ITS UCDAVIS INSTITUTE OF TRANSPORTATION STUDIES

Research Report – UCD-ITS-RR-11-28

Quality Assurance for Caltrans Bridge Shear Retrofit Projects

January 2011

B.-W. Tsai
C.L. Monismith

Institute of Transportation Studies ◦ University of California, Davis
One Shields Avenue ◦ Davis, California 95616
PHONE (530) 752-6548 ◦ FAX (530) 752-6572
www.its.ucdavis.edu

January 2011
Research Report: UCPRC-RR-2011-01

Quality Assurance for Caltrans Bridge Shear Retrofit Projects

Authors:
B.-W. Tsai and C.L. Monismith

PREPARED FOR:

California Department of Transportation
Division of Engineering Services
Office of Structure Construction

PREPARED BY:

University of California
Pavement Research Center
UC Davis, UC Berkeley

DOCUMENT RETRIEVAL PAGE	**Research Report Number: UCPRC-RR-2011-01**

Title: Quality Assurance for Caltrans Bridge Shear Retrofit Projects

Authors: Bor-Wen Tsai and Carl L. Monismith

Prepared for:	**FHWA No.:**	**Work Submitted:**	**Date:**
Caltrans Division of Engineering Services	CA139999A	January 25, 2011	January 2011

Strategic Plan Element No.:	**Status:**	**Version No:**
2.3	Stage 6, final version	1

Abstract:
A quality assurance (QA) process is proposed to prevent the possibility of legal issues for a multibridge retrofit project recently conducted by the California Department of Transportation (Caltrans) to increase the horizontal shear resistance of decks. Providing a solution requires determining an appropriate number of samples with the corresponding sampling scheme to ensure 95% compliance with the specification requirements. The determination of the appropriate sample size must recognize the practical considerations of cost, time, statistical simulation results, and the binomial distribution theory. It is not uncommon for many agencies to base quality assurance on just three samples. Using a binomial distribution, a discussion is also presented why it is not appropriate to take this limited number of samples for quality assurance. With the selection of sample size for each bridge, a representative sampling scheme that is random and unbiased was developed using the concept of uniform design as sampling strategy. The recommended acceptance criteria are specified based on the hypothesis testing results with the normal approximation of a binomial distribution.

Keywords: Quality assurance, sample size, sampling scheme, binomial distribution, Uniform Design.

Proposals for implementation:

Related documents: See References

Signatures:

B.-W. Tsai **First Author**	C.L. Monismith **Technical Review**	D. Spinner **Editor**	C. L. Monismith **Principal Investigator**	T. J. Holland **Caltrans Contract Manager**

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in Braille, large print, audiocassette, or compact disk. To obtain a copy of this document in one of these alternate formats, please contact: the Division of Research and Innovation, MS-83, California Department of Transportation, P.O. Box 942873, Sacramento, CA 94273-0001.

PROJECT OBJECTIVES

1. In order to prevent potential legal issues arising from a recent bridge retrofit project conducted by the California Department of Transportation (Caltrans), provide a quality assurance (QA) solution that requires determination of an appropriate number of samples and a corresponding scheme to ensure 95 percent compliance with specification requirements.
2. For each bridge, provide a selected sample size and a representative sampling scheme that is random and unbiased, and which uses uniform design as a sampling strategy.
3. Recommend final acceptance criteria/specifications for each bridge based on the hypothesis testing results with the normal approximation of a binomial distribution.

ACKNOWLEDGMENTS

This report describes research activities requested and sponsored by the California Department of Transportation (Caltrans), Division of Engineering Services, Office of Structure Construction. Caltrans sponsorship is gratefully acknowledged. The authors also thank Joe Holland, John Babcock, and Rosme Aguilar of Caltrans for their support in completion of this report.

EXECUTIVE SUMMARY

A multibridge retrofit project was recently conducted by the California Department of Transportation (Caltrans) to increase the horizontal shear resistance of the decks. This required drilling and bonding #5 rebar as dowels in 6 inch-deep holes along the center line of the existing girders of eight bridges. The contractor performing the work was required to completely fill the area around the dowels with epoxy in order to provide sufficient bonding to meet the specification. Unfortunately, an inspection of limited sample size performed after completion of the dowel installation process revealed that many had been improperly grouted. This resulted in rejection of the work on seven of the eight bridges and a request from Caltrans that the contractor repair the rejected sections. In addition, in order to prevent the possibility of legal issues, Caltrans decided to establish a testing procedure that would statistically determine the reliability of the work, and asked the University of California Pavement Research Center (UCPRC) to undertake the task.

Providing the solution requires determination of an appropriate number of samples (sample size) with a corresponding sampling scheme to ensure 95 percent compliance with the specification requirement that dowels be either fully bonded or not, with none partially bonded. Essentially, this becomes a case of 0 (failure) or 1 (success) with an inherent population (or contractor) proportion *p*. The approach selected was to assume each sampled dowel is a Bernoulli random variable and each dowel inspection is a Bernoulli trial. The count of successes/failures from *n* Bernoulli trials (i.e., sample size = *n*) is designated as a binomial random variable (*X*). The probability associated with a specific outcome $X = x$ is given by a binomial density function expressed as follows:

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The research objectives described in this report are: (1) to determine the appropriate sample size, recognizing the practical considerations of cost, time, statistical simulation results, and the normal approximation of a binomial distribution; (2) to develop the most representative sampling scheme with the specified sample size; and, (3) to provide performance specifications (or acceptance criteria) for each bridge. To obtain a solution, the associated sample size determination, hypothesis testing, and performance specification were developed based on binomial distribution theory and the normal approximation of a binominal distribution.

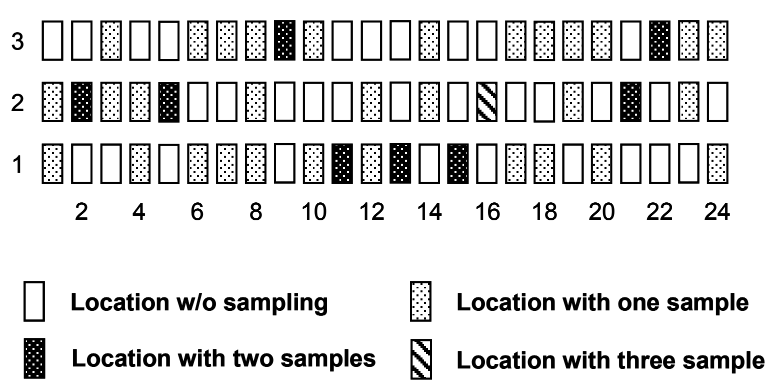
Determination of sample size for quality assurance (QA) is based primarily on an acceptable error level $E = |\hat{p} - p_0|$ for a performance parameter specified by the agency. It is necessary to have the sample size "large enough" so that the sampling error will be within a reasonable level of accuracy. If the sample size is too small, it is not worthwhile gathering data; the results will tend to be too imprecise to be of value. To investigate the sample size effect, a binomial sampling simulation was conducted. The binomial population was randomly generated based on the quality of the contractor (proportion) and the assumption of 8 dowels for each of 642 locations, i.e., a binomial population with 5,136 dowels. Results of this extensive statistical simulation suggest that there is a critical point of diminishing returns (probably around 100 ~ 200, which is the "large enough" sample size range for the binomial distribution) where increasing sample size provides little benefit.

It is not uncommon for agencies to base QA on three samples. However, a discussion using binomial distribution is presented showing why it is inappropriate to take only this number of samples for quality assurance. For example, basing a large project on only three samples provides the agency with insufficient power to reject the null hypothesis—given that this hypothesis is false unless a project delivered is of such poor quality that the agency is confident it can reject it.

For this project, considering the time and cost that the agency may be willing to spend, it is recommended that one-tenth of the number of dowels for each bridge should be sampled; that is, the quality assurance is based on each bridge rather than based on the whole project. The sample size of each bridge is summarized as follows:

Bridge Name	**# of Dowels per Bridge**	**# of Samples per Bridge**
Van Winkle Wash Bridges (Left [Lt.] and Right [Rt.])	534	50
Haller Wash Bridges (Lt. and Rt.)	282	30
Rojo Wash Bridges (Lt. and Rt.)	318	30
Clipper Valley Wash Bridges (Lt. and Rt.)	993	100

With the selection of sample size for each bridge, a representative sampling scheme that is random and unbiased was developed. This made use of uniform design (UD) as sampling strategy to ensure that the most representative sampling scheme can be achieved and applied to each bridge; for example, the sampling scheme for the Van Winkle Wash Bridge (Rt.) is illustrated in the following figure.



For this sampling scheme, the following recommendations should be adhered to:

1. The contractor must follow the specified sampling scheme; if it is determined that changes to the sampling scheme are necessary, the agency (Caltrans) must grant permission for them to be made.
2. The dowel (or dowels) sampled per location must be randomly selected with the approval of the agency (Caltrans).
3. The agency (Caltrans) is responsible for inspecting whether the dowels are fully bonded or not fully bonded.

The recommended acceptance criteria are based on hypothesis testing results with the normal approximation of a binomial distribution. The hypothesis testing of the null hypothesis $H_0 : p = 0.95$ and an alternative hypothesis $H_1 : p < 0.95$ with the conventional α value of 0.05 is utilized to develop the acceptance criteria at various sample sizes (*n* = 30, 50, 100). The conventional power level 0.8, where power is defined as the probability to correctly reject H_0 if H_0 is not true, is specified to establish the acceptance criteria. Accordingly, the acceptance criterion, $Y \le m$, is established for each bridge, where *Y* is the count of failures and *m* is the specified upper bound with sample size *n*. If $Y > m$, then the project is rejected and a power level at least 0.8 is guaranteed for the agency; otherwise, if $Y \le m$, then the project is not going to be rejected. The acceptance criteria for each bridge are summarized as follows:

Bridge Name	No. of Locations	No. of Dowels	Sample Size	Acceptance Criterion	
				Proportion	Count of Failures
Van Winkle Wash Bridges (Rt. And Lt.)	72	534	50	**P ≥ 0.858**	**Y ≤ 7**
Haller Wash Bridges (Rt. And Lt.)	42	282	30	**P ≥ 0.826**	**Y ≤ 5**
Rojo Wash Bridges (Rt. And Lt.)	45	318	30	**P ≥ 0.826**	**Y ≤ 5**
Clipper Valley Wash Bridges (Rt. and Lt.)	162	993	100	**P ≥ 0.888**	**Y ≤ 11**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Problem Statement

A multibridge retrofit project was recently conducted by the California Department of Transportation (Caltrans) to increase the horizontal shear resistance of the decks. The retrofit required drilling and bonding #5 rebar dowels in 6-inch deep holes along the center line of the existing girders of eight bridges. The contractor performing the work was required to completely fill the area around the dowels with epoxy in order to provide sufficient bonding to meet the specification. Unfortunately, an inspection of limited sample size performed after completion of the dowel installation process revealed that many had been improperly grouted. This resulted in rejection of the work on seven of the eight bridges and a request from Caltrans that the contractor repair the rejected sections. In addition, in order to prevent the possibility of legal issues, Caltrans decided to establish a testing procedure that would statistically determine the reliability of the work, and asked the University of California Pavement Research Center (UCPRC) to undertake the task.

The problem lies in determining an appropriate number of samples (sample size) with a corresponding sampling scheme in order to ensure 95 percent compliance with the specification requirement that dowels be completely bonded, i.e., the dowels should be either fully bonded or unbonded, and not partially bonded.

1.2 Objectives

The research objectives described in this report are: (1) to determine the appropriate sample size, recognizing the practical considerations of cost, time, statistical simulation results, and the normal approximation of a binomial distribution; (2) to develop the most representative sampling scheme with the specified sample size; and, (3) to provide performance specifications (or acceptance criteria) for each bridge.

1.3 Background

The eight box-girder bridges to be retrofitted were located on Interstate Highway 40 (I-40, three lanes in each direction) in San Bernardino County (Caltrans District 8) from 7.0 miles east of the Kelbaker Road undercrossing to the Clipper Valley Wash Bridge (Figure D.1). The bridges included were the Van Winkle Wash (Right [Rt.] and Left [Lt.], Figure D.2), Haller Wash (Rt. and Lt., Figure D.3), Rojo Wash (Rt. And Lt., Figure D.4), and Clipper Valley Wash (Rt. and Lt., Figure D.5) bridges.

The main purpose of the dowel bar retrofit project was to increase horizontal shear resistance at the deck girder joint of these eight box-girder bridges by drilling and bonding #5 rebar dowels through the joint between the deck and girder. The dowel bars had to be fully epoxy-encased, and no partially epoxy-encased dowels were permitted.

Each bridge has three girders that required retrofitting, which was accomplished by following this sequence of tasks: (1) removal of alternating 8 ft.-by-16 in. wide pieces of the deck (also designated as *locations* in this report) of the deck within the work area (an enclosure of two lanes), (2) drilling and bonding the dowels, and (3) replacement of the deck concrete. Construction staging included two stages (Figure 1.1), each of which consisted of two phases. For each stage, the work area enclosed two lanes. Stage 1A started with the inner two girders and followed a zigzag construction pattern; Stage 1B fixed the rest of the alternating pieces (locations) of the inner two girders, also following a zigzag construction pattern. Retrofit of the outermost girder included Stages 2A and 2B, following the alternating pattern also shown in Figure 1.1 and Figure D.8. Appendix D shows details of the girder repairs (Figure D.6 and Figure D.7) and the temporary deck access opening (Figure D.9). Table 1.1 contains a summary of the bridge locations and the number of dowel bars placed in the various construction stages for each bridge.

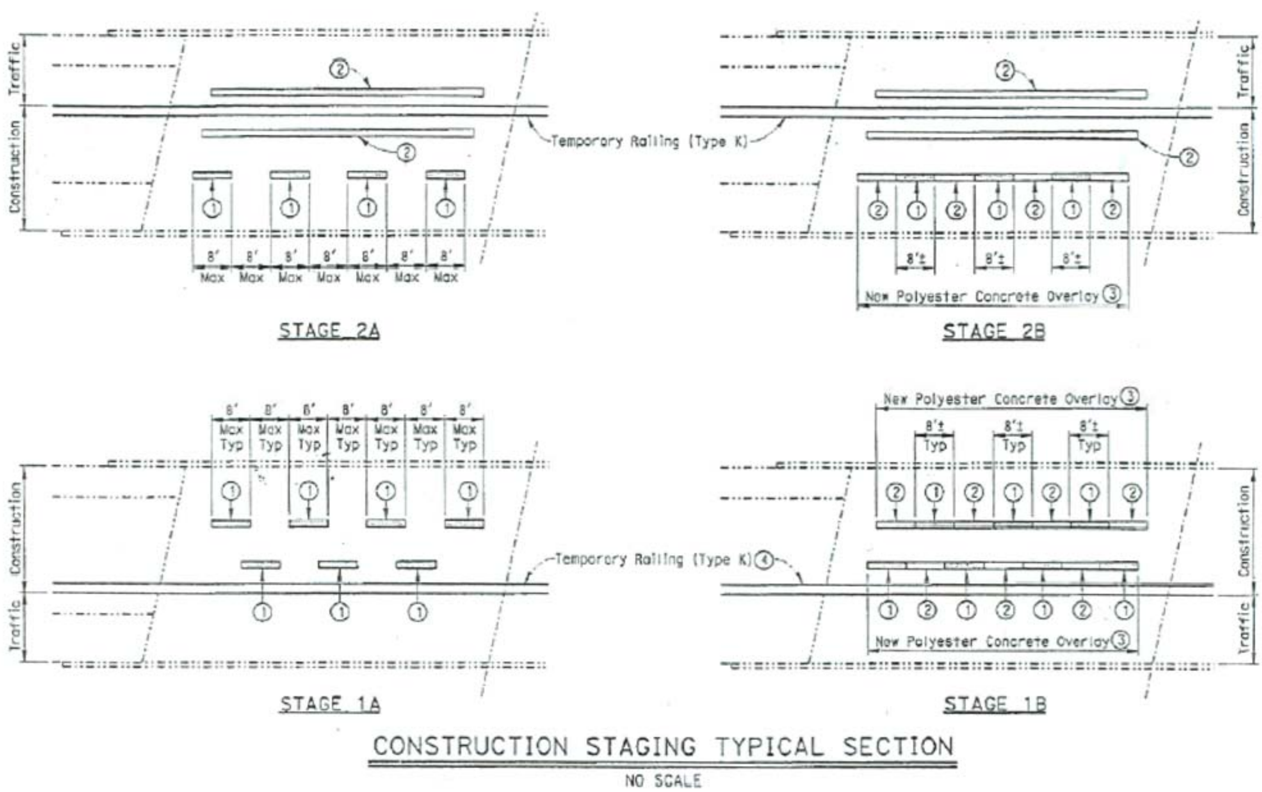


Figure 1.1: Typical bridge construction staging.

