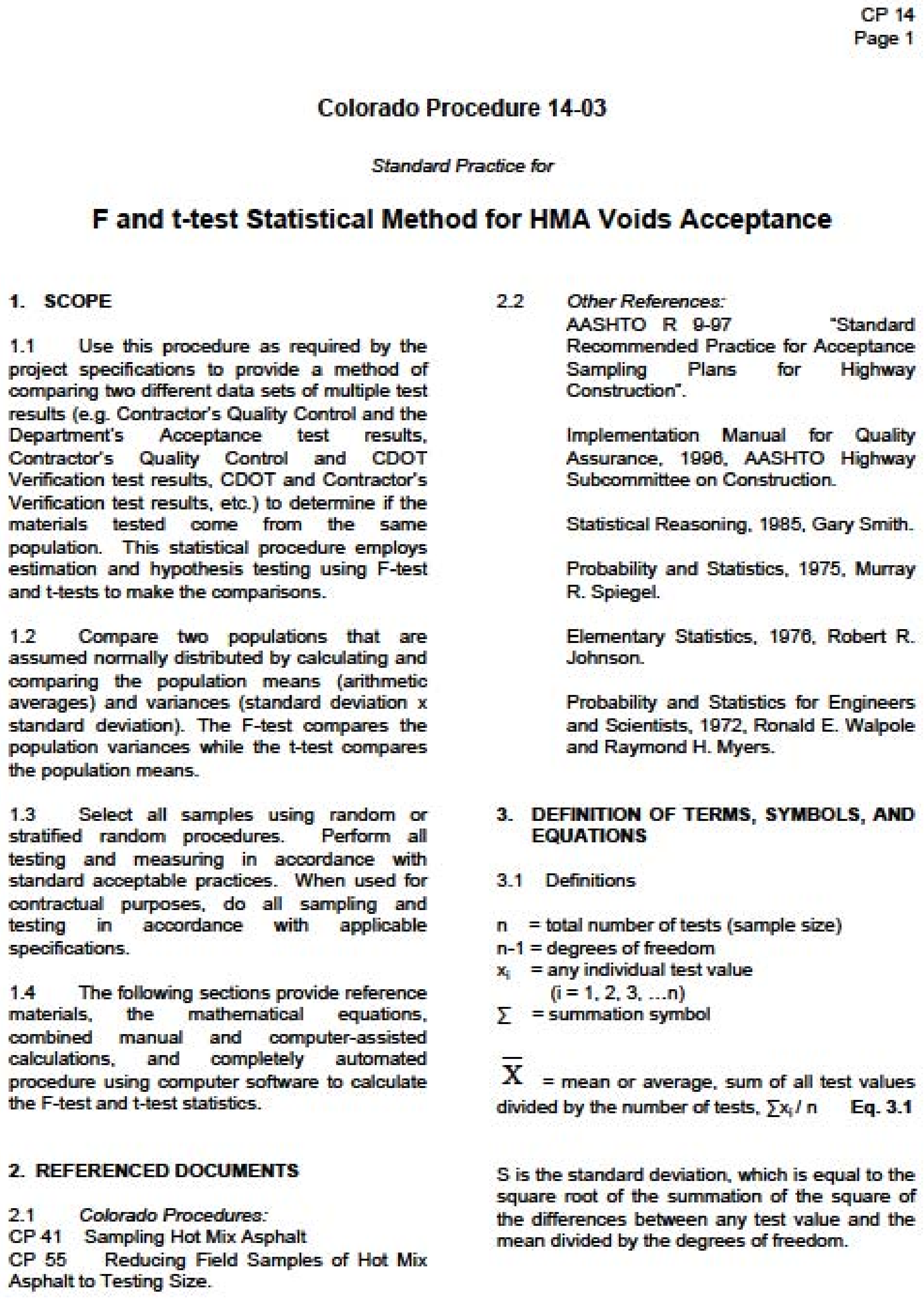
degrees of freedom	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	63.657	12.706	6.314
2	9.925	4.303	2.920
3	5.841	3.182	2.353
4	4.604	2.776	2.132
5	4.032	2.571	2.015
6	3.707	2.447	1.943
7	3.499	2.365	1.895
8	3.355	2.306	1.860
9	3.250	2.262	1.833
10	3.169	2.228	1.812
11	3.106	2.201	1.796
12	3.055	2.179	1.782
13	3.012	2.160	1.771
14	2.977	2.145	1.761
15	2.947	2.131	1.753
16	2.921	2.120	1.746
17	2.898	2.110	1.740
18	2.878	2.101	1.734
19	2.861	2.093	1.729
20	2.845	2.086	1.725
21	2.831	2.080	1.721
22	2.819	2.074	1.717
23	2.807	2.069	1.714
24	2.797	2.064	1.711
25	2.787	2.060	1.708
26	2.779	2.056	1.706
27	2.771	2.052	1.703
28	2.763	2.048	1.701
29	2.756	2.045	1.699
30	2.750	2.042	1.697
40	2.704	2.021	1.684
60	2.660	2.000	1.671
120	2.617	1.980	1.658
∞	2.576	1.960	1.645