Table 1.1: Summary of Bridge Locations and Number of Dowels at Various Construction Stages

Bridge Name	Stage #	Span #	Girder #	# of Locations	# of Dowels
Van Winkle Wash Bridge (Rt.) (with polyester overlay)	1A	1,2,3,4	2,3	24	178
	1B	1,2,3,4	2,3	24	178
	2A	1,2,3,4	4	12	178
	2B	1,2,3,4	4	12	
			Σ	***72***	***534***
Van Winkle Wash Bridge (Lt.) (with polyester overlay)	1A	1,2,3,4	3,4	24	178
	1B	1,2,3,4	3,4	24	178
	2A	1,2,3,4	2	12	178
	2B	1,2,3,4	2	12	
			Σ	***72***	***534***
Haller Wash Bridge (Rt.) (without polyester overlay)	1A	1,2	2,3	14	94
	1B	1,2	2,3	14	94
	2A	1,2	4	7	94
	2B	1,2	4	7	
			Σ	***42***	***282***
Haller Wash Bridge (Lt.) (without polyester overlay)	1A	1,2	2,3	14	94
	1B	1,2	2,3	14	94
	2A	1,2	4	7	94
	2B	1,2	4	7	
			Σ	***42***	***282***
Rojo Wash Bridge (Rt.) (with polyester overlay)	1A	1,2	2,3	15	106
	1B	1,2	2,3	15	106
	2A	1,2	4	8	106
	2B	1,2	4	7	
			Σ	***45***	***318***
Rojo Wash Bridge (Lt.) (with polyester overlay)	1A	1,2	4,3	15	106
	1B	1,2	4,3	15	106
	2A	1,2	2	8	106
	2B	1,2	2	7	
			Σ	***45***	***318***
Clipper Valley Wash Bridge (Rt.) (with polyester overlay)	1A	1,2,3,4,5,6,7,8,9	2,3	54	331
	1B	1,2,3,4,5,6,7,8,9	2,3	54	331
	2A	1,2,3,4,5,6,7,8,9	4	27	331
	2B	1,2,3,4,5,6,7,8,9	4	27	
			Σ	***162***	***993***
Clipper Valley Wash Bridge (Lt.) (with polyester overlay)	1A	1,2,3,4,5,6,7,8,9	4,3	54	331
	1B	1,2,3,4,5,6,7,8,9	4,3	54	331
	2A	1,2,3,4,5,6,7,8,9	2	27	331
	2B	1,2,3,4,5,6,7,8,9	2	27	
			Σ	***162***	***993***
			Total	***642***	***4,254***

2 SAMPLE SIZE, SAMPLING SCHEME, AND ACCEPTANCE CRITERIA OF A QA PROCESS

2.1 Construction of Statistical Hypothesis Testing

For each dowel retrofit, the agency required that each dowel be fully bonded and that none be partially bonded, i.e., a case of 0 (failure) or 1 (success) with an inherent population (or contractor) proportion p. Each sampled dowel is termed a *Bernoulli random variable* and each dowel inspection is termed a *Bernoulli trial.* The count of successes/failures from n Bernoulli trials (i.e., sample size = n) is designated as a binomial random variable (X). The probability associated with a specific outcome $X = x$ is given by a binomial density function $P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ (see Appendix A.1).

The agency required the contractor to ensure 95 percent compliance with the specification requirements that dowels be completely bonded and none be partially bonded. The equivalent of a statistical statement of hypothesis testing based on the binomial distribution with parameters n (sample size) and p (proportion) is then the null hypothesis $H_0 : p = 0.95$ (see Appendix A.2). In this case, the use of alternative hypothesis $H_1 : p < 0.95$ to establish the performance specification of a quality assurance (QA) process seems to be more appropriate than the other two alternative hypotheses: $H_1 : p \neq 0.95$ and $H_1 : p > 0.95$.

A QA process using binomial distribution established for the agency should include the following steps:

1. Determination of sample size,
2. Development of a sampling scheme, and
3. Determination of the acceptance criteria for a QA process.

2.2 Determination of Sample Size

The determination of sample size for QA is primarily based on an acceptable error level $E = |\hat{p} - p_0|$ for a performance parameter specified by the agency, as illustrated in Appendix A. In general, the larger the sample size n, the smaller the sampling error $E = |\hat{p} - p_0|$ tends to be. It is therefore necessary to have the sample size "large enough" so that the sampling error will tend to be at a reasonable level of accuracy. If the sample size is too small, there is not much point in gathering the data because the results will tend to be too imprecise to be of use.

A binomial sampling simulation was conducted to investigate the effects of the quality of contractor (proportion), samples per location, number of locations, and number of total sample size (as shown in Appendix B). The binomial population was randomly generated based on the quality of contractor (proportion) and the assumption of 8 rebar dowels for each of 642 locations, i.e., a binomial population with 5,136 rebar dowels. The factors and their corresponding factor levels in the experimental design of this sampling simulation include: (1) factor **Contractor**, i.e., quality of contractor, with four proportion levels: 0.85, 0.90, 0.95, and 0.98; (2) factor **SamplesPerLocation** with four levels: 1, 2, 3, and 4; and (3) factor **Locations** with 7 levels: 10, 20, 50, 100, 200, 500, and 642. Each of the 112 cases (4 × 4 × 7) was simulated 500 times. For each simulation, the proportion was calculated; hence, the proportion distribution was generated after 500 simulations. The standard deviation *S* was used to characterize the dispersion of the proportion distribution. Figure 2.1 (also shown in Appendix B.1, Figure B.2d) illustrates the simulation results, in terms of box plots, of the standard deviation *S* versus **TotalSamples**, which is the product of the two factors **Locations** and **SamplePerLocation**. It is apparent that there is a critical point of diminishing return (probably around 100 ~ 200, which is the "large enough" sample size range for the binomial distribution) where increasing sample size provides little benefit.

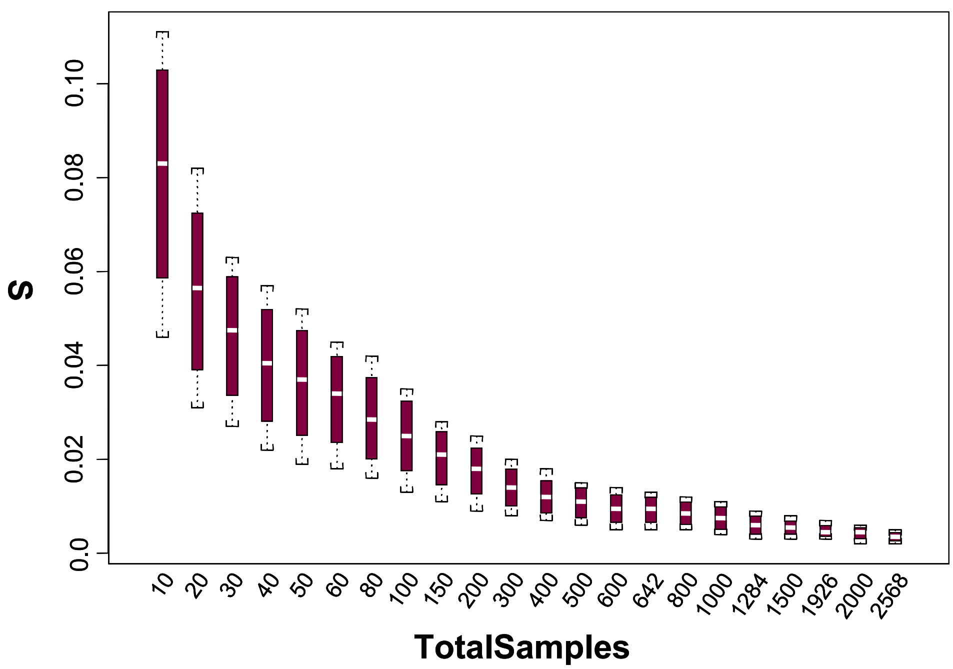


Figure 2.1: Sampling simulation results of standard deviation *S* versus total sample size.

The normal approximation for a discrete binomial distribution was applied in developing the acceptance criteria of a QA process. A frequently used rule of thumb *(1)* is that the approximation is reasonable when $np > 5$ and $n(1-p) > 5$, which is especially appropriate for large values of *n*. Accordingly, if $p = 0.95$, then *n* has to be 100 to fulfill the rule of thumb (Appendix A.2.1).

It is not uncommon for agencies to base QA on three samples. However, a discussion using binomial distribution is presented in Appendix B.2 to show why it is inappropriate to take only this number of samples for quality assurance. For example, basing a large project on only three samples provides the agency with insufficient power to reject the null hypothesis—given that this hypothesis is false unless a project delivered is of such poor quality that the agency is confident it can reject it.

A sample size of 100 was determined to be a reasonable minimum based on the foregoing discussion. However, 100 samples is more than one-third of the number of dowels of the Haller Wash Bridges (Lt. and Rt.; each bridge has 282 dowels) and about one-tenth of the number of dowels of the Clipper Valley Wash Bridges (Lt. and Rt.; each bridge has 993 dowels). In consideration of the time and cost to the agency, it is suggested that one-tenth of the number of dowels for each bridge should be sampled.

Based on the previous discussion, the following recommendations are made for sample size determination:

Recommendations

1. **Sample size should be determined for each bridge rather than on the whole project.**
2. **Approximately one-tenth of the number of dowels of each bridge should be obtained for quality assurance. Accordingly, the sample size for each bridge is summarized as follows:**

Bridge Name	# of Dowels per Bridge	# of Samples per Bridge
Van Winkle Wash Bridges (Lt. and Rt.)	534	50
Haller Wash Bridges (Lt. and Rt.)	282	30
Rojo Wash Bridges (Lt. and Rt.)	318	30
Clipper Valley Wash Bridges (Lt. and Rt.)	993	100

2.3 Development of a Sampling Scheme

After the sample size for each bridge was determined, the next step was to develop the most representative random and unbiased sampling scheme. Thus *Uniform Design* (UD)—which ensures that the most representative sampling scheme can be achieved—was applied as a sampling strategy to each bridge (see Appendix C).

Generally speaking, uniform design is a space-filling experimental design that allocates experimental points uniformly scattered in the domain. The fundamental concept of UD is to choose a set of experimental points with the smallest discrepancy among all the possible designs for a given number of factors and experimental runs *(2,3,4)*.

Given that the strength of UD is that it provides a series of uniformly scattered experimental points over the domain, this homogeneity in two factors has physically become the spatial uniformity of sampling from a bridge section in *x* and *y* directions. The application of uniform design to this multibridge retrofit project resulted in the generation of sampling scheme with a UD table for each bridge consisting of pairs of (x, y) coordinates. The unit of the *x*-axis is the number of locations and the unit of the y-axis is the number of girders.

A prospective bridge was divided into $n(X)$ (*x*-direction) × $n(Y)$ (*y*-direction) cells (or locations). The $n(X)$ represents the number of locations in the *x*-direction and the $n(Y)$ is the number of girders in the *y*-direction. N points (sample size) were then assigned to these $n(X) \times n(Y)$ cells. Hence, a sampling scheme was defined by $n(X)$, $n(Y)$, and N. For instance, *x24y3n50* (as illustrated in Figure 2.2 and Figure C.1) represents 50 samples that were assigned to 50 cells of the 24 × 3 cells. It should be noted that it is possible to assign more than one sample per sampled location.

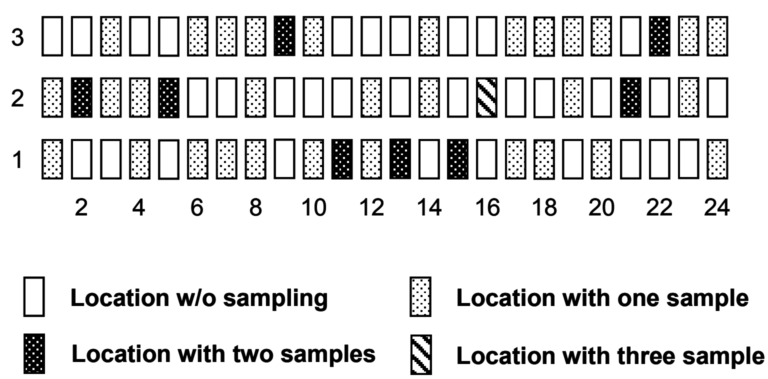


Figure 2.2: Sampling scheme for the Van Winkle Wash Bridge (Rt.) (*x24y3n50*).

The UD table not only provides the most representative sampling scheme but it also provides the agency an unbiased and random sampling scheme that the contractor can follow in the quality assurance process. The bridge sampling schemes generated by UD tables are plotted in Appendix C.2. In addition to the specified sampling scheme for each bridge, the following recommendations are made in formulating the sampling scheme:

Recommendations
1. **The specified sampling scheme must be followed by contractor; if it is determined that changes to the sampling scheme are necessary, the agency (Caltrans) must grant permission for them to be made.** 2. **The dowel (or dowels) sampled per location must be randomly selected with the approval of the agency (Caltrans).** 3. **The agency (Caltrans) is responsible for the inspecting whether or not the dowel bars are fully bonded.**

2.4 Acceptance Criteria of a QA Process

Once the sample size and sampling scheme are determined, development of the acceptance criteria for the QA process is needed to ensure that the acceptance level is obtained. The acceptance criteria are based on the hypothesis testing results with the normal approximation of a binomial distribution. For the sample sizes selected for this project (*n* = 30, 50, and 100), it is demonstrated in Appendix C.3 (Figure C.9) that the normal approximation of a binomial distribution seems to be rational, and the normal approximation is more apparent as the sample size increases. The hypothesis testing of $H_0 : p = 0.95$ and $H_1 : p < 0.95$ was utilized to develop the acceptance criteria. Figure 2.3 plots the relationship of power versus estimated proportion at various sample sizes (*n* = 30, 50, 100) under the hypothesis testing $H_0 : p = 0.95$ and $H_1 : p < 0.95$, and $\alpha = 0.05$.

To establish the acceptance criterion, the agency first has to determine the power level in order to be confident enough to correctly reject H_0 if H_0 is not true. It is recommended that power = 0.8 be specified to establish the acceptance criteria. Let $\hat{p}$ be the estimated proportion and *Y* the count of failures based on the sampling result of a QA process from the specified bridge sampling scheme. For example, the interpretation of Figure 2.3 under the hypothesis testing $H_0 : p = 0.95$ and $H_1 : p < 0.95$ at *n* = 30 and power = 0.8 is that the agency will have at least 0.8 power to reject the null hypothesis $H_0 : p = 0.95$ and favor the alternative hypothesis $H_1 : p < 0.95$ if $\hat{p} < 0.826$ (i.e., $Y > 5$); otherwise, if $p \geq 0.826$ (i.e., $Y \leq 5$), then the agency will have insufficient power to reject the null hypothesis. Therefore, the acceptance criterion is specified such that if there are more than 5 failures, then the agency has a power greater than 0.8 to reject $H_0 : p = 0.95$ and favors $H_1 : p < 0.95$. The acceptance criteria for each bridge are summarized in Table 2.1. As for the acceptance criteria, the following recommendations are made:

Recommendations
1. It is recommended that power = 0.8 be specified to establish the acceptance criterion. 2. The acceptance criterion: $Y \leq m$, where Y is the count of failures and m is the specified lower bound with sample size n. If $Y > m$, then the project should be rejected; however, if $Y \leq m$, then the project need not be rejected.

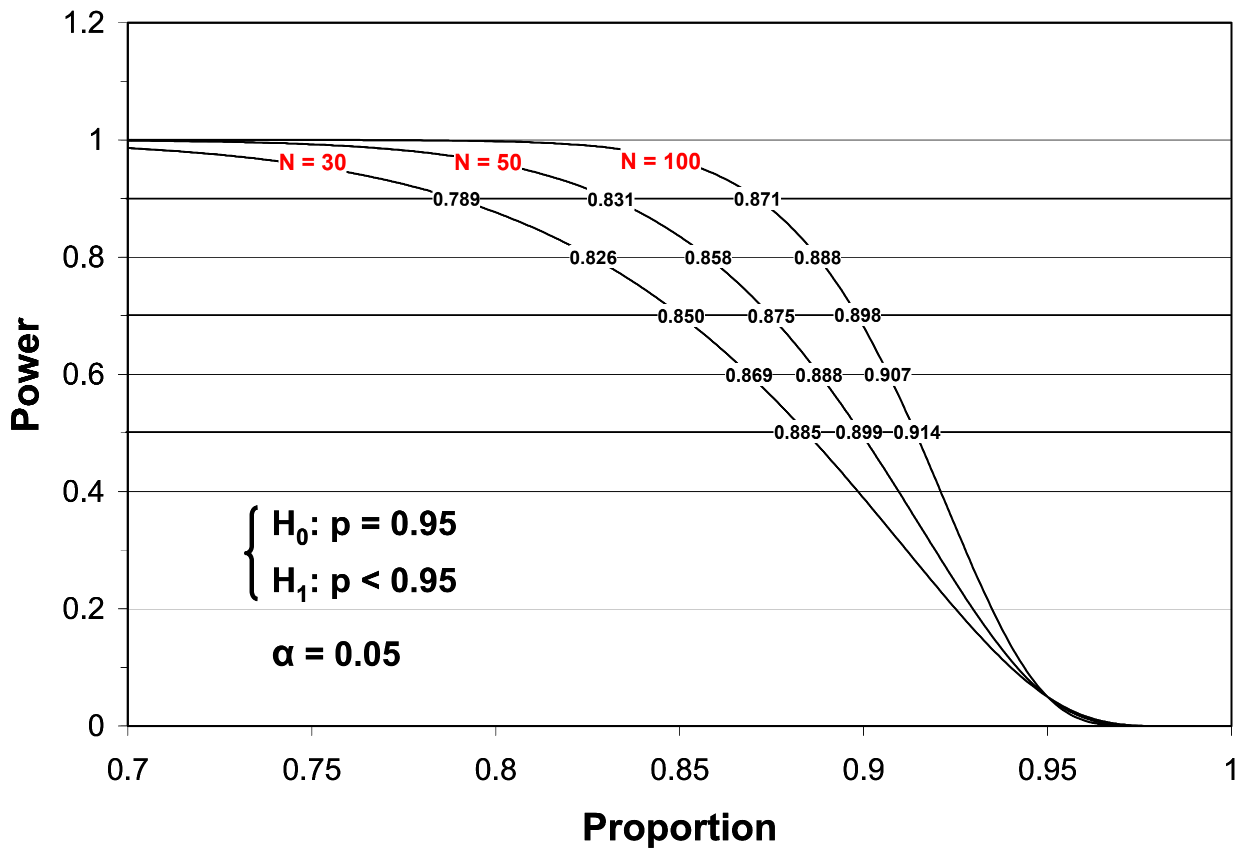


Figure 2.3: The relationship of power versus estimated proportion at various sample sizes (n = 30, 50, and 100) under the hypothesis testing $H_0: p = 0.95$ and $H_1: p < 0.95$ with α = 0.05.

Table 2.1: Acceptance Criteria at Various Power Levels for Each Bridge (also shown in Table C.1)

Bridge Name	No. of Locations	No. of Dowels	Sample Size	Sampling Scheme	Power Level	Acceptance Criterion	
						Proportion	Count of Failures
Van Winkle Bridge (Rt.) (with polyester overlay)	72	534	50	Figure C.1	0.5	P ≥ 0.899	Y ≤ 5
					0.6	P ≥ 0.888	Y ≤ 5
					0.7	P ≥ 0.875	Y ≤ 6
					0.8	**P ≥ 0.858**	**Y ≤ 7**
					0.9	P ≥ 0.831	Y ≤ 8
Van Winkle Bridge (Lt.) (with polyester overlay)	72	534	50	Figure C.2	0.5	P ≥ 0.899	Y ≤ 5
					0.6	P ≥ 0.888	Y ≤ 5
					0.7	P ≥ 0.875	Y ≤ 6
					0.8	**P ≥ 0.858**	**Y ≤ 7**
					0.9	P ≥ 0.831	Y ≤ 8
Haller Bridge (Rt.) (without polyester overlay)	42	282	30	Figure C.3	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	**P ≥ 0.826**	**Y ≤ 5**
					0.9	P ≥ 0.789	Y ≤ 6
Haller Bridge (Lt.) (without polyester overlay)	42	282	30	Figure C.4	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	**P ≥ 0.826**	**Y ≤ 5**
					0.9	P ≥ 0.789	Y ≤ 6
Rojo Bridge (Rt.) (with polyester overlay)	45	318	30	Figure C.5	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	**P ≥ 0.826**	**Y ≤ 5**
					0.9	P ≥ 0.789	Y ≤ 6
Rojo Bridge (Lt.) (with polyester overlay)	45	318	30	Figure C.6	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	**P ≥ 0.826**	**Y ≤ 5**
					0.9	P ≥ 0.789	Y ≤ 6
Clipper Valley Bridge (Rt.) (with polyester overlay)	162	993	100	Figure C.7	0.5	P ≥ 0.914	Y ≤ 8
					0.6	P ≥ 0.907	Y ≤ 9
					0.7	P ≥ 0.898	Y ≤ 10
					0.8	**P ≥ 0.888**	**Y ≤ 11**
					0.9	P ≥ 0.871	Y ≤ 12
Clipper Valley Bridge (Lt.) (with polyester overlay)	162	993	100	Figure C.8	0.5	P ≥ 0.914	Y ≤ 8
					0.6	P ≥ 0.907	Y ≤ 9
					0.7	P ≥ 0.898	Y ≤ 10
					0.8	**P ≥ 0.888**	**Y ≤ 11**
					0.9	P ≥ 0.871	Y ≤ 12

REFERENCES

1. Montgomery, D. C., and G. C. Runger. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, Inc. Fifth Edition, 2010.
2. Fang, K. T. The Uniform Design: Application of Number Theoretical Methods in Experimental Design. *Acta Mathematicae Applagatae Sinica.* Vol. 3, pp. 353 – 372, 1980.
3. Wang, Y., and K. T. Fang. A Note on Uniform Distribution and Experimental Design, *KeXue TongBao*, Vol. 26, pp. 485 – 489, 1981.
4. Fang, Kai-Tai, Dennis K. J. Lin, Peter Winker, and Yong Zhang. Uniform Design: Theory and Application. *Technometrics*. Vol. 42, No. 3, pp. 237 – 248, 2000.

APPENDIX A: FUNDAMENTAL STATISTICS

A.1 Bernoulli Random Variable and Binomial Distribution

A.1.1 Bernoulli Random Variables

The random variable X_i is called a *Bernoulli random variable* if the random variable X_i follows the following probability function

$$p(x) = \begin{cases} p & , X_i = 1 \\ 1 - p & , X_i = 0 \end{cases} \qquad \textbf{(A.1)}$$

That is, X_i takes on value 1 with probability *p* and value 0 with probability $1 - p$. The realization of this random variable is called a *Bernoulli trial.* The sequence of Bernoulli trials $X_1, X_2, \cdots$, is a *Bernoulli process.* The outcome $X_i = 1$ is often referred to "success" or "conforming," and $X_i = 0$ is often called "failure" or "nonconforming." Suppose that a random sample of *n* observations, $X_1, X_2, \cdots, X_n$, is taken from a Bernoulli process with constant probability of success *p*. Then the sum of the sample observations $X = X_1 + X_2 + \cdots + X_n$ follows a binomial distribution with parameters *n* and *p*.

A.1.2 Binomial Distribution

If a random experiment consists of *n* Bernoulli trials (X_i) such that,

1. Each X_i is statistically independent,
2. Each X_i is either 1 or 0 with probability *p* or 1 – *p* respectively, and
3. The probability of success *p* is the same of all X_i values,

then, a binomial random variable $X = X_1 + X_2 + \cdots + X_n$ is defined as the sum of *n* X_i values, i.e., *X* represents the number of trials that result in a success. The probability associated with a specific outcome $X = x$ is given by $P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}$. $\binom{n}{x}$ stands for the total number of different sequences of Bernoulli trials that contain *x* successes and *n* – *x* failures. The name of the distribution is obtained from the binomial expression; for constants *a* and *b*, the binomial expression is $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

According to Equation A.1, the mean and the variance of each X_i can be easily derived as follows:

$$E(X_i) = \sum xf(x) = 1p + 0(1-p) = p \qquad \textbf{(A.2)}$$

$$\begin{aligned} Var(X_i) &= E(X_i - \mu)^2 = \sum (x-\mu)^2 f(x) \\ &= (1-p)^2 p + (0-p)^2 (1-p) \\ &= p(1-p) \end{aligned} \qquad \textbf{(A.3)}$$

Thus, we have

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

$$Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p)$$

A.1.3 An Example of Binomial Distribution

As an example, a fair coin is tossed 5 times and the total number of heads is observed. The probability of a fair coin to have a head (H; value 1) or a tail (T; value 0) is 0.5. The sequence of Bernoulli trials {H, T, H, H, T} is called a Bernoulli process. The probability associated with a specific outcome *X* = *x*, where *x* is the count of heads *x* = 0, 1, 2, 3, 4, and 5, is listed in Table A.1; as a result, the binomial density function can be plotted in Figure A.1.

Table A.1: Probabilities Resulting from Tossing a Fair Coin Five Times

Number of Head Counts	Outcome Set	Probability
$X = 0$	{T, T, T, T, T}	$P(X=0) = \binom{5}{0} 0.5^0 (1-0.5)^{5-0} = \frac{1}{32}$
$X = 1$	{H, T, T, T, T}	$P(X=1) = \binom{5}{1} 0.5^1 (1-0.5)^{5-1} = \frac{5}{32}$
$X = 2$	{H, H, T, T, T}	$P(X=2) = \binom{5}{2} 0.5^2 (1-0.5)^{5-2} = \frac{10}{32}$
$X = 3$	{H, H, H, T, T}	$P(X=3) = \binom{5}{3} 0.5^3 (1-0.5)^{5-3} = \frac{10}{32}$
$X = 4$	{H, H, H, H, T}	$P(X=4) = \binom{5}{4} 0.5^4 (1-0.5)^{5-4} = \frac{5}{32}$
$X = 5$	{H, H, H, H, H}	$P(X=5) = \binom{5}{5} 0.5^5 (1-0.5)^{5-5} = \frac{1}{32}$

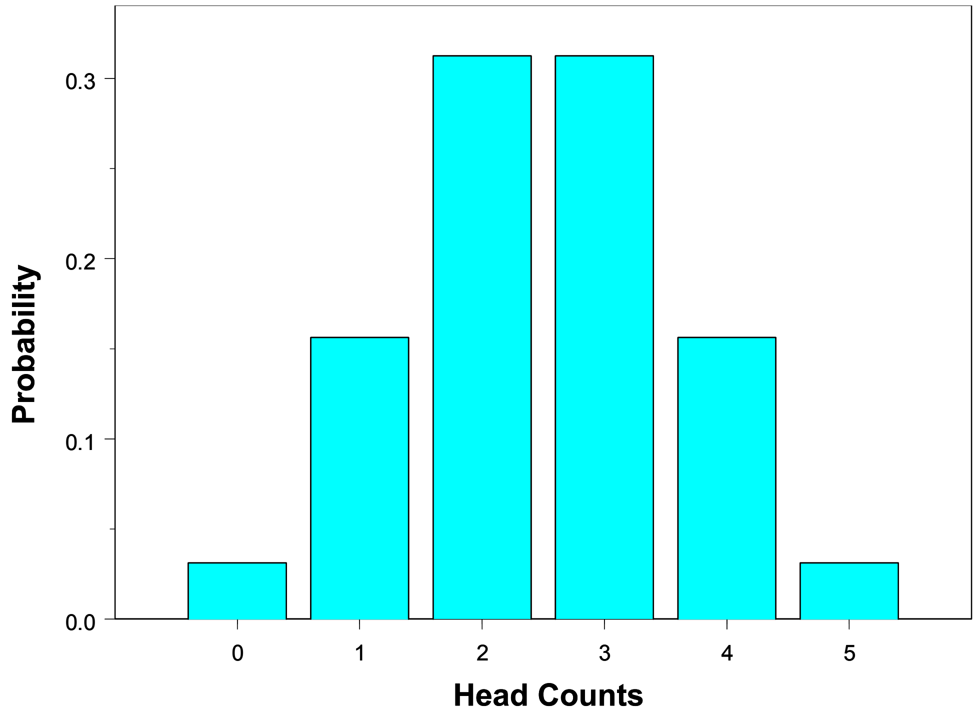


Figure A.1: Probability density function resulting from tossing a fair coin five times (binomial distribution with parameters n = 5 and p = 0.5)

A.2 Large-Sample Confidence Interval of a Population Proportion

A.2.1 Normal Approximation for a Binomial Proportion

The *central limit theorem* can be described as follows:

Central Limit Theorem
If $x_1, x_2, \cdots, x_n$ are independent random variables with mean μ_i and variance σ_i^2, and if $y = x_1 + x_2 + \cdots + x_n$, then the distribution $\dfrac{y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$ approaches the $N(0,1)$ distribution as *n* approaches infinity.

It is recognized that the binomial random variable *X* (≡ *y*) is the sum of independent Bernoulli random variables X_i s with μ_i = p and $\sigma_i^2 = p(1- p)$ for each X_i (Equations A.2 and A.3); hence, its distribution can be approximated by a normal distribution, that is,

Normal Approximation for a Binomial Proportion
If n is larger, the distribution of $$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$ is approximately standard normal, where $\hat{p} = X/n$.

A frequently used rule of thumb is that the approximation is reasonable when $np > 5$ and $n(1-p) > 5$, which is especially appropriate for large values of n. Accordingly, if $p = 0.95$, then n has to be 100 to fulfill the rule of thumb. The following discussion assumes that a binomial proportion can be approximated by the standard normal distribution.

A.2.2 Approximate Confidence Interval on a Binomial Proportion

If $\hat{p}$ is the conforming proportion (proportion of "success") of observations in a random sample of size n, then an approximate $100(1-\alpha)$ percent confidence interval on the conforming proportion p of the population is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \textbf{(A.4)}$$

Thus, a 95% two-sided confidence interval ($\alpha = 0.05$) for the true proportion p can be computed from Equation A.4 with $z_{\alpha/2} = z_{0.025} \cong 1.96$.

Figure A.2 and Figure A.3 plot respectively the 95% and 90% confidence intervals versus conforming proportions at various sample sizes. Table A.2 lists the associated upper and lower bounds at various sample sizes and conforming proportions for both 95% and 90% confidence intervals. Several observations of confidence interval can be addressed as follows:

1. From Equation A.4, it is apparent that $\hat{p} \to p$ when $n \to \infty$; therefore, from Figure A.2 and Figure A.3, it is apparent that the upper and lower bounds of the confidence intervals are symmetrical to $p = \hat{p}$ line.
2. For a specified sample size, the bandwidth (the distance between upper and lower limits) increases as the conforming (or estimated) proportion decreases.
3. For any conforming proportion, the larger the sample size the narrower the bandwidth.

4. The 90% bandwidth is smaller than the 95% bandwidth at a given conforming proportion and sample size. As an example, the 90% confidence interval is (0.914, 0.986) at $\hat{p} = 0.95$ and n = 100 compared with the 95% confidence interval (0.907, 0.993).

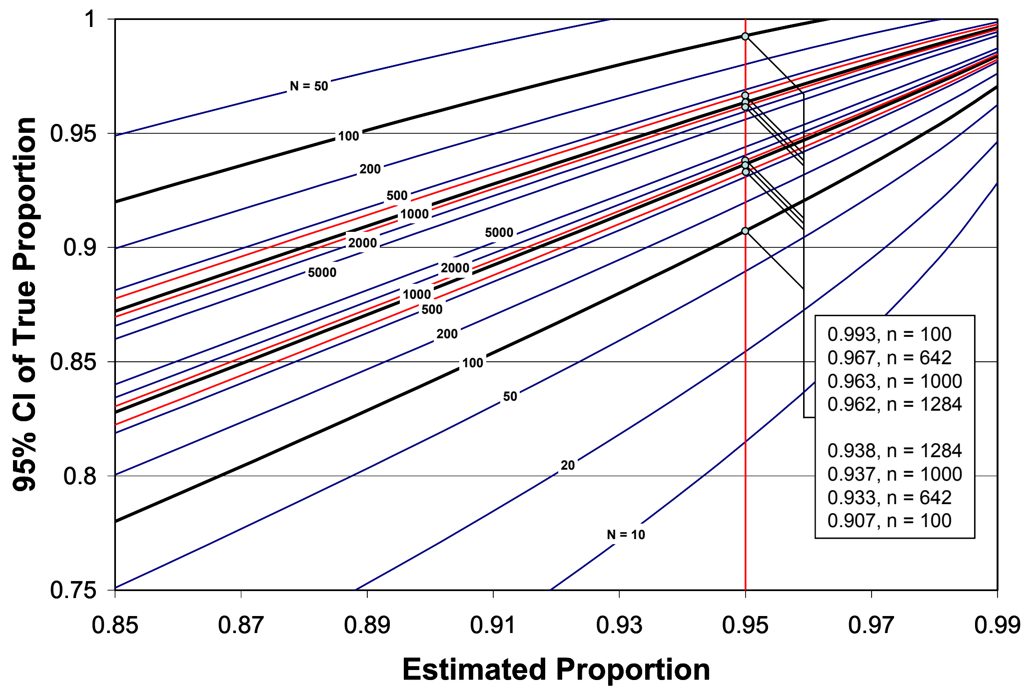


Figure A.2: 95% confidence interval as a function of proportion and sample size.

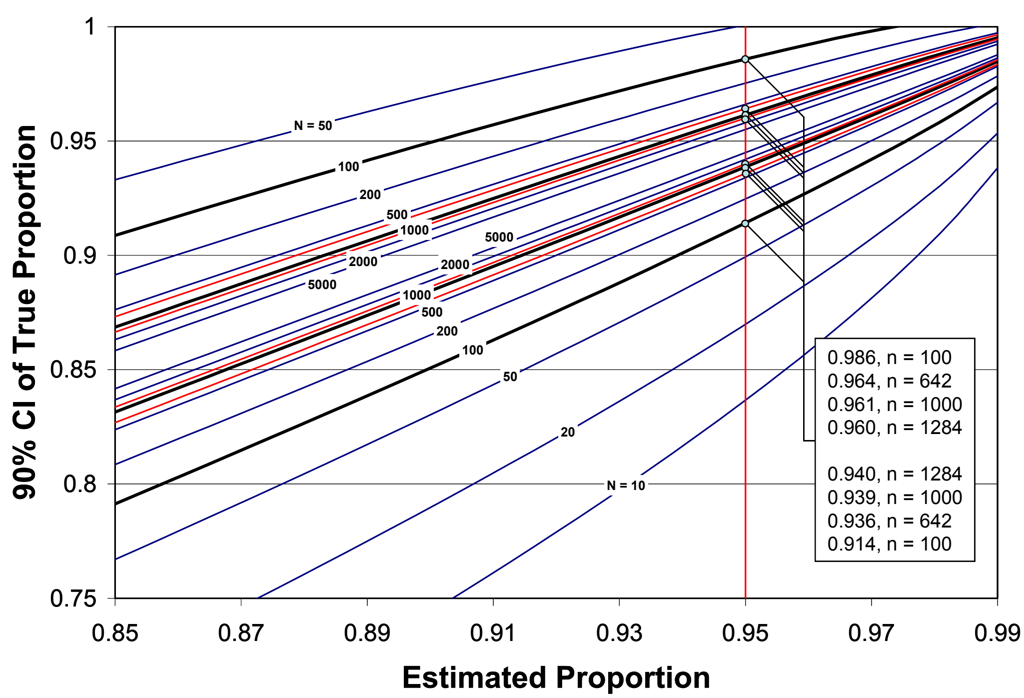


Figure A.3: 90% confidence interval as a function of proportion and sample size.