NOTE : This is for a two-tailed test with the null and alternate hypotheses shown below :

H_o : $\overline{X}_c = \overline{X}_v$

H_a : $\overline{X}_c \neq \overline{X}_v$

APPENDIX H: COLORADO F- AND T-TEST STATISTICAL METHOD FOR HMA VOIDS ACCEPTANCE



Colorado Procedure 14-03

Standard Practice for

F and t-test Statistical Method for HMA Voids Acceptance

1. SCOPE

1.1 Use this procedure as required by the project specifications to provide a method of comparing two different data sets of multiple test results (e.g. Contractor's Quality Control and the Department's Acceptance test results, Contractor's Quality Control and CDOT Verification test results, CDOT and Contractor's Verification test results, etc.) to determine if the materials tested come from the same population. This statistical procedure employs estimation and hypothesis testing using F-test and t-tests to make the comparisons.

1.2 Compare two populations that are assumed normally distributed by calculating and comparing the population means (arithmetic averages) and variances (standard deviation x standard deviation). The F-test compares the population variances while the t-test compares the population means.

1.3 Select all samples using random or stratified random procedures. Perform all testing and measuring in accordance with standard acceptable practices. When used for contractual purposes, do all sampling and testing in accordance with applicable specifications.

1.4 The following sections provide reference materials, the mathematical equations, combined manual and computer-assisted calculations, and completely automated procedure using computer software to calculate the F-test and t-test statistics.

2. REFERENCED DOCUMENTS

2.1 *Colorado Procedures:*
CP 41 Sampling Hot Mix Asphalt
CP 55 Reducing Field Samples of Hot Mix Asphalt to Testing Size.

2.2 *Other References:*
AASHTO R 9-97 "Standard Recommended Practice for Acceptance Sampling Plans for Highway Construction".

Implementation Manual for Quality Assurance, 1996, AASHTO Highway Subcommittee on Construction.

Statistical Reasoning, 1985, Gary Smith.

Probability and Statistics, 1975, Murray R. Spiegel.

Elementary Statistics, 1976, Robert R. Johnson.

Probability and Statistics for Engineers and Scientists, 1972, Ronald E. Walpole and Raymond H. Myers.