Table A.2: Lower and Upper Bounds of 95% and 90% Confidence Intervals at Various Proportions and Sample Sizes

Proportion	*n*	95% CI Lower Bound	95% CI Upper Bound	90% CI Lower Bound	90% CI Upper Bound	Proportion	n	95% CI Lower Bound	95% CI Upper Bound	90% CI Lower Bound	90% CI Upper Bound
	3	0.8216	1.1384	0.8470	1.1130		3	0.6413	1.2187	0.6877	1.1723
	5	0.8573	1.1027	0.8770	1.0830		5	0.7064	1.1536	0.7423	1.1177
	10	0.8932	1.0668	0.9072	1.0528		10	0.7719	1.0881	0.7973	1.0627
	20	0.9186	1.0414	0.9285	1.0315		20	0.8182	1.0418	0.8362	1.0238
	50	0.9412	1.0188	0.9474	1.0126		50	0.8593	1.0007	0.8706	0.9894
	100	0.9526	1.0074	0.9570	1.0030		100	0.8800	0.9800	0.8880	0.9720
0.98	200	0.9606	0.9994	0.9637	0.9963	0.93	200	0.8946	0.9654	0.9003	0.9597
	500	0.9677	0.9923	0.9697	0.9903		500	0.9076	0.9524	0.9112	0.9488
	642	0.9692	0.9908	0.9709	0.9891		642	0.9103	0.9497	0.9134	0.9466
	1000	0.9713	0.9887	0.9727	0.9873		1000	0.9142	0.9458	0.9167	0.9433
	1284	0.9723	0.9877	0.9736	0.9864		1284	0.9160	0.9440	0.9183	0.9417
	2000	0.9739	0.9861	0.9749	0.9851		2000	0.9188	0.9412	0.9206	0.9394
	5000	0.9761	0.9839	0.9767	0.9833		5000	0.9229	0.9371	0.9241	0.9359
	3	0.7770	1.1630	0.8080	1.1320		3	0.6130	1.2270	0.6624	1.1766
	5	0.8205	1.1195	0.8445	1.0955		5	0.6822	1.1578	0.7204	1.1196
	10	0.8643	1.0757	0.8813	1.0587		10	0.7519	1.0881	0.7789	1.0611
	20	0.8952	1.0448	0.9073	1.0327		20	0.8011	1.0389	0.8202	1.0198
	50	0.9227	1.0173	0.9303	1.0097		50	0.8448	0.9952	0.8569	0.9831
	100	0.9366	1.0034	0.9419	0.9981		100	0.8668	0.9732	0.8754	0.9646
0.97	200	0.9464	0.9936	0.9502	0.9898	0.92	200	0.8824	0.9576	0.8884	0.9516
	500	0.9550	0.9850	0.9575	0.9825		500	0.8962	0.9438	0.9000	0.9400
	642	0.9568	0.9832	0.9589	0.9811		642	0.8990	0.9410	0.9024	0.9376
	1000	0.9594	0.9806	0.9611	0.9789		1000	0.9032	0.9368	0.9059	0.9341
	1284	0.9607	0.9793	0.9622	0.9778		1284	0.9052	0.9348	0.9075	0.9325
	2000	0.9625	0.9775	0.9637	0.9763		2000	0.9081	0.9319	0.9100	0.9300
	5000	0.9653	0.9747	0.9660	0.9740		5000	0.9125	0.9275	0.9137	0.9263
	3	0.7383	1.1817	0.7739	1.1461		3	0.5862	1.2338	0.6382	1.1818
	5	0.7882	1.1318	0.8159	1.1041		5	0.6592	1.1608	0.6995	1.1205
	10	0.8385	1.0815	0.8581	1.0619		10	0.7326	1.0874	0.7611	1.0589
	20	0.8741	1.0459	0.8879	1.0321		20	0.7846	1.0354	0.8047	1.0153
	50	0.9057	1.0143	0.9144	1.0056		50	0.8307	0.9893	0.8434	0.9766
	100	0.9216	0.9984	0.9278	0.9922		100	0.8539	0.9661	0.8629	0.9571
0.96	200	0.9328	0.9872	0.9372	0.9828	0.91	200	0.8703	0.9497	0.8767	0.9433
	500	0.9428	0.9772	0.9456	0.9744		500	0.8849	0.9351	0.8889	0.9311
	642	0.9448	0.9752	0.9473	0.9727		642	0.8879	0.9321	0.8914	0.9286
	1000	0.9479	0.9721	0.9498	0.9702		1000	0.8923	0.9277	0.8951	0.9249
	1284	0.9493	0.9707	0.9510	0.9690		1284	0.8943	0.9257	0.8969	0.9231
	2000	0.9514	0.9686	0.9528	0.9672		2000	0.8975	0.9225	0.8995	0.9205
	5000	0.9546	0.9654	0.9554	0.9646		5000	0.9021	0.9179	0.9033	0.9167
	3	0.7034	1.1966	0.7430	1.1570		3	0.5605	1.2395	0.6151	1.1849
	5	0.7590	1.1410	0.7897	1.1103		5	0.6370	1.1630	0.6793	1.1207
	10	0.8149	1.0851	0.8366	1.0634		10	0.7141	1.0859	0.7440	1.0560
	20	0.8545	1.0455	0.8698	1.0302		20	0.7685	1.0315	0.7897	1.0103
	50	0.8896	1.0104	0.8993	1.0007		50	0.8168	0.9832	0.8302	0.9698
	100	0.9073	0.9927	0.9142	0.9858		100	0.8412	0.9588	0.8507	0.9493
0.95	200	0.9198	0.9802	0.9247	0.9753	0.90	200	0.8584	0.9416	0.8651	0.9349
	500	0.9309	0.9691	0.9340	0.9660		500	0.8737	0.9263	0.8779	0.9221
	642	0.9331	0.9669	0.9359	0.9641		642	0.8768	0.9232	0.8805	0.9195
	1000	0.9365	0.9635	0.9387	0.9613		1000	0.8814	0.9186	0.8844	0.9156
	1284	0.9381	0.9619	0.9400	0.9600		1284	0.8836	0.9164	0.8862	0.9138
	2000	0.9404	0.9596	0.9420	0.9580		2000	0.8869	0.9131	0.8890	0.9110
	5000	0.9440	0.9560	0.9449	0.9551		5000	0.8917	0.9083	0.8930	0.9070
	3	0.6713	1.2087	0.7145	1.1655		3	0.5359	1.2441	0.5929	1.1871
	5	0.7318	1.1482	0.7653	1.1147		5	0.6157	1.1643	0.6598	1.1202
	10	0.7928	1.0872	0.8165	1.0635		10	0.6961	1.0839	0.7273	1.0527
	20	0.8359	1.0441	0.8527	1.0273		20	0.7529	1.0271	0.7749	1.0051
	50	0.8742	1.0058	0.8848	0.9952		50	0.8033	0.9767	0.8172	0.9628
	100	0.8935	0.9865	0.9009	0.9791		100	0.8287	0.9513	0.8385	0.9415
0.94	200	0.9071	0.9729	0.9124	0.9676	0.89	200	0.8466	0.9334	0.8536	0.9264
	500	0.9192	0.9608	0.9225	0.9575		500	0.8626	0.9174	0.8670	0.9130
	642	0.9216	0.9584	0.9246	0.9554		642	0.8658	0.9142	0.8697	0.9103
	1000	0.9253	0.9547	0.9276	0.9524		1000	0.8706	0.9094	0.8737	0.9063
	1284	0.9270	0.9530	0.9291	0.9509		1284	0.8729	0.9071	0.8756	0.9044
	2000	0.9296	0.9504	0.9313	0.9487		2000	0.8763	0.9037	0.8785	0.9015
	5000	0.9334	0.9466	0.9345	0.9455		5000	0.8813	0.8987	0.8827	0.8973

A.3 Large-Sample Test on a Proportion

A.3.1 Test on Binomial Proportion

Let X be the number of observations in a random sample of size n that belongs to the class associated with proportion p (in this case, the success rate) of a binomial distribution. Then, if the null hypothesis $H_0 : p = p_0$ is true, the distribution is approximately $X \sim N[np_0, np_0(1 - p_0)]$. It should be noted that this approximation procedure will be valid as long as p is not extremely close to zero or one, and if the sample size is relatively larger. To test a two-sided hypothesis $H_0 : p = p_0$ and $H_1 : p \neq p_0$, where p_0 is the true proportion, the test statistic based on the normal approximation to the binomial is then,

$$Z_0 = \frac{np - np_0}{\sqrt{np_0(1 - p_0)}}$$

The null hypothesis $H_0 : p = p_0$ is rejected if $|Z_0| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the percentile of the $N(0,1)$ distribution such that $P\{z \geq Z_{\alpha/2}\} = \alpha/2$.

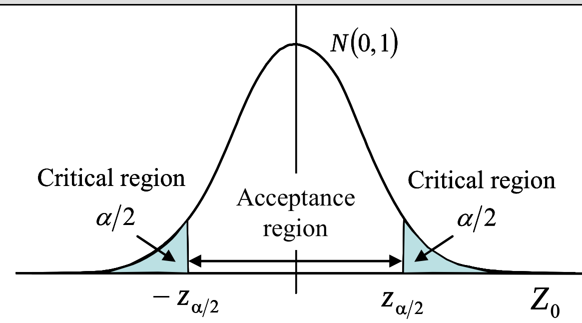
For testing a one-sided hypothesis $H_0 : p = p_0$ and $H_1 : p < p_0$, the H_0 is rejected if the value of Z_0 is too small. Thus, H_0, if $Z_0 < -Z_\alpha$, would be rejected in favor of $H_1 : p < p_0$.

For testing a one-sided hypothesis $H_0 : p = p_0$ and $H_1 : p > p_0$, the H_0 is rejected if the value of Z_0 is too large. Thus H_0, if $Z_0 > Z_\alpha$, would be rejected in favor of $H_1 : p > p_0$.

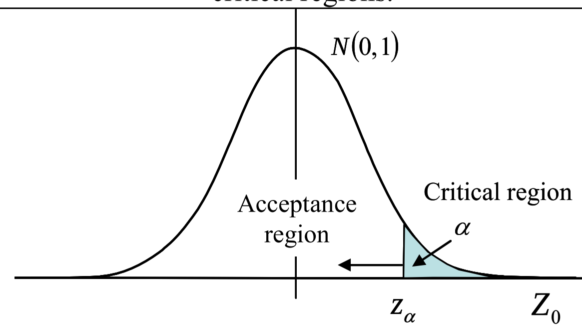
Table A.3 summarizes various testing hypotheses on a binomial proportion and schematically illustrates the rejection criteria for fixed-level tests.

Table A.3: Testing Hypotheses on a Binomial Proportion

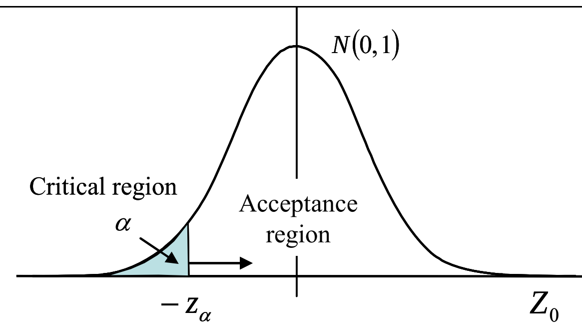
<table>
<tr><th>Testing Hypotheses</th><th>P-Value</th><th>Rejection Criterion
For Fixed-Level Tests</th></tr>
<tr><td>$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$</td><td>Probability above $|Z_0|$ and Probability below $-|Z_0|$

P-value = $2[1-\Phi(Z_0)]$</td><td>


The H_0 is rejected if the value of Z_0 is in the critical regions.</td></tr>
<tr><td>$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$</td><td>Probability above Z_0

P-value = $1 - \Phi(Z_0)$</td><td>


The H_0 is rejected if the value of Z_0 is too large.</td></tr>
<tr><td>$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$</td><td>Probability below Z_0

P-value = $\Phi(Z_0)$</td><td>


The H_0 is rejected if the value of Z_0 is too small.</td></tr>
</table>

A.3.2 Probability of Type II Error β on the Mean

The acceptance or rejection of the null hypothesis H_0 is referred to as a *decision.* Therefore, a correct decision is made in situations in which (1) H_0 is correctly accepted if H_0 is true and (2) H_0 is correctly rejected if H_0 is not true. As shown in Table A.4 for a decision based on a sample, when the null hypothesis is valid, the probability α of erroneously rejecting it is designated as the *Type I* error (or seller's risk); when the null hypothesis is not true, the probability β of erroneously accepting it is named the *Type II* error (or buyer's risk).

Table A.4: Decision-Making in Hypothesis Testing

	Truth about the population	
	H_0 True	H_0 Not True
Reject H_0	**Type I error (α)**	**Correct decision**
Accept H_0	**Correct decision**	**Type II error (β)**

Power is defined as the probability 1 – β of correctly rejecting H_0 if H_0 is not true. Therefore, the definitions of Type I error, Type II error, and Power can be summarized as in the following table:

Type I Error, Type II Error, and Power
Seller's Risk: $\alpha = P\{\text{Type I error}\} = P\{\text{reject } H_0 \mid H_0 \text{ is true}\}$
Buyer's Risk: $\beta = P\{\text{Type II error}\} = P\{\text{fail to reject } H_0 \mid H_0 \text{ is false}\}$
Power $= 1 - \beta = P\{\text{reject } H_0 \mid H_0 \text{ is false}\}$

In general, the contractor and the agency benefit by keeping the Type I error (α) and Type II error (β) low, respectively. From the viewpoint of the agency (the buyer), it is necessary to have the power as high as possible. Conventionally, the Type I error must be kept at or below 0.05 and the statistical power (1 – β) must be kept correspondingly high. To detect a reasonable departure from the null hypothesis, the power should be ideally at least 0.80.

In the following sections, the power calculation and the relationship of power versus sample size versus proportion will be discussed for a two-sided hypothesis and two one-sided hypotheses of proportion.

A.3.2.1 Two-Sided Hypothesis ($H_0: p = p_0$ and $H_1: p \neq p_0$)

At first, consider the two-sided hypothesis, $H_0: \mu = \mu_0$ and $H_1: \mu \neq \mu_0$, then suppose that the null hypothesis is false and that the true value of the mean (or the proportion) is $\mu = \mu_0 + \delta$ where $\delta > 0$, the test statistic is then

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - \mu_0 - \delta + \delta}{\sigma/\sqrt{n}} = \frac{\overline{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}$$

That is, the distribution of Z_0 when H_1 is true follows $Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right)$. Figure A.4 illustrates the distribution of Z_0 under both the null hypothesis and an alternative hypothesis. Based on the definition of

Type II error: $P\{\text{fail to reject } H_0 \mid H_0 \text{ is false}\}$, a Type II error is made only if $-z_{\alpha/2} \le Z_0 \le z_{\alpha/2}$ where $Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma},1\right)$. Hence, we have the probability of Type II error (*1*)

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \qquad \textbf{(A.5)}$$

where Φ is the distribution function of a standard normal distribution. Note that from Figure A.4 the Type II error β is going to increase (that is, the power is to be reduced) as the value of δ decreases.

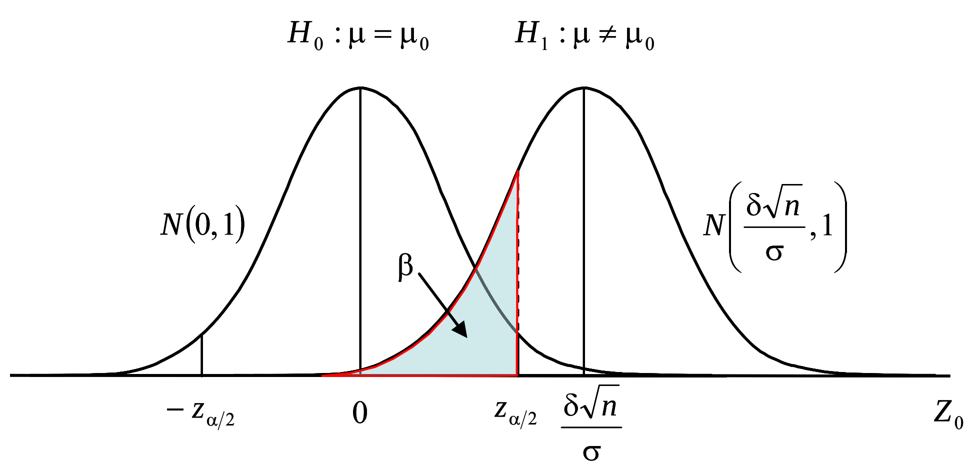


Figure A.4: The distribution of Z_0 under $H_0 : \mu = \mu_0$ and $H_1 : \mu \ne \mu_0$.

Now consider the case of a binomial population proportion. If X is the number of observations in a random sample of size n that belongs to a class of interest, then $\hat{p} = X/n$ is the sample proportion that belongs to that class and the distribution of X is approximately $X \sim N[np_0, np_0(1-p_0)]$; hence, $\hat{p} = X/n$ has the distribution $X/n \sim N[p_0, p_0(1-p_0)]$. Then the test statistic for a binomial proportion is,

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Therefore, the β of Equation A.5 can be converted as follows for the two-sided hypothesis $H_0 : p = p_0$ and $H_1 : p \ne p_0$.

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \qquad \textbf{(A.6)}$$

According to Equation A.6, Figure A.5 through Figure A.7 plot the power (1 - β) as a function of proportion *p*, p_0, and sample size *n* at various α levels (seller's risks). Several findings can be addressed from these figures as follows,

1. At first, the interpretation of Figure A.5 under the hypothesis testing $H_0 : p = p_0$ and $H_1 : p \neq p_0$ is that, for example at *n* = 642 and power = 0.90, the agency will have at least 0.90 power to reject the null hypothesis $H_0 : p = 0.95$ if $p \le 0.916$ (lower bound) or $p \ge 0.978$ (upper bound); however, if $0.916 < p < 0.978$, then the agency will have insufficient power to reject the null hypothesis.
2. The lower and upper bounds at various power and α levels are listed in Table A.5.
3. As expected, at a specified power level, the distance between lower and upper bounds decreases, i.e., the lower and upper bounds will approach *p* = 0.95 as the sample size increases.
4. For a specified power level and sample size, the larger the α level or the higher the seller's risk, the closer the lower and upper bounds. In other words, for a given sample size and estimated proportion, the power of the agency is increased due to the increase of seller's risk.

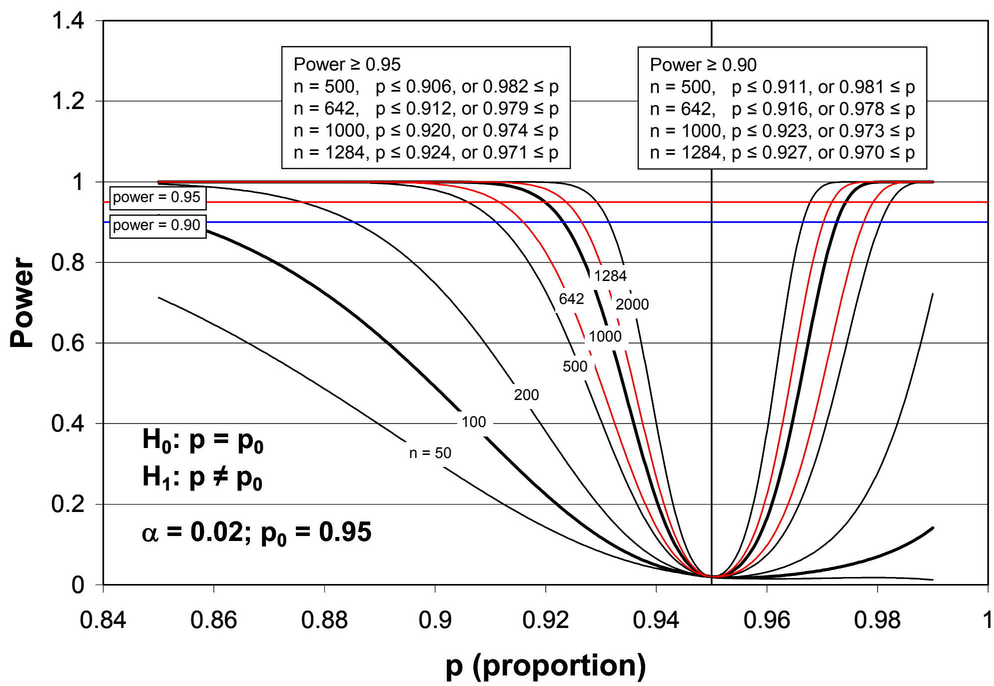


Figure A.5: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ and $H_1 : p \neq p_0$; $p_0 = 0.95$; α = 0.02).

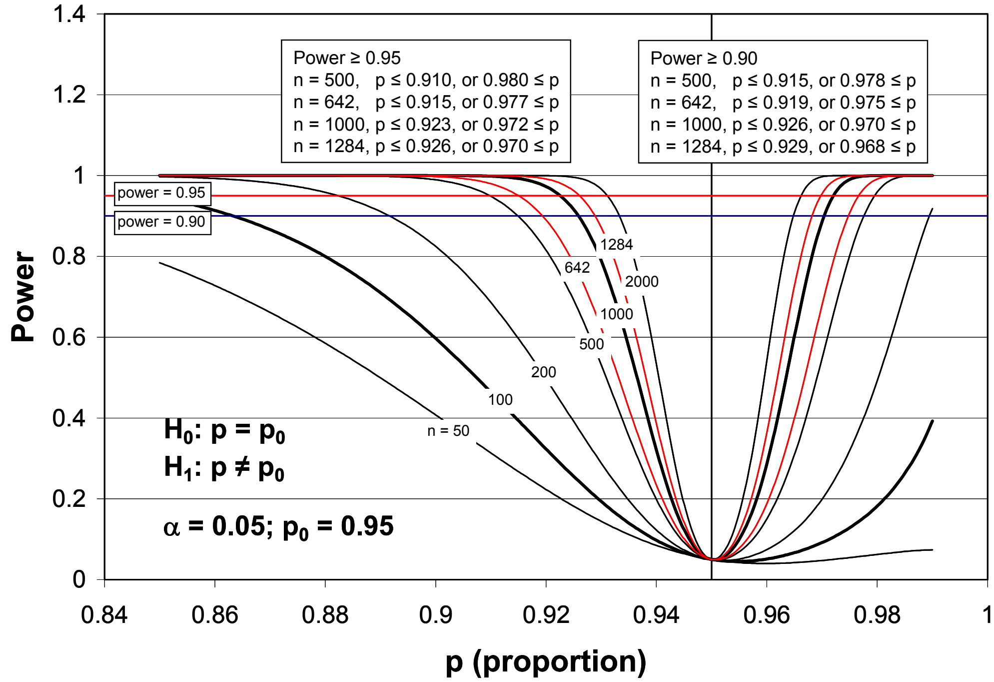


Figure A.6 : The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ **and** $H_1 : p \neq p_0$**;** $p_0 = 0.95$**; α = 0.05).**

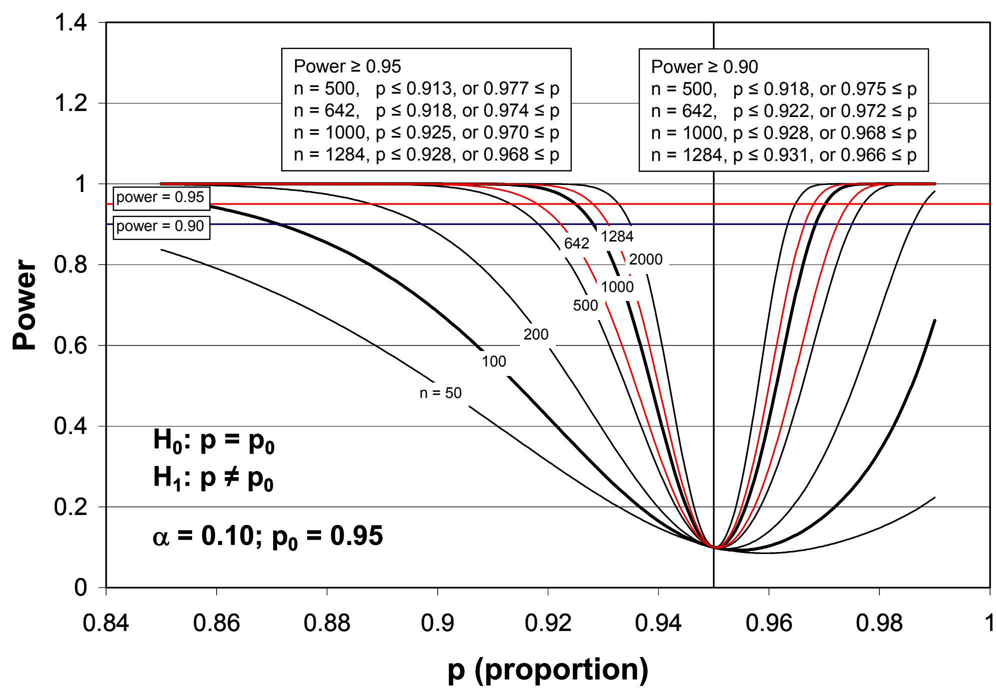


Figure A.7: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ **and** $H_1 : p \neq p_0$**;** $p_0 = 0.95$**; α = 0.10).**

Table A.5: The Lower and Upper Bounds of Two-Sided Hypothesis Testing at Various Powers, α Levels, and Sample Sizes ($H_0 : p = p_0$ and $H_1 : p \neq p_0$; $p_0 = 0.95$)

Power	α = 0.02			α = 0.05			α = 0.10		
	N	Lower Bound	Upper Bound	N	Lower Bound	Upper Bound	N	Lower Bound	Upper Bound
0.95	50			50			50		
	100			100			100	0.857	
	200	0.876		200	0.882		200	0.888	0.988
	500	0.906	0.982	500	0.910	0.980	500	0.913	0.977
	642	0.912	0.979	642	0.915	0.977	642	0.918	0.974
	1000	0.920	0.974	1000	0.923	0.972	1000	0.925	0.970
	1284	0.924	0.972	1284	0.926	0.970	1284	0.928	0.968
	2000	0.929	0.968	2000	0.931	0.966	2000	0.933	0.965
0.90	50			50			50		
	100	0.854		100	0.863		100	0.871	
	200	0.885		200	0.892	0.989	200	0.897	0.986
	500	0.911	0.981	500	0.915	0.978	500	0.918	0.975
	642	0.916	0.978	642	0.919	0.975	642	0.922	0.972
	1000	0.923	0.973	1000	0.926	0.970	1000	0.928	0.968
	1284	0.927	0.970	1284	0.929	0.968	1284	0.931	0.966
	2000	0.931	0.966	2000	0.933	0.965	2000	0.935	0.963
0.80	50			50			50	0.858	
	100	0.871		100	0.880		100	0.888	
	200	0.895		200	0.902	0.987	200	0.907	0.983
	500	0.917	0.978	500	0.921	0.975	500	0.924	0.972
	642	0.921	0.975	642	0.924	0.972	642	0.927	0.970
	1000	0.927	0.971	1000	0.930	0.968	1000	0.932	0.966
	1284	0.930	0.968	1284	0.932	0.966	1284	0.934	0.964
	2000	0.934	0.965	2000	0.936	0.963	2000	0.937	0.962
0.50	50	0.878		50	0.890		50	0.900	
	100	0.899		100	0.907		100	0.915	0.986
	200	0.914	0.986	200	0.920	0.980	200	0.925	0.975
	500	0.927	0.973	500	0.931	0.969	500	0.934	0.966
	642	0.930	0.970	642	0.933	0.967	642	0.936	0.964
	1000	0.934	0.966	1000	0.936	0.964	1000	0.939	0.961
	1284	0.936	0.964	1284	0.938	0.962	1284	0.940	0.960
	2000	0.937	0.961	2000	0.940	0.960	2000	0.942	0.958
0.30	50	0.901		50	0.911		50	0.921	
	100	0.914		100	0.922	0.987	100	0.929	0.978
	200	0.924	0.981	200	0.929	0.974	200	0.934	0.969
	500	0.933	0.969	500	0.937	0.965	500	0.940	0.962
	642	0.935	0.966	642	0.938	0.963	642	0.941	0.960
	1000	0.938	0.963	1000	0.940	0.960	1000	0.943	0.958
	1284	0.939	0.961	1284	0.942	0.959	1284	0.943	0.957
	2000	0.941	0.959	2000	0.943	0.957	2000	0.945	0.956

A.3.2.2 One-Sided Hypothesis ($H_0 : p = p_0$ and $H_1 : p > p_0$)

The same argument that is applied in Section A.3.2.1 regarding power calculation can be applied to the one-sided hypothesis $H_0 : p = p_0$ and $H_1 : p > p_0$. Based on the definition of Type II error, $P\{$fail to reject H_0 | H_0 is false$\}$, a Type II error is made only if $Z_0 \le z_\alpha$ where $Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma},1\right)$, i.e., the shaded area of Figure A.8.

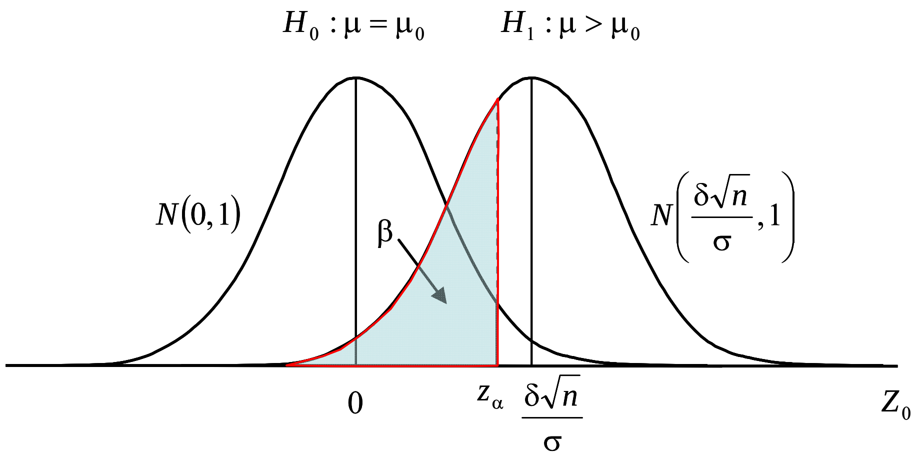


Figure A.8: The distribution of Z_0 under $H_0 : \mu = \mu_0$ and $H_1 : \mu > \mu_0$.

Thus,

$$\beta = \Phi\left(z_\alpha - \frac{\delta\sqrt{n}}{\sigma}\right) = \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \quad \textbf{(A.7)}$$

According to Equation A.7, Figure A.9 through Figure A.11 plot the power (1 - β) as a function of proportion *p*, p_0, and sample size *n* at various α levels (seller's risks). Several findings can be concluded from these figures as follows:

1. The interpretation of Figure A.9 under the hypothesis testing $H_0 : p = p_0$ and $H_1 : p > p_0$ at $n = 642$ and power = 0.90 is that the agency will have at least 0.90 power to reject the null hypothesis $H_0 : p = p_0$ and favor the alternative hypothesis $H_1 : p > p_0$ if $p \ge 0.975$ (lower bound); otherwise, if $p \le 0.975$, then the agency will have insufficient power to reject the null hypothesis.
2. The corresponding lower bounds at various power and α levels are listed in Table A.6.
3. As expected, the lower bound will approach $p = 0.95$ as the sample size increases.
4. For a specified power level and sample size, the larger the α level or the higher the seller's risk, the closer the lower bounds to $p = 0.95$. That is to say, for a given sample size and estimated proportion, the power of the agency is increased due to the increase of seller's risk.

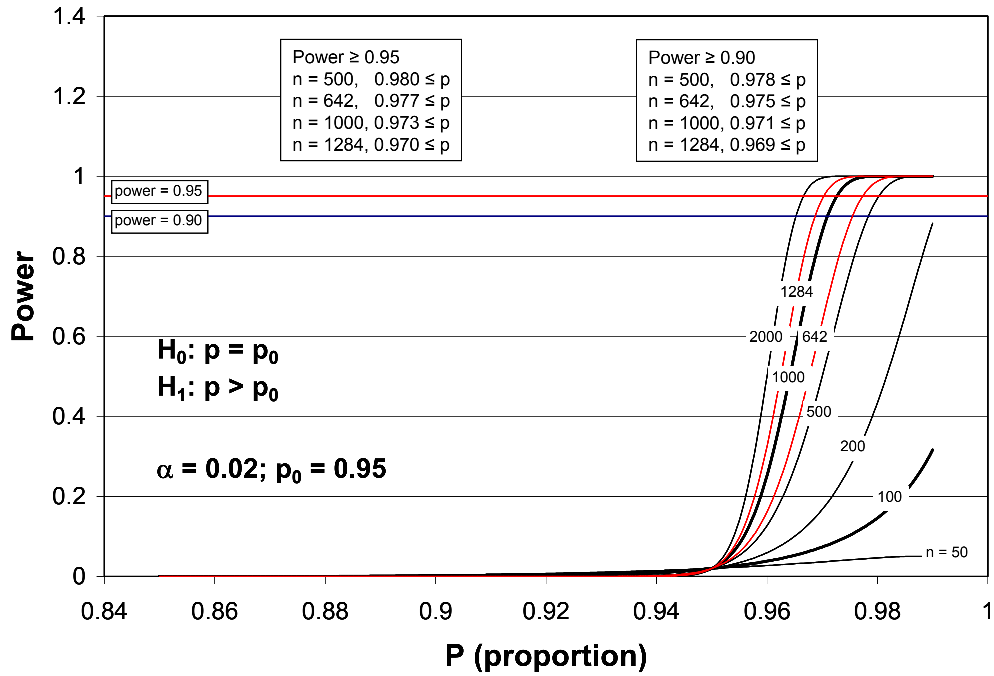


Figure A.9: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ and $H_1 : p > p_0$; $p_0 = 0.95$; α = 0.02).

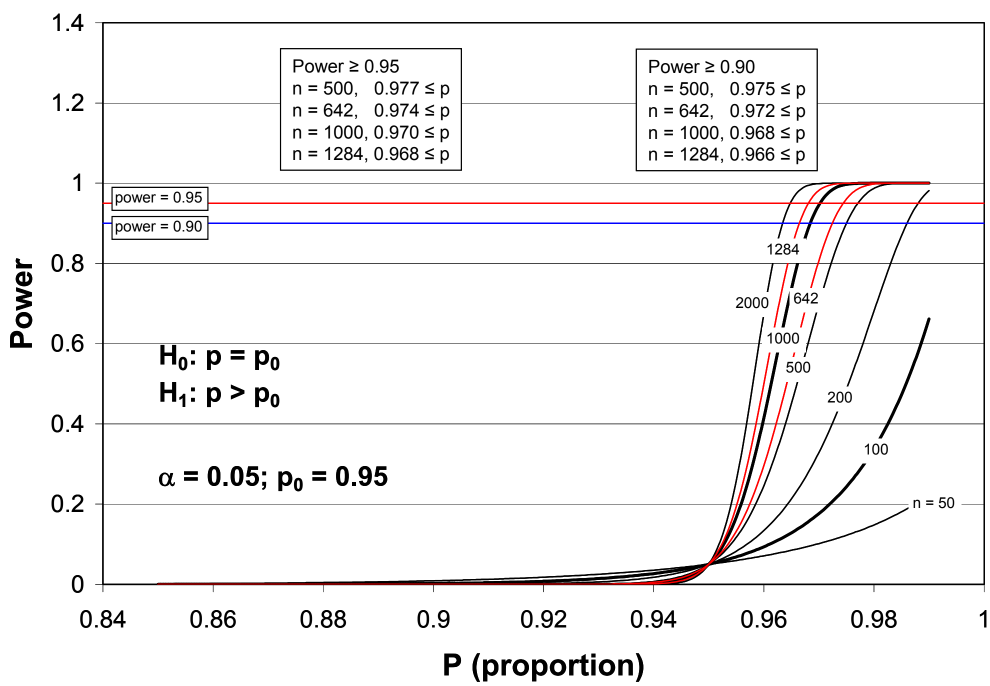


Figure A.10: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ and $H_1 : p > p_0$; $p_0 = 0.95$; α = 0.05).