3. DEFINITION OF TERMS, SYMBOLS, AND EQUATIONS

3.1 Definitions

n = total number of tests (sample size)
$n-1$ = degrees of freedom
x_i = any individual test value ($i = 1, 2, 3, \ldots n$)
$\sum$ = summation symbol

$\overline{X}$ = mean or average, sum of all test values divided by the number of tests, $\sum x_i / n$ **Eq. 3.1**

S is the standard deviation, which is equal to the square root of the summation of the square of the differences between any test value and the mean divided by the degrees of freedom.

$$S=\sqrt{\frac{\Sigma(Xi-\overline{X})^{2}}{n-1}} \qquad \textbf{Eq. 3.2}$$

V = sample variance, S^2 **Eq. 3.3**

α = level of significance or critical region. This is the probability of incorrectly deciding the data sets are different when they actually come from the same population. In either the construction or the manufacturing industry, α is the risk of rejecting a good material or product. The critical region, α (critical area) in the F and t probability distribution curves is equivalent to the rejection area. Since the total area bounded by either the F or t distribution curve is equal to 1 - α, the acceptance region is 1 - α. When α = 0.05, there is a probability of 95 percent that the two data sets are from the same population. When α = 0, it means no rejection and 100 percent of either the data or product is acceptable.

The two-tailed test hypothesis method tests if the population parameters (variances or averages) are either equal or not equal. All the values of α obtained from this procedure are based on the two-tailed testing approach (use two-tailed statistical tables or functions).

F = the ratio of the variance from each data set (larger variance divided by smaller variance: S_1^2/S_2^2 or S_2^2/S_1^2), **Eq. 3.4** depending on which ratio yields a value equal to or greater than 1; S1 and S2 are respective variances of the first and second data set;

$$|t|=\overline{D}/\left(S_d/\sqrt{n}\right),$$

paired t-test for paired observations (split samples). The variability is attributed to the test method only because of using identical sampling procedure and specimen. **Eq. 3.5**

$\overline{D}$ = average of the differences between paired observations or test results,

$$\frac{\Sigma(x_1-x_2)}{n}, \qquad \textbf{Eq. 3.6}$$

where x_1 and x_2 are paired observations.

S_d = standard deviation of the differences between paired observations or test results, similar to Eq. 3.2 with D_i and $\overline{D}$ replacing x_i and $\overline{x}$ respectively.

$$S_d=\sqrt{\frac{\Sigma\left(D_i-\overline{D}\right)^2}{n-1}} \qquad \textbf{Eq. 3.7}$$

where D_i = difference between paired observations, $x_1 - x_2$.

$$|t|=\left(\overline{X}_1-\overline{X}_2\right)/S_p\sqrt{(1/n_1)+(1/n_2)}$$

for independent samples. The variability is attributed to material, sampling, and test method. **Eq. 3.8**

$\overline{X}_1$ and $\overline{X}_2$ are respective averages of the first and second data set, and

n_1 and n_2 are respective sample sizes of the first and second data set.

$|\;|$ = absolute value (disregard sign)

S_p = pooled estimate for the variance,

$$\sqrt{\frac{(n_1-1)S_1^{2}+(n_2-1)S_2^{2}}{(n_1+n_2-2)}} \qquad \textbf{Eq. 3.9}$$

4. SUMMARY OF METHOD

4.1 The method involves calculating sample statistics from three or more representative measurements, test results, or values, for each specified element in a lot or sample. The specimen may either be a split or independent sample. The statistical variables to be calculated include the mean, standard deviation, variance, F and t values, and the α -value. The following sections summarize the F-test and t-test method to be employed in this procedure.

4.2 Determine the appropriate population parameters and sample statistics to be used in estimation and hypothesis testing (F & t-tests). For the F-test calculation, test the assumption that the population variances are equal against the assumption that they are not equal (use a two-tailed F-test). For the t-test calculation, assume the population variances are equal and test the assumption that the population means are equal against the assumption that they are not equal (use a two-tailed t-test).

4.3 Choose a level of significance or critical region (α) for each of the F-test and t-test calculations. AASHTO R 9-97 provides suggested critical values of α used in the highway construction industry. CDOT typically uses values of 0.10, 0.05, 0.01, and 0.005. In this procedure, use α -values as specified in the project specifications.

4.4 Calculate all the required variables in the appropriate F-test and t-test equations (split or independent samples for t-test) and compare the calculated α -value with the level of significance chosen in the previous subsection.