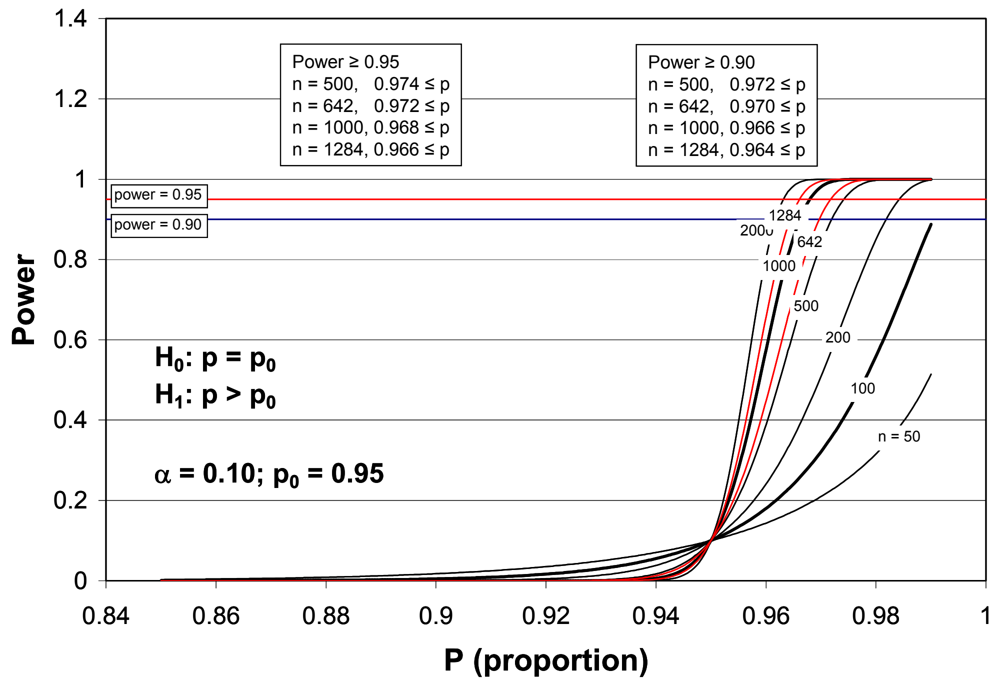


Figure A.11: The relationship of power versus proportion with various sample sizes ($H_0: p = p_0$ and $H_1: p > p_0$; $p_0 = 0.95$; α = 0.10).

A.3.2.3 One-Sided Hypothesis ($H_0: p = p_0$ and $H_1: p < p_0$)

The same argument that was applied is Section A.3.2.1 regarding power calculation can be also applied to the one-sided hypothesis $H_0: p = p_0$ and $H_1: p < p_0$. Based on the definition of Type II error, P{fail to reject $H_0 \mid H_0$ is false}, the Type II error is made only if $-z_\alpha \le Z_0$ where $Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right)$, i.e., the shaded area of Figure A.12.

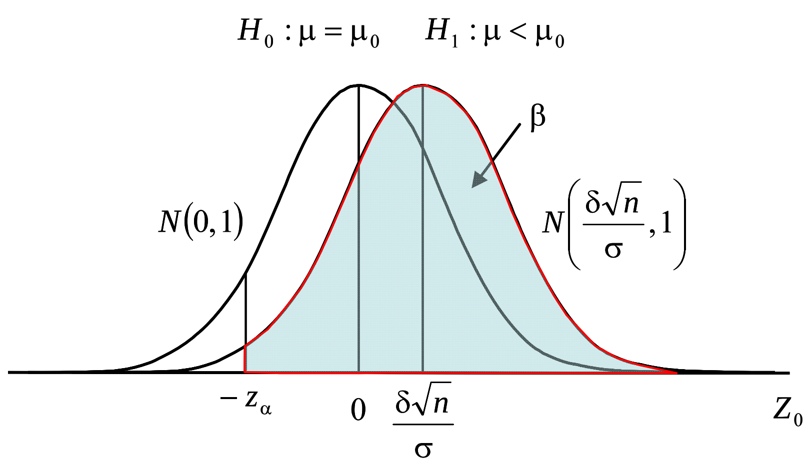


Figure A.12 The distribution of Z_0 under $H_0: \mu = \mu_0$ and $H_1: \mu < \mu_0$.

Table A.6: The Lower Bounds of a One-Sided Hypothesis Test at Various Powers, α Levels, and Sample Sizes ($H_0 : p = p_0$ and $H_1 : p > p_0$; $p_0 = 0.95$)

Power	α = 0.02			α = 0.05			α = 0.10		
	N	Lower Bound	Upper Bound	N	Lower Bound	Upper Bound	N	Lower Bound	Upper Bound
0.95	50			50			50		
	100			100			100		
	200			200	0.988		200	0.984	
	500	0.980		500	0.977		500	0.974	
	642	0.977		642	0.974		642	0.972	
	1000	0.973		1000	0.970		1000	0.968	
	1284	0.970		1284	0.968		1284	0.966	
	2000	0.967		2000	0.965		2000	0.963	
0.90	50			50			50		
	100			100			100		
	200			200	0.986		200	0.982	
	500	0.978		500	0.975		500	0.972	
	642	0.975		642	0.972		642	0.970	
	1000	0.971		1000	0.968		1000	0.966	
	1284	0.969		1284	0.966		1284	0.964	
	2000	0.965		2000	0.963		2000	0.962	
0.80	50			50			50		
	100			100			100	0.987	
	200	0.988		200	0.983		200	0.978	
	500	0.976		500	0.972		500	0.969	
	642	0.973		642	0.970		642	0.967	
	1000	0.969		1000	0.966		1000	0.964	
	1284	0.967		1284	0.964		1284	0.962	
	2000	0.964		2000	0.962		2000	0.960	
0.50	50			50			50	0.989	
	100			100	0.986		100	0.978	
	200	0.982		200	0.975		200	0.970	
	500	0.970		500	0.966		500	0.962	
	642	0.968		642	0.964		642	0.961	
	1000	0.964		1000	0.961		1000	0.959	
	1284	0.962		1284	0.960		1284	0.958	
	2000	0.960		2000	0.958		2000	0.956	
0.30	50			50			50	0.979	
	100	0.989		100	0.978		100	0.969	
	200	0.976		200	0.969		200	0.963	
	500	0.966		500	0.962		500	0.958	
	642	0.964		642	0.960		642	0.957	
	1000	0.961		1000	0.958		1000	0.955	
	1284	0.960		1284	0.957		1284	0.955	
	2000	0.958		2000	0.956		2000	0.954	

Thus,

$$\beta = 1 - \Phi\left(-z_\alpha - \frac{\delta\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(\frac{p_0 - p - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \quad \textbf{(A.8)}$$

According to Equation A.8, Figure A.13 through Figure A.15 plot the power (1 - β) as a function of proportion p, p_0, and sample size n at various α levels (seller's risks). Several findings can be concluded from these figures, as follows:

1. The interpretation of Figure A.13 under the hypothesis testing $H_0 : p = p_0$ and $H_1 : p < p_0$ at $n = 642$ and power = 0.90 is that the agency will have at least 0.90 power to reject the null hypothesis $H_0 : p = p_0$ and favor the alternative hypothesis $H_1 : p < p_0$ if $p \leq 0.914$ (upper bound); otherwise, if $p \geq 0.914$, then the agency will not have enough power to reject the null hypothesis.
2. The corresponding upper bounds at various power and α levels are listed in Table A.7.
3. As expected, the upper bound will approach $p = 0.95$ as the sample size increases.
4. For a specified power level and sample size, the larger the α level or the higher the seller's risk, the larger the upper bounds. That is to say, for a given sample size and estimated proportion, the power of the agency is increased due to the increase of seller's risk.

It is recognized that a minimum population proportion of 0.95 is required by the agency in this bridge project. It is then in the agency's best interest to test the hypotheses $H_0 : p = p_0$ and $H_1 : p < p_0$ rather than $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$ or $H_0 : p = p_0$ and $H_1 : p > p_0$. Therefore, the hypothesis testing of $H_0 : p = p_0$ and $H_1 : p < p_0$ will be used for the development of the acceptance criteria for a quality assurance process.

For a given α level and a specified β risk (or power level), Equations A.6, A.7, and A.8 can be solved to find the approximate sample size. Hence, the approximate sample size equation for a two-sided test on a binomial proportion is

$$n = \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0}\right)^2.$$

The approximate sample size equation for a one-sided test on a binomial proportion is

$$n = \left(\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0}\right)^2$$

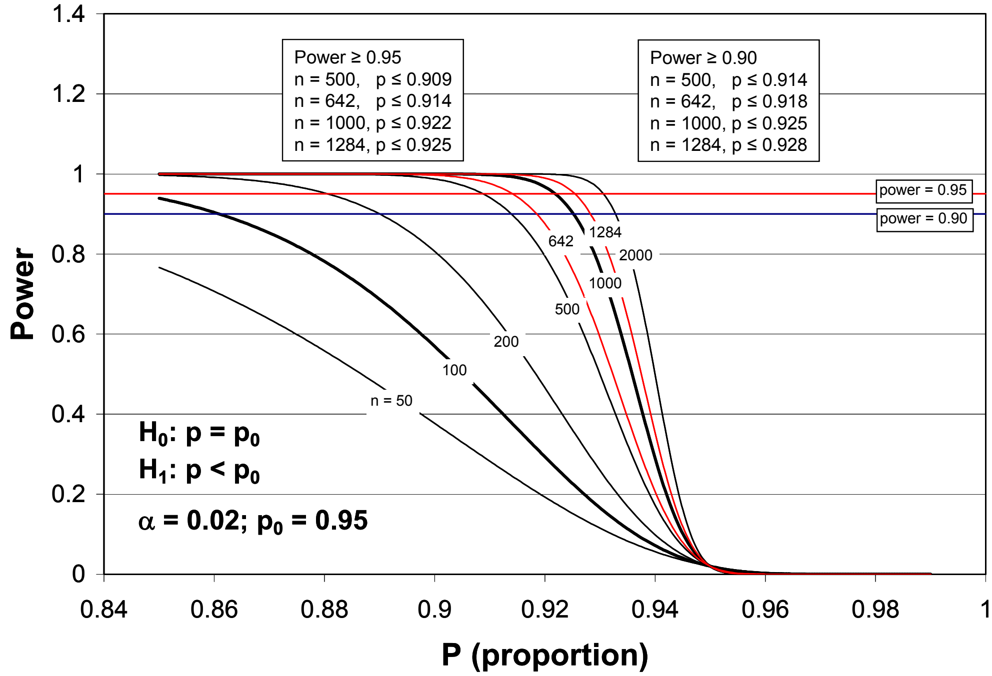


Figure A.13: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ and $H_1 : p < p_0$; $p_0 = 0.95$; α = 0.02).

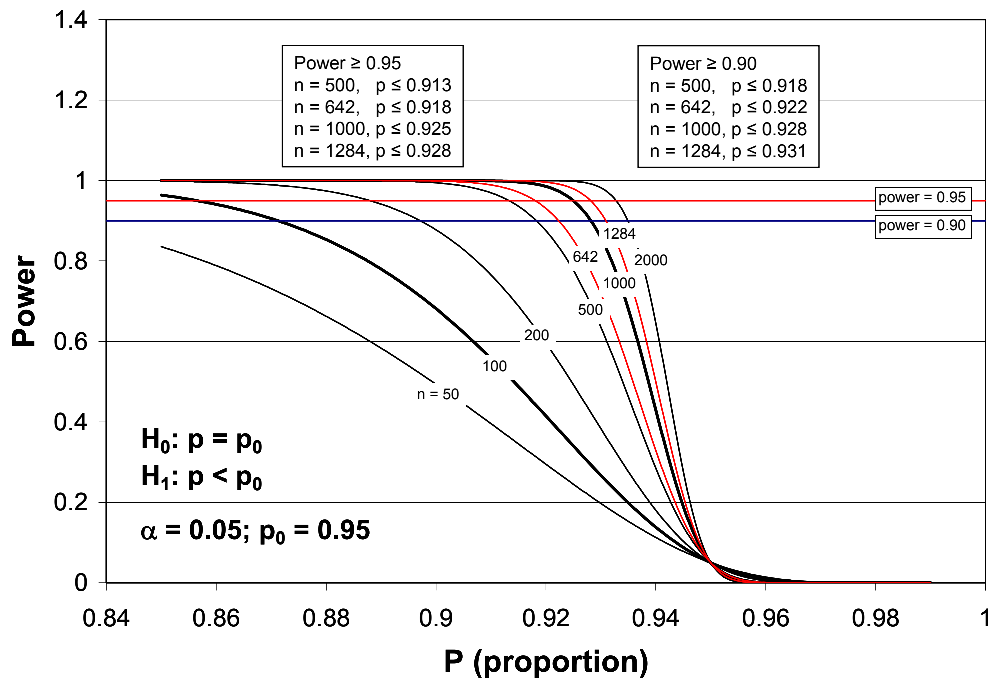


Figure A.14: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ and $H_1 : p < p_0$; $p_0 = 0.95$; α = 0.05).

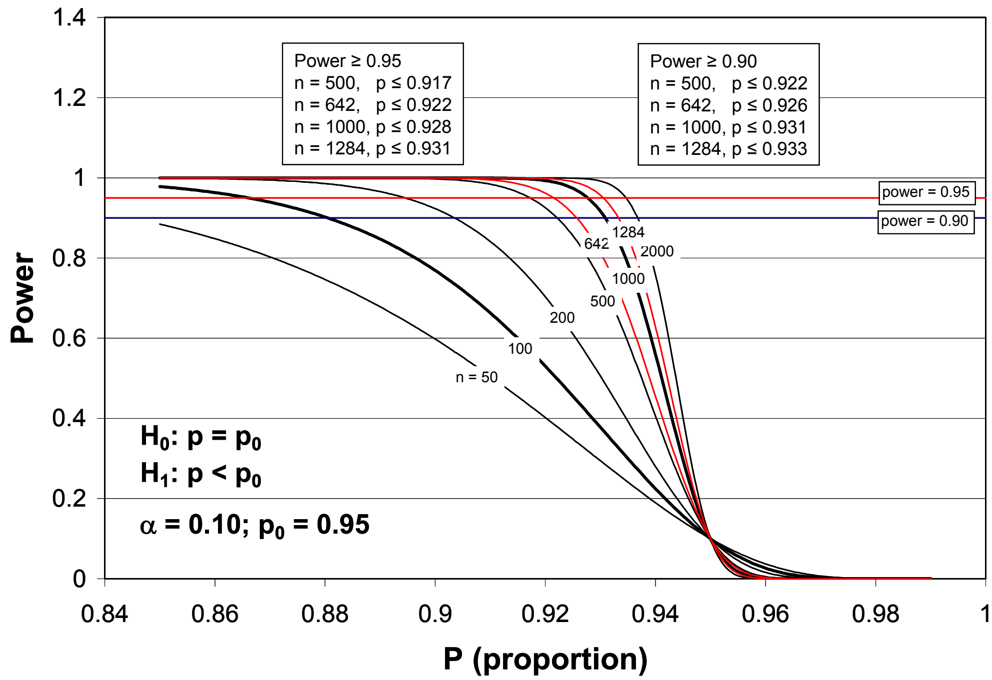


Figure A.15: The relationship of power versus proportion with various sample sizes ($H_0 : p = p_0$ and $H_1 : p < p_0$; $p_0 = 0.95$; α = 0.10).

Table A.7: Upper Bounds of a One-Sided Hypothesis Test at Various Powers, α Levels, and Sample Sizes ($H_0 : p = p_0$ and $H_1 : p < p_0$; $p_0 = 0.95$)

Power	α = 0.02			α = 0.05			α = 0.10		
	N	Lower Bound	Upper Bound	N	Lower Bound	Upper Bound	N	Lower Bound	Upper Bound
0.95	50			50			50		
	100			100		0.856	100		0.866
	200		0.881	200		0.888	200		0.895
	500		0.909	500		0.913	500		0.917
	642		0.914	642		0.918	642		0.922
	1000		0.922	1000		0.925	1000		0.928
	1284		0.925	1284		0.928	1284		0.931
	2000		0.931	2000		0.933	2000		0.935
0.90	50			50			50		
	100		0.861	100		0.871	100		0.880
	200		0.890	200		0.897	200		0.903
	500		0.914	500		0.918	500		0.922
	642		0.918	642		0.922	642		0.926
	1000		0.925	1000		0.928	1000		0.931
	1284		0.928	1284		0.931	1284		0.933
	2000		0.933	2000		0.935	2000		0.937
0.80	50			50		0.858	50		0.871
	100		0.878	100		0.888	100		0.896
	200		0.901	200		0.907	200		0.914
	500		0.920	500		0.924	500		0.928
	642		0.924	642		0.927	642		0.931
	1000		0.929	1000		0.932	1000		0.935
	1284		0.932	1284		0.934	1284		0.936
	2000		0.935	2000		0.937	2000		0.939
0.50	50		0.887	50		0.899	50		0.910
	100		0.905	100		0.914	100		0.922
	200		0.918	200		0.925	200		0.930
	500		0.930	500		0.934	500		0.938
	642		0.932	642		0.936	642		0.939
	1000		0.934	1000		0.939	1000		0.941
	1284		0.938	1284		0.940	1284		0.942
	2000		0.940	2000		0.942	2000		0.944
0.30	50		0.908	50		0.919	50		0.929
	100		0.920	100		0.928	100		0.935
	200		0.928	200		0.934	200		0.939
	500		0.936	500		0.940	500		0.943
	642		0.937	642		0.941	642		0.944
	1000		0.940	1000		0.943	1000		0.945
	1284		0.941	1284		0.943	1284		0.946
	2000		0.943	2000		0.945	2000		0.946

APPENDIX B: SAMPLE SIZE DETERMINATION

B.1 Statistical Sampling Simulation

The purpose of this sampling simulation is to investigate the effects of the quality of contractor (proportion), samples per location, number of locations, and total number in the sample. In Table B.1, there are 642 locations containing a total of 4,254 rebar dowels for all the bridges. For this sampling simulation, the binomial population was randomly generated based on the quality of contractor (proportion) and the assumption of 8 rebar dowels for each of the 642 locations, i.e., a binomial population with 5,136 rebar dowels. The factors and their corresponding factor levels in the experimental design of the sampling simulation are as follows:

Factor **Contractor**, i.e., quality of contractor, with four proportion levels: 0.85, 0.90, 0.95, 0.98;

Factor **SamplesPerLocation** with four levels: 1, 2, 3, 4; and

Factor **Locations** with seven levels: 10, 20, 50, 100, 200, 500, 642.