4.5 Conclude that the measurements, test results, or test values come from the same population if the calculated α -value is greater than the selected level of significance. Conclude that the measurements, test results, or test values do not come from the same population if the calculated α -value is less than the selected level of significance.

5. COMBINED MANUAL AND COMPUTER-ASSISTED PROCEDURE

5.1 Determine the appropriate arithmetic mean, standard deviation, and F-value for the test results from the lot (either split or independent samples) for each element being evaluated using Equations 3.1, 3.2, 3.3, and 3.4. Given the F-value and the degrees of freedom for the numerator and the denominator, calculate the α -value using either any applicable software function or Microsoft Excel statistical function FDIST. The function FDIST calculates the one-tailed α -value. This α -value should be multiplied by 2 to obtain the α -value for the two-tailed test. The FDIST function has the command format FDIST [calculated F-value, numerator degrees of freedom, denominator degrees of freedom]. Compare this result with the level of significance previously selected for the F-test. Conclude that the test data are not from the same population if the calculated α -value is less than the selected level of significance. If the calculated α -value is greater than the selected level of significance, perform a t-test calculation assuming equal population variances.

5.2 Determine the absolute t-value using Equations 3.5 to 3.7 for split samples and Equations 3.8 to 3.9 for independent samples. Given the t-value, the degrees of freedom, and the number of tails equal to 2, calculate the α -value using either any applicable software function or Microsoft Excel statistical function TDIST. The TDIST function has the command format TDIST [calculated t-value, degrees of freedom, tails]. Compare this result with the level of significance previously selected for the t-test. Conclude that the test data are not from the same population if the calculated α -value is less than the selected level of significance. Conclude that the test data are from the same population if the calculated α -value is greater than the previously selected level of significance.

6. COMPUTER-ASSISTED PROCEDURE

6.1 Any applicable computer software with statistical functions may be used to conduct F-test and t-test calculations. The Microsoft Excel statistical function FTEST can be used to calculate the α-value for the F-test while the Microsoft Excel statistical function TTEST can be used to calculate the α-value for the t-test. The FTEST function has the command format FTEST [array1, array2]. Array1 is the first data set and array2 is the second data set. The FTEST function directly calculates the two-tailed α-value. Compare this value with the selected level of significance. Conclude that the test data are not from the same population if the result of the FTEST calculation is less than the selected level of significance. Proceed to conducting a t-test assuming equal population variances if the result of the FTEST calculation is greater than the selected level of significance.

6.2 The Microsoft Excel TTEST function has the command format TTEST [array1, array2, tails, type]. Array1 is the first data set and array2 is the second data set. The tails parameter specifies the number of distribution tails and type means the kind of t-test to perform. The type can be 1 (paired t-test), 2 (two-sample equal variance) and 3 (two-sample unequal variance). Type 3 is not used in this procedure because the test data sets are automatically concluded to be not from the same population if the sample variances are found to be unequal. The t-test directly calculates the α-value, given the required values of the variables in the TTEST function. Compare this value with the selected level of significance. Conclude that the test data are from the same population if the result of the TTEST calculation is greater than the selected level of significance. Conclude that the test data are not from the same population if the result of TTEST calculation is less than the selected level of confidence.

6.3 The Department has software to perform F-test and t-test analysis. The software calculates the F-test and t-test values and compares them with the selected level of significance. The software automatically indicates if the test data are either from the same population or not using appropriate label or designation.

7. F AND t-TESTS SAMPLE CALCULATIONS USING COMBINED MANUAL AND COMPUTER-ASSISTED PROCEDURES, AND MICROSOFT EXCEL STATISTICAL FUNCTIONS

7.1 Combined Manual and Computer-Assisted Procedures

7.1.1 Split Samples (Paired Observations)

This example will demonstrate the combined manual and computer-assisted procedures to conduct F-test and paired t-test calculations for split samples.

Problem Statement:

Using the ignition furnace method to determine the asphalt content of a mix, the following test results were obtained for split samples A and B:

Test Number	Sample A	Sample B
1	4.79	4.88
2	4.74	4.84
3	4.41	4.82
4	4.77	4.71
5	4.58	4.79

Using F-test and t-test method, determine if sample A and sample B are from the same population.

Solution:

a) Select the level of significance (α) at which to evaluate the F and t statistics. Use the level specified in the project special provisions. Assuming that α =0.05 is specified, determine the F-value using Eq. 3.4 which derives its value from Eq. 3.3 (variance), Eq. 3.2 (standard deviation), and Eq. 3.1 (mean) in each data set.

	Sample A	Sample B
Arithmetic Average	4.66	4.81
Standard Deviation	0.161	0.064
Variance	0.02607	0.00407

F-value (larger variance is divided by smaller variance, 0.02607 / 0.00407)	6.40541
Degrees of freedom, n-1, (numerator, 5-1)	4
Degrees of freedom, n-1, (denominator, 5-1)	4

b) Calculate the α-value using the Microsoft Excel function FDIST [calculated F-value, numerator degrees of freedom, denominator degrees of freedom] which translates into α = FDIST [6.40541, 4, 4] and yields α/2-value = 0.0498 (one-tailed test result). Multiply this value by 2 to give the two-tailed α-value = 0.100.

c) Compare this calculated α-value with the level of significance, = 0.05, as specified in Step 2. Since the calculated α-value is greater than the selected α level (0.100 > 0.05), **conclude that the sample variances are equal and proceed to conducting a t-test.**

d) Calculate the difference (D_i) between sample A and sample B for each set of test number. Calculate the arithmetic average ($\overline{D}$) of the differences between sample A and sample B (last column). Also, calculate the standard deviation of the differences between sample A and sample B. Calculate the absolute t-value using Equations 3.5 and 3.6. The sample size for observed differences is n = 5.

Test Number	Sample A	Sample B	(Sample A – Sample B)
1	4.79	4.88	-0.09
2	4.74	4.84	-0.10
3	4.41	4.82	-0.41
4	4.77	4.71	0.06
5	4.58	4.79	-0.21
	Average (D) of A-B		-0.15
	Standard Deviation, S_d		0.1742

Calculating the absolute value of t yields $|t| = -0.15/(0.1742/\sqrt{5}) = 1.925$

e) Calculate the α-value using the Microsoft Excel function TDIST [calculated t-value, degrees of freedom, number of tails] which translates into α = TDIST [1.925,4,2] and yields an α-value = 0.126. Since this value is larger than the selected level of significance, α = 0.05, **conclude that the two data sets are from the same population.**

7.1.2 Independent Samples (Non-paired Observations)

This example will demonstrate the combined manual and computer-assisted procedures to conduct F-test and t-test calculations for independent samples.

Problem Statement:

Using the ignition furnace method to determine the asphalt content of a mix, the following test results were obtained for independent samples A and B:

Test Number	Sample A	Sample B
1	4.65	4.75
2	4.84	4.79
3	4.59	4.74
4	4.75	4.41
5	4.63	4.77
6	4.75	4.58
7	4.58	4.81
8	4.82	
9	4.86	
10	4.70	
11	4.60	
12	4.77	
13	4.65	
14	4.80	

Using F-test and t-test method, determine if sample A and sample B are from the same population.

Solution:

a) Select the level of significance (α) at which to evaluate the F and t statistics. Use the level specified in the project special provisions. Assuming that α =0.05 is specified, determine the F-value using Eq. 3.4 which derives its value from Eq. 3.3 (variance), Eq. 3.2 (standard deviation), and Eq. 3.1 (mean) in each data set.