Each of the 112 cases (4 × 4 × 7) was simulated 500 times. The proportion was calculated for each simulation, hence the proportion distribution was generated after 500 simulations. The standard deviation *S* was used to characterize the proportion distribution dispersion, and the design plot in Figure B.1 shows the main effects of the factors on the standard deviation. In the figure, the horizontal line represents the grand mean of the response variable (i.e., the standard deviation *S*) and the vertical line with short sticks indicates the means of factor levels for a specific factor. Therefore, the farther apart the marked factor levels on the vertical line are, the more significant the effect of the factor on the response variable. It should be noted that the factor **TotalSamples** is the product of two factors, **Locations** and **SamplePerLocation**.

In addition to the design plot, the factor plots shown in Figure B.2 display the effects of factor levels to the response variable in terms of box plots. The box plot illustrates a measure of location (the median [white strip]), a measure of dispersion (the interquartile range IQR [lower quartile: bottom-edge of box; upper quartile: top-edge of box]), and the possible outliers (data points with a horizontal line outside the 1.5IQR distance from the edges of box; the most extreme data points within 1.5 IQR distance are marked with a bracket), and also gives an indication of the symmetry or skewness of the distribution.

Several findings from Figure B.1 and Figure B.2 can be addressed in the following:

1. The higher the values of quality of contractor (proportion), samples per location, number of locations, and sample size, the lower the standard deviation *S*.
2. It seems that sample size has the most significant effect on standard deviation *S*.
3. From Figure B.2(d), it is apparent that there is a critical point of diminishing returns (probably around 100 ~ 200) where increasing sample size provides little benefit. It is necessary to have the sample size "large enough" so that sampling error will tend to be on a reasonable level of accuracy. Otherwise, if the

sample size is too small, there is no point in gathering the data because the results will tend to be too imprecise to be of use.

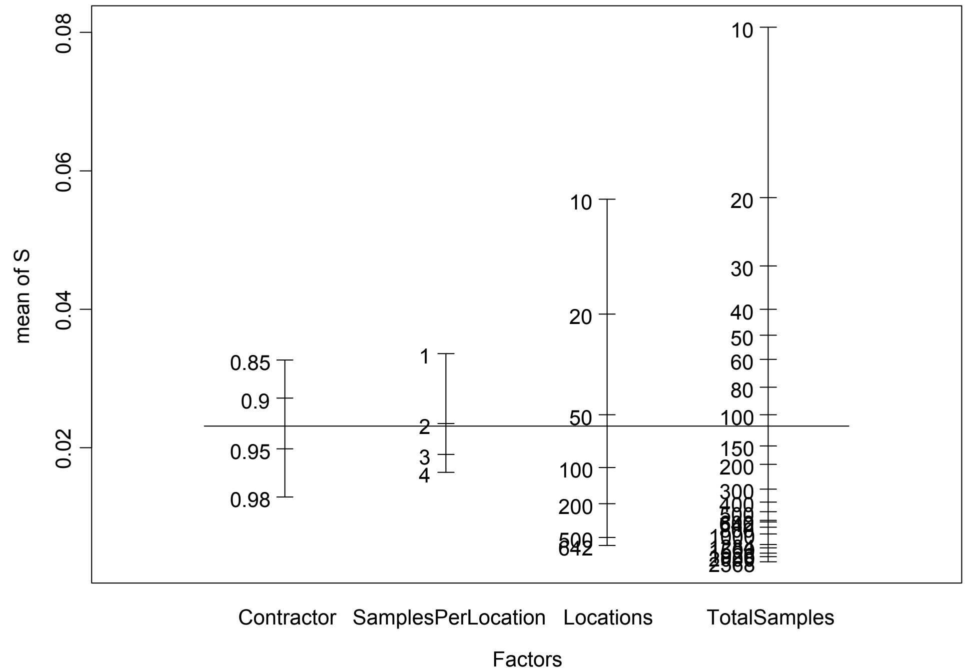


Figure B.1: Design plot of the main effects of the sampling simulation results.

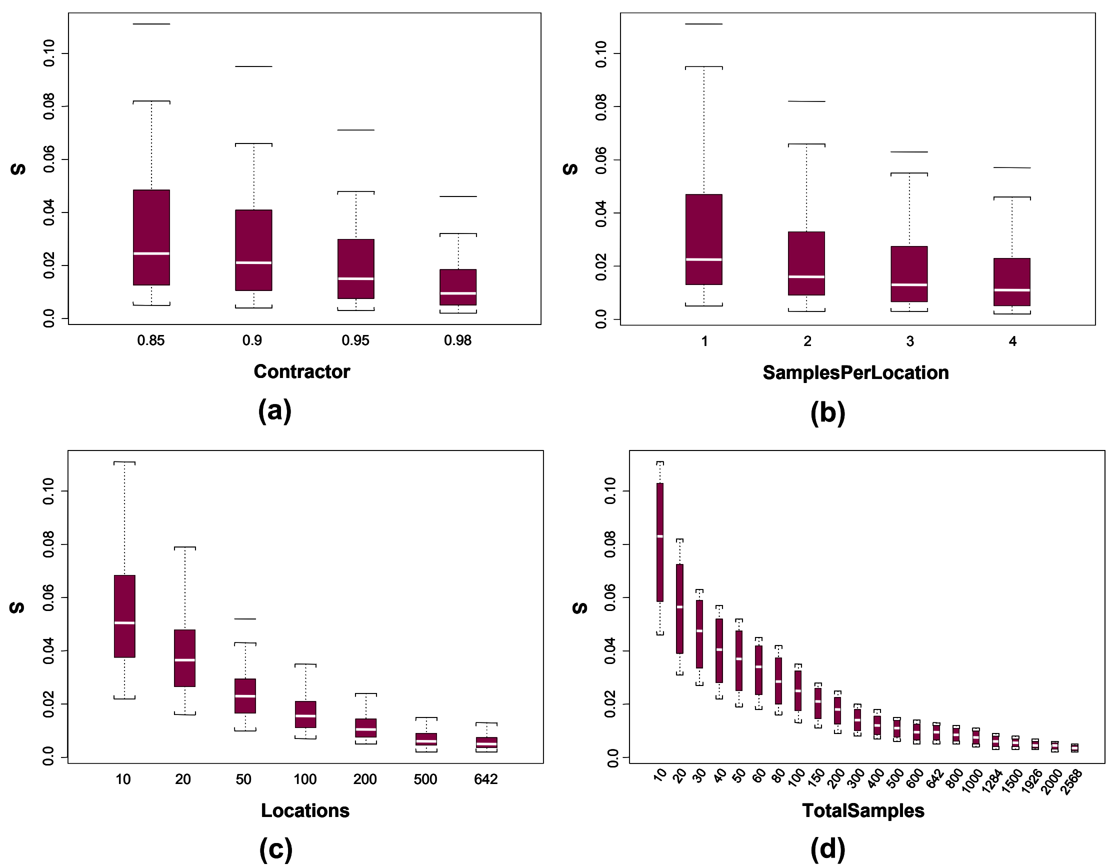


Figure B.2: Factor plots of sampling simulation results.

B.2 Why is it inappropriate to take only three samples?

The statistical simulation indicates that the sample size drawn from a binomial population has to be large enough to produce a reasonable level of accuracy, and that making the sample larger simply wastes time and money. It is not uncommon for agencies to base quality assurance on three samples. The discussion of binomial distribution presented here shows why it is inappropriate to only take this number of samples for quality assurance. The performance index obtained from these three samples could be calculated by taking their average or could be counted by the number of successes/failures, as in the binomial distribution presented in Figure B.3 and Figure B.4. The performance index would then be compared to the performance specification to statistically accept or reject the project through hypothesis testing. The question raised then is, how confident will the agency be by relying on such tiny fraction of samples?

To answer the question, two major factors considered in the following binomial example are the statistical power of hypothesis testing and the performance specification. Recall that *power* is defined as the probability of correctly rejecting the null hypothesis given that the null hypothesis is wrong, i.e., $P\{\text{reject } H_0 \mid H_0 \text{ is false}\}$. Recall too that the two parameters for determining the binomial distribution are sample size *n* and population proportion *p*. The event $Y \geq 2$ with three samples will be inspected under the one-sided hypothesis $\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$, where p_0 is 0.95. Figure B.3 and Figure B.4, respectively, plot the binomial probability distributions in terms of *X* (count of success) and *Y* (count of failures) with $n = 3$ and various population proportions.

Table B.1 lists not only the probabilities of counts of failure/success but also the cumulative probabilities of the event $Y \geq 2$ in a binomial distribution with $n = 3$ and various proportions.

Several findings can be addressed from the binomial distributions with parameters $n = 3$ and various proportions as presented in Figure B.3 through Figure B.4, and Table B.1.

1. The probability of being correct on all three trials $P(X = 3)$ or $P(Y = 0)$ decreases as the proportion decreases. Note that even when $p = 0.80$ (which is far from H_0: $p = 0.95$) the probability of being correct on all three trials is still 0.51.
2. As Figure B.3 and Figure B.4 show, the binomial distributions with parameters $n = 3$ cannot be approximated by a normal distribution.
3. The paired binomial distributions of [$P(X = 1)$, $P(X = 2)$] and [$P(X = 3)$, $P(X = 0)$] are symmetrical at $p = 0.5$ (Figure B.3).
4. As noted in Figure B.5, the binomial probability distributions of $P(X = 3)$ and $P(X = 0)$ are monotonically increasing and decreasing as *p* increases; however, there are peak values or modes for $P(X = 2)$ or $P(X = 1)$ that occurred roughly at $p = 0.68$ and $p = 0.32$ separately.

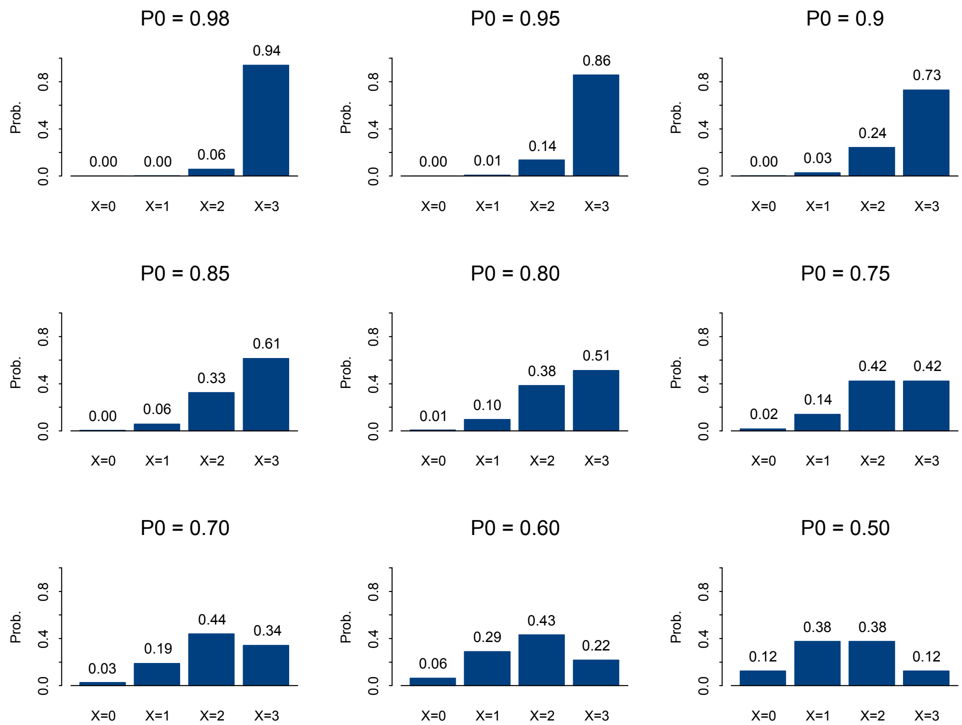


Figure B.3: Binomial distributions with various population proportions and $n = 3$. (X stands for number of successes).

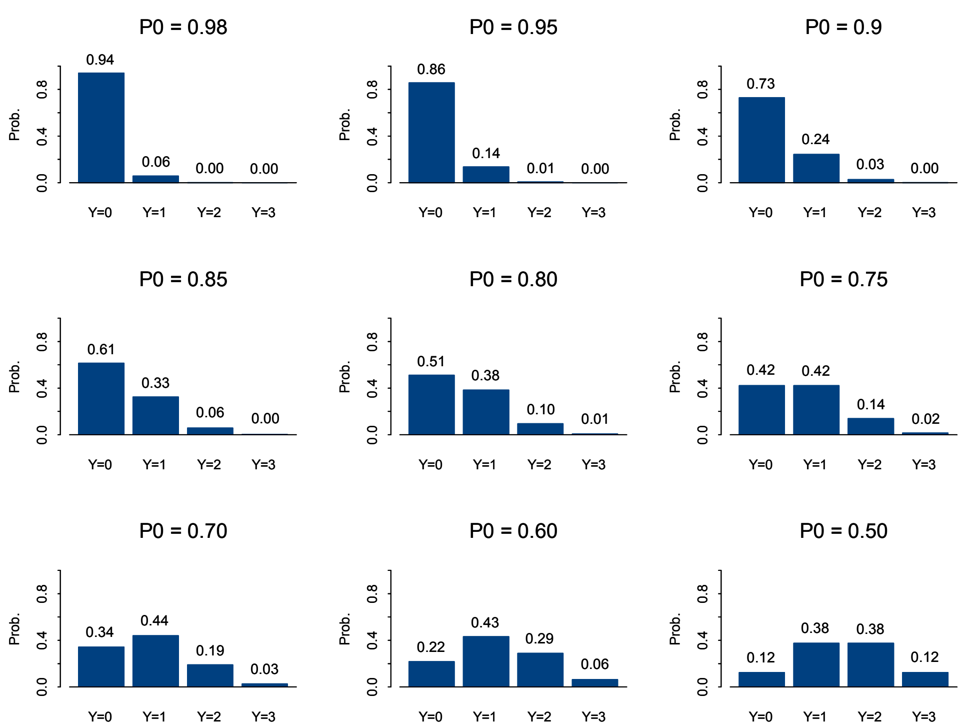


Figure B.4: Binomial distributions with various population proportions and $n = 3$. (Y stands for number of failures).

Table B.1: Probabilities of a Binomial Distribution with *n* = 3 and Various Proportions

P0	1 - P0	Probability of Count of Failure/Success				Cumulative Probability
		P(Y = 0) or P(X = 3)	P(Y = 1) or P(X = 2)	P(Y = 2) or P(X = 1)	P(Y = 3) or P (X = 0)	P(Y ≥ 2) Or P(X ≤ 1)
0.98	0.02	0.9412	0.0576	0.0012	0.0000	0.0012
0.95	0.05	0.8574	0.1354	0.0071	0.0001	0.0072
0.90	0.10	0.7290	0.2430	0.0270	0.0010	0.0280
0.85	0.15	0.6141	0.3251	0.0574	0.0034	0.0574
0.80	0.20	0.5120	0.3840	0.0960	0.0080	0.1040
0.75	0.25	0.4219	0.4219	0.1406	0.0156	0.1562
0.70	0.30	0.3430	0.4410	0.1890	0.0270	0.2160
0.65	0.35	0.2746	0.4436	0.2389	0.0429	0.2818
0.60	0.40	0.2160	0.4320	0.2880	0.0640	0.3520
0.55	0.45	0.1664	0.4084	0.3341	0.0911	0.4252
0.50	0.50	0.1250	0.3750	0.3750	0.1250	0.5000
0.45	0.55	0.0911	0.3341	0.4084	0.1664	0.5748
0.40	0.60	0.0640	0.2880	0.4320	0.2160	0.6480
0.35	0.65	0.0429	0.2389	0.4436	0.2746	0.7182
0.30	0.70	0.0270	0.1890	0.4410	0.3430	0.7840
0.25	0.75	0.0156	0.1406	0.4219	0.4219	0.8438
0.20	0.80	0.0080	0.0960	0.3840	0.5120	0.8960
0.15	0.85	0.0034	0.0574	0.3251	0.6141	0.9392
0.10	0.90	0.0010	0.0270	0.2430	0.7290	0.9720
0.05	0.95	0.0001	0.0071	0.1354	0.8574	0.9928

A one-tailed hypothesis test $\begin{cases} H_0 : p = 0.95 \\ H_1 : p < 0.95 \end{cases}$ was conducted at a 5% significance level. If the *p*-value is less than 5%, then the null hypothesis will be rejected in favor of alternative hypothesis H_1; otherwise, the null hypothesis will not be rejected due to the lack of strong evidence.

According to the binomial distribution with parameters n =3 and p = 0.95 (as plotted in Figure B.4), the probability of failing one or more trials, P($Y \geq 1$) = P(Y = 1) + P(Y = 2) + P(Y = 3) or P($X \leq 2$), is 0.1426, which is larger than the 5% significance level. Thus, in order to conduct the hypothesis testing $\begin{cases} H_0 : p = 0.95 \\ H_1 : p < 0.95 \end{cases}$ at 5% significance level, the event $Y \geq 2$ was used to establish the critical region. The probability of failing on 2 or more counts P($Y \geq 2$) (Figure B.4)—which is equivalent to the statement that the probability of being correct on 1 or fewer trials (P($X \leq 1$) (Figure B.3), given that the null hypothesis, H_0: p = 0.95—is true is less than 0.01.

Recall that *power* is defined as the probability of correctly rejecting H_0: p = 0.95 given that H_0 is false, i.e., power = $P\{\text{reject } H_0 \mid H_0 \text{ is false}\}$. As an example, considering the binomial distribution with n = 3 and p = 0.85, it is apparent that H_0 is now false; hence, the power = P($Y \geq 2 \mid p = 0.85$) = 0.0574 (Table B.1), i.e., the probability of $Y \geq 2$ given that p = 0.85 is 0.0574.