	Sample A	Sample B
Arithmetic Average	4.71	4.69
Standard Deviation	0.0974	0.1457
Variance	0.009486	0.021224

F-value (larger variance is divided by smaller variance, 0.021224 / 0.009486)	2.23732
Degrees of freedom, n-1, (numerator, 7-1)	6
Degrees of freedom, n-1, (denominator, 14-1)	13

b) Calculate the α-value using the Microsoft Excel function FDIST [calculated F-value, numerator degrees of freedom, denominator degrees of freedom] which translates into α = FDIST [2.23732, 6, 13] and yields α/2-value = 0.1054 (one-tailed test result). Multiply this value by 2 to give the two-tailed α-value = 0.211.

c) Compare this calculated α-value with the level of significance, α = 0.05, as specified in Step 2. Since the calculated α-value is greater than the selected α level (0.211 > 0.05), **conclude that the sample variances are equal and proceed to conducting a t-test.**

d) Calculate the arithmetic averages ($\overline{X}_1$ and $\overline{X}_2$) and variances (S_1^2 and S_2^2) for each data set. Calculate the pooled variance, Sp for both data sets using Eq. 3.9. Calculate the absolute t-value using Eq. 3.8. The sample size for sample A is n = 14 and the sample size for sample B is n = 7.

Test Number	Sample A		Sample B
1	4.65		4.75
2	4.84		4.79
3	4.59		4.74
4	4.75		4.41
5	4.63		4.77
6	4.75		4.58
7	4.58		4.81
8	4.82		
9	4.86		
10	4.70		
11	4.60		
12	4.77		
13	4.65		
14	4.80		
Arithmetic Average ($\overline{X}_1$ or $\overline{X}_2$)	4.71		4.69
Variance (S_1^2 or S_2^2)	0.00949		0.02122
Pooled Variance (S_p)		0.11486	

Calculating the absolute value of t yields:

$$|t| = (4.71 - 4.69)/(0.11486)(\sqrt{(1/14)+(1/7)})$$
$$|t| = 0.38959$$

e) Calculate the α-value using the Microsoft Excel function TDIST [calculated t-value, degrees of freedom, number of tails] which translates into α = TDIST [0.38959,19,2] and yields an α-value = 0.701. Since this value is larger than the selected level of significance, α = 0.05, **conclude that the two data sets are from the same population.**

7.2 Microsoft Excel Statistical Functions

7.2.1 Split Samples (Paired Observations)

This example will demonstrate the use of Microsoft Excel statistical functions to conduct F-test and paired t-test calculations for split samples.

Problem Statement:

Using the ignition furnace method to determine the asphalt content of a mix, the following test results were obtained for split samples A and B:

Test Number	Sample A	Sample B
1	4.79	4.88
2	4.74	4.84
3	4.41	4.82
4	4.77	4.71
5	4.58	4.79

Using F-test and t-test method, determine if sample A and sample B are from the same population.

<u>Solution:</u>

a) Select the level of significance (α) at which to evaluate the F and t statistics. Use the level specified in the project special provisions. Assuming that α =0.05 is specified, determine the α-value by using the Microsoft Excel function FTEST [array1, array2]. Array1 and array2 represent the first and second sets of data respectively. They can be interchanged and still yield the same result. Applying the FTEST function gives α = FTEST ({4.79, 4.74, 4.41, 4.77, 4.58}, {4.88, 4.84, 4.82, 4.71, 4.79}) which yields 0.100. Since this value is greater than the specified level of significance (0.100 > 0.05), **conclude that the sample variances are equal and proceed to conducting a t-test.**

b) Calculate the α-value for the t distribution using the Microsoft Excel function TTEST [array1, array2, tails, type]. Array1 and array2 parameters are as defined above. Tails = 2, for a two-tailed test and type = 1, for paired observation (split samples). Applying the TTEST function gives = TTEST ({4.79, 4.74, 4.41, 4.77, 4.58}, {4.88, 4.84, 4.82, 4.71, 4.79}, 2, 1) which yields 0.126. Since this value is larger than the specified level of significance, α = 0.05, **conclude that the two data sets are from the same population.**

7.2.2 Independent Samples

The same procedure as above should apply to conducting F-test and t-tests for independent samples. In the calculation routine, the only difference of independent samples from split samples is the value of the type parameter in the TTEST function is type = 2 instead of type = 1.

APPENDIX I: SURVEY OF QA/QC SOFTWARE

State	Software	How Acquired?	Comments from Respondent	Related Info (M&T obtained)
Oklahoma	1. LIMS	-Online Vendor	http://online-lims.com/ins_onlqc.htm	Offers exportable data to Excel, automatic linking to online worksheets and online LIMS reports, includes user interactive charts to aid in QC decision process, and a user configurable SRM (Standard Reference Model) database.
	2. LABsys	-Online Vendor	http://ontekonline.com/e/download.asp#	A single application that allows you to collect QC test results as well as manage lab equipment and projects in a database.
	3. Asphalt-It	-Online Vendor	http://qcqa.com/product/product.asp?prodPID=338	Design and Control Asphalt paving mixtures only.
	4. SQCpack	-Online Vendor	http://pqsystems.com *Note: Does not indicate whether Oklahoma has used any of these programs	Creates charts, histograms, and Pareto charts. Can easily adapt to different types of data sets.
New York	1. Access/Excel Hybrid	-In House	Currently in use by the NYSDOT.	
	2. Clough-Harbour Assoc.	-Vendor	Viewed demonstration of the web-based QC/QA program developed for the MassDOT. (Contact: William Ashford of Concord, MA 978-369-2890)	
Louisiana	None	-		
New Jersey	Unspecified In-house	-In House	Applications developed by Richard Weed rmweed@usa.net	
New Hampshire	None	-		
South Carolina	None	-		

Virginia	1. Access/Excel Hybrid	-In House	Developed a QC/QA program using Access and an SQL server.	
	2. Other States	-In House (other state)	Mentioned work Texas and Washington had done and linked to a web site http://fhwa.dot.gov/pavement/concrete/pubs/07019/chapt2.cfm	
Washington	1. ATSER	-Not Given	Used on one design-build and found the acceptable. No information given.	
	2. CATS	-In House	Construction Audit Tracking System used to inspect design-builds and track non-conformance issues	
	3. SAMS	-In House	Statistical Analysis of Materials System used in design-builders QC/QA analysis and Quality Verification by applying f and T-tests. *Note: To learn more about CATS and SAMS contact Bob Briggs 360-709-5411.	
FHWA	1. QL-PAY	-Public Domain	Computes pay factors based on PWL or PD and compares contractor test results and agency results using f and T-tests for both paired and split samples. Can plot run charts and histograms.	
	2. PWL-PAY	-Public Domain (not yet available)	Input PWL acceptance criteria in form of pay table or pay equations and using f and T-tests for paired/split as well as independent samples. Check skewness and Kurtosis. Can provide a risk assessment of the specification (alpha and beat risks). Currently not available.	
	3. Tie-in to Access		Recommends contacting Matt Witzac at the University of Arizona or someone at ARA and CenterSpace (http://www.centerspace.net) that can tie in to any stand alone program	CenterSpace Software runs on the .NET platform. Includes matrix and vector classes, linear algebra, random number generators, numerical integration methods, interpolation, statistics, biostatistics, multiple linear regression, analysis of variance, optimization and object oriented interfaces.

Texas	1. Excel/Access Hybrid	-In House with online graphical tools	Use in house analysis utilizing Kaleidagraph (http://synergy.com/) and SPC (http://pister.com/SPC.html) as graphical analysis tools.	
Arkansas	Unspecified In House	-In House		
Maine	Excel	-In House		
Florida	Unspecified	-In House	Still under development	
Kentucky	None	-		

Summary Highlights:

Number of Responses = 13 States
(1) Number of States using In House = 7 States (54%)
(2) Number of States not implemented = 6 States (46%)
(3) Number of States using Vendor software = 1 State (9%)

*Notes:

Washington used Vendor software for a Design/Build project and uses in-house software for the normal Design/Bid/Build projects, therefore percentages will sum above 100%.

FHWA has developed two programs open to Public Domain.