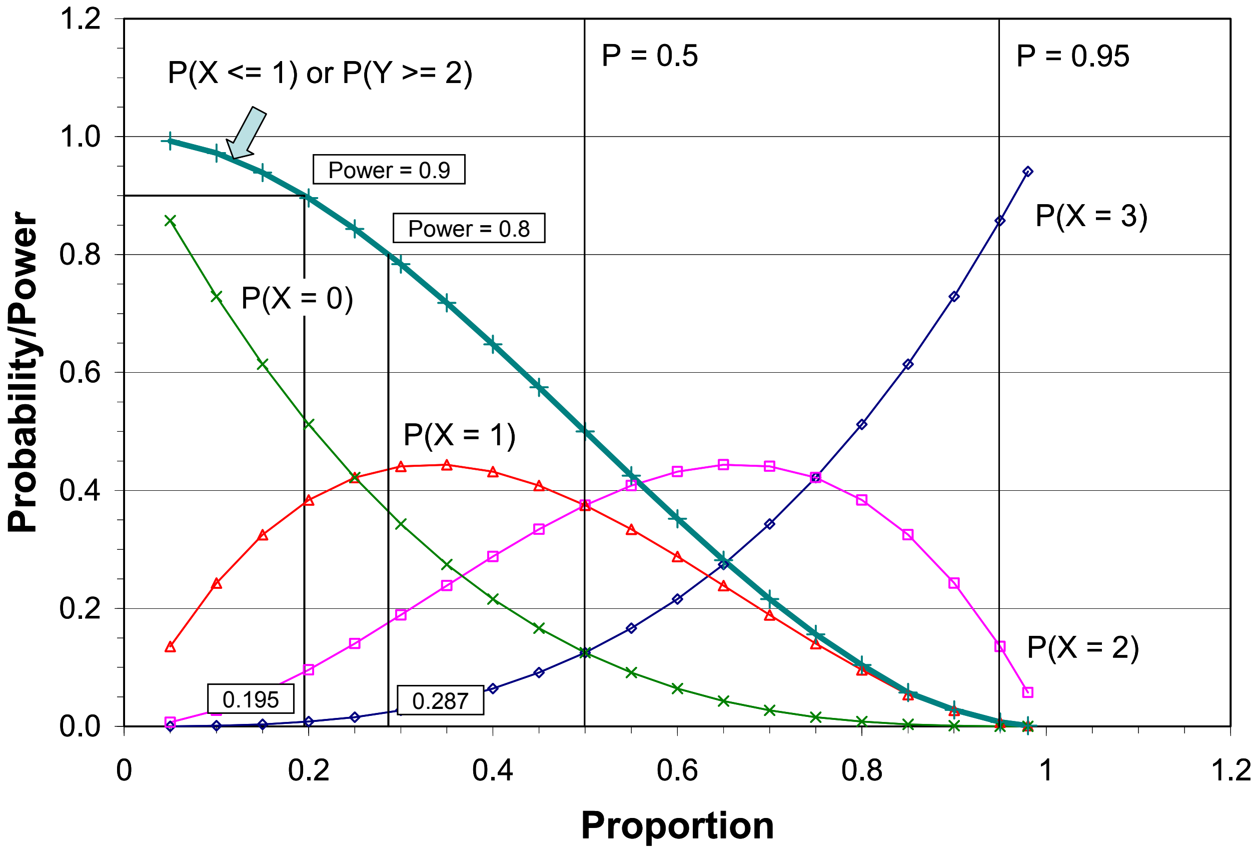


Figure B.5: Probability/power versus proportion of a binomial distribution with n = 3.

Accordingly, the relationship of power versus proportion can be presented as in Figure B.5. The power is monotonically increasing as the proportion decreases. Conventionally, the acceptable power level ranges from 0.8 through 0.9. The corresponding proportions are 0.287 for power 0.8 and 0.195 for power 0.9. That is to say, if the agency wants to achieve a power level of 0.8, the sampling proportion value of contractor must be smaller than 0.287 so that the agency has enough power to reject the null hypothesis H_0: p = 0.95 and thus favor H_1: $p < 0.95$.

In sum, by taking only three samples out of a project, the agency will have insufficient power to reject H_0: p = 0.95 given that H_0 is false unless the quality of the project delivered by the contractor is so poor that the agency is confident enough to reject the project.

APPENDIX C: ACCEPTANCE CRITERIA IN A QA PROCESS

The binomial distribution quality assurance (QA) procedure established should include the following steps:

1. Determination of sample size,
2. Development of a sampling scheme, and
3. Determination of QA process acceptance criteria

C.1 Determination of Sample Size

The determination of sample size is compromised by the following considerations:

1. The determination of sample size for quality assurance (QA) of hot-mix asphalt (HMA) construction is primarily based on an acceptable error level $E = |\hat{p} - p_0|$ for an HMA parameter specified by the agency, as illustrated in Appendix *A.3.2*.
2. As noted in the discussions of statistical simulations, the "large enough" sample size for the binomial distribution is in the range of approximately 100 ~ 200 (Appendix B.1: Figure B.2[d]).
3. The cost and time that the agency is willing to spend will be the primary considerations.
4. As proven in Appendix B.2, by taking only three samples of a project, the agency will have insufficient power to reject H_0: p = 0.95 given that H_0 is false unless the quality of project produced from the contractor is so poor that the agency is confident enough to reject the project.
5. A frequently used rule of thumb is that the approximation is reasonable when $np > 5$ and $n(1-p) > 5$, especially for large values of n. Accordingly, if p = 0.95, then n has to be 100 to fulfill the rule of thumb. (Appendix *A.2.1*)
6. From the above discussions, it seems that sample size 100 is the most compromised size. However, 100 samples is more than one-third the number of dowels of the Haller Wash Bridges (Lt. and Rt.; each bridge has 282 dowels) and about one-tenth of the number of dowels of the Clipper Valley Wash Bridges (Lt. and Rt.; each bridge has 993 dowels).

Accordingly, the decision on sample size is made as follows:

1. The sample size determination is based on each bridge rather than on the whole project.
2. Approximately, one-tenth of the number of dowels of each bridge will be taken for the purpose of quality assurance. The sample size of each bridge is summarized in the following (based on Table B.1):

Bridge Name	# of Dowels per Bridge	# of Samples per Bridge
Van Winkle Wash Bridges (Lt. and Rt.)	534	50
Haller Wash Bridges (Lt. and Rt.)	282	30
Rojo Wash Bridges (Lt. and Rt.)	318	30
Clipper Valley Wash Bridges (Lt. and Rt.)	993	100

After the sample size for each bridge is determined, the next step is to develop the most representative sampling scheme that is random and unbiased. The use of uniform design (UD) as sampling strategy to ensure that the most representative sampling scheme can be achieved is demonstrated in the following sections.

C.2 Development of a Sampling Scheme

C.2.1 Uniform Experimental Design

Statisticians have developed a variety of experimental design methods for different purposes, with the expectation that use of these methods will result in increased yields from experiments, quality improvements, and reduced development time or overall costs. Popular experimental design methods include full factorial designs, fractional factorial designs, block designs, orthogonal arrays, Latin squares, supersaturated designs, etc. One relatively new design method is called *Uniform Design* (UD). Since it was proposed by Fang and Wang in the 1980s *(2, 3),* UD has been successfully used in various fields, such as chemistry and chemical engineering, quality and system engineering, computer sciences, survey design, pharmaceuticals, and the natural sciences, etc.

Generally speaking, uniform design is a space-filling experimental design that allocates experimental points uniformly scattered in the domain. The fundamental concept of UD is to choose a set of experimental points with the smallest discrepancy among all the possible designs for a given number of factors and experimental runs (*4*). For a given measure of uniformity M, a uniform design has the smallest M-value over all fractional factorial designs with n runs and m q-level factors. There are several methods to construct uniform designs such as the good lattice, Latin square method, expending orthogonal design, optimization searching method, etc.

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

Given that the strength of uniform design is that it provides a series of uniformly scattered experimental points over the domain, this homogeneity in two factors has physically become the spatial uniformity of sampling from a bridge section in x and y directions. The application of uniform design to this multibridge retrofit project resulted in the generation of sampling scheme with a UD table for each bridge consisting of pairs of (x, y) coordinates. The unit of the x-axis is the number of locations and the unit of the y-axis is the number of girders.

C.2.2 Bridge Sampling Schemes

The prospective bridge was divided into $n(X)$ (x-direction) × $n(Y)$ (y-direction) cells (or locations). The $n(X)$ represents the number of locations in the x-direction and the $n(Y)$ is the number of girders in the y-direction. N points (sample size) were then assigned to these $n(X) \times n(Y)$ cells according to the table generated by the UD design software. Hence, a sampling scheme was defined by $n(X)$, $n(Y)$, and N. For instance, *x24y3n50* represents 50 samples that were assigned to 50 cells of the 24 × 3 cells. It should be noted that it is possible to assign more than one sample per sampled location. *Note: The dowel (or dowels) sampled per location must be randomly selected with the approval of the agency (Caltrans).*

The UD table not only issues the most representative sampling scheme, but it also gives the agency a more unbiased and random sampling scheme that can be followed in the quality assurance process. The bridge sampling schemes generated by UD tables are plotted in the following:

- Van Winkle Wash Bridge (Rt.): Figure C.1 (*x24y3n50*)
- Van Winkle Wash Bridge (Lt.): Figure C.2(*x24y3n50*)
- Haller Wash Bridge (Rt.): Figure C.3 (*x14y3n30*)
- Haller Wash Bridge (Lt.): Figure C.4 (*x14y3n30*)
- Rojo Wash Bridge (Rt.): Figure C.5 (*x15y3n30*)
- Rojo Wash Bridge (Lt.): Figure C.6 (*x15y3n30*)
- Clipper Valley Wash Bridge (Rt.): Figure C.7 (*x54y3n100*)
- Clipper Valley Wash Bridge (Lt.): Figure C.8 (*x54y3n100*)

Note: The agency is responsible for inspecting whether the dowel bars are or are not fully bonded.

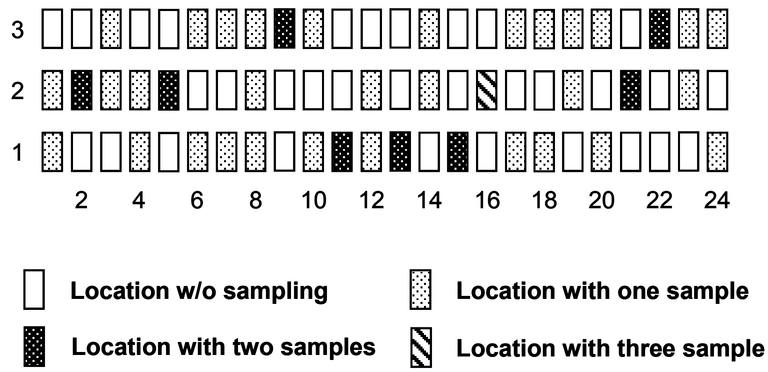


Figure C.1: Sampling scheme for the Van Winkle Wash Bridge (Rt.) (*x24y3n50*).

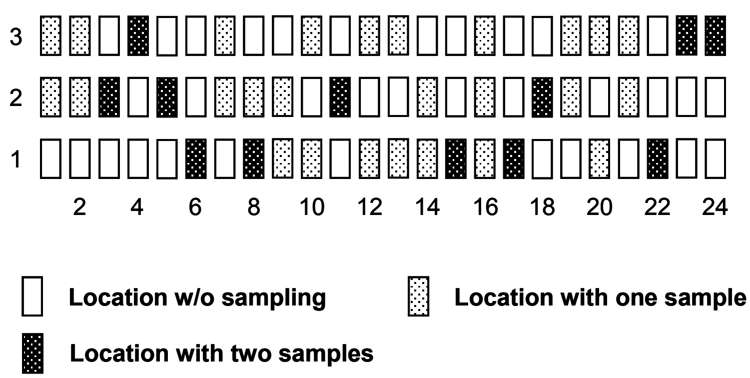


Figure C.2: Sampling scheme for the Van Winkle Wash Bridge (Lt.) (*x24y3n50*).

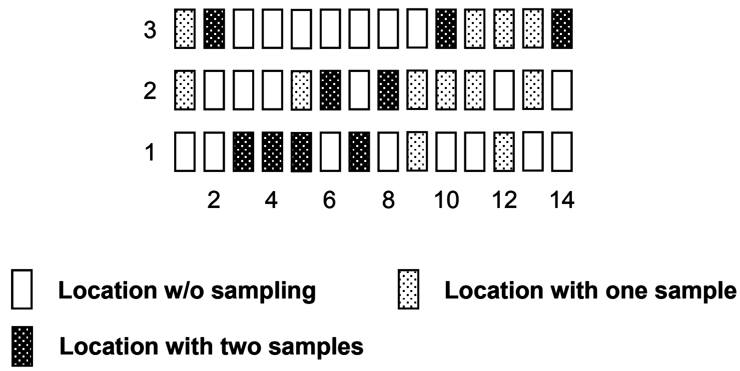


Figure C.3: Sampling scheme for the Haller Wash Bridge (Rt.) (*x14y3n30*).

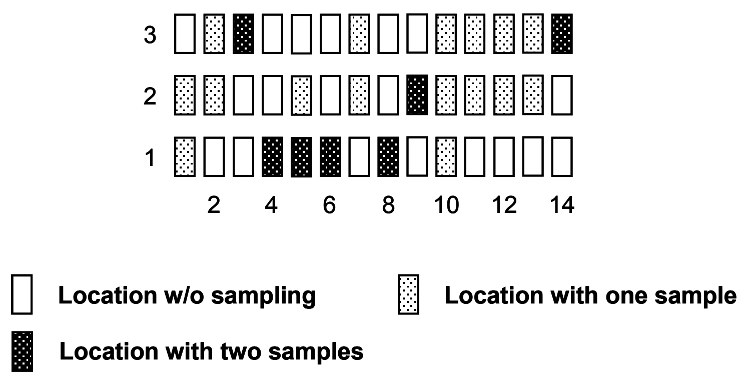


Figure C.4: Sampling scheme for the Haller Wash Bridge (Lt.) (*x14y3n30*).

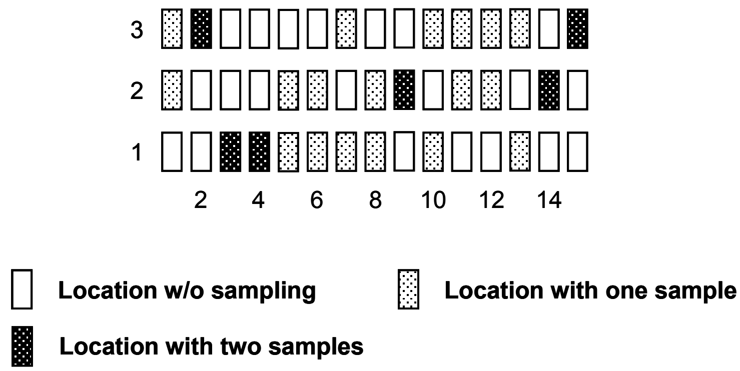


Figure C.5: Sampling scheme for the Rojo Wash Bridge (Rt.) (*x15y3n30*).

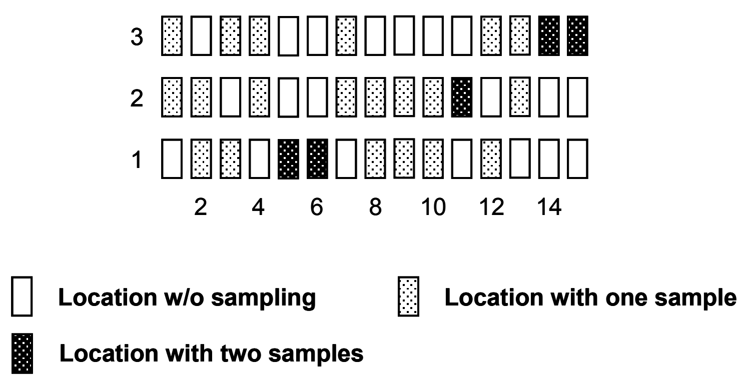


Figure C.6: Sampling scheme for the Rojo Wash Bridge (Lt.) (*x15y3n30*).

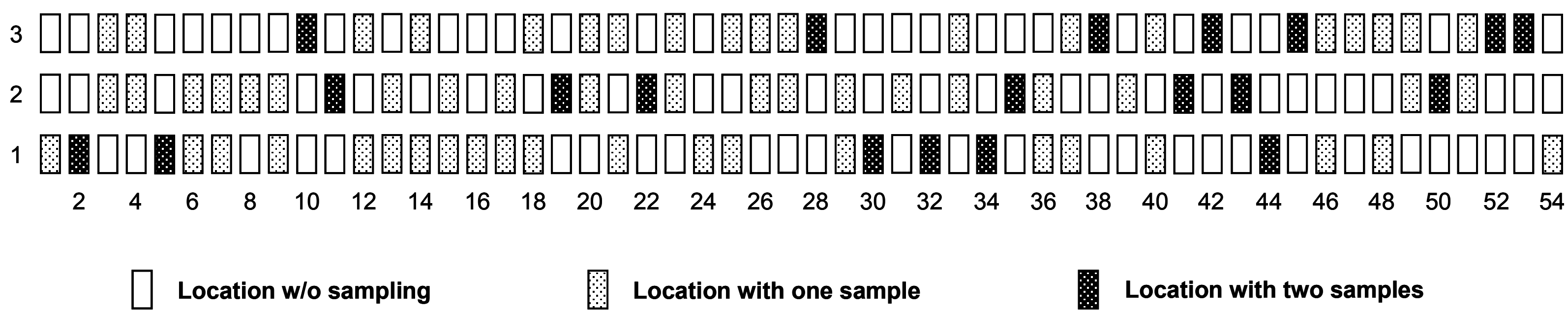


Figure C.7: Sampling scheme for the Clipper Valley Wash Bridge (Rt.) (*x54y3n100*).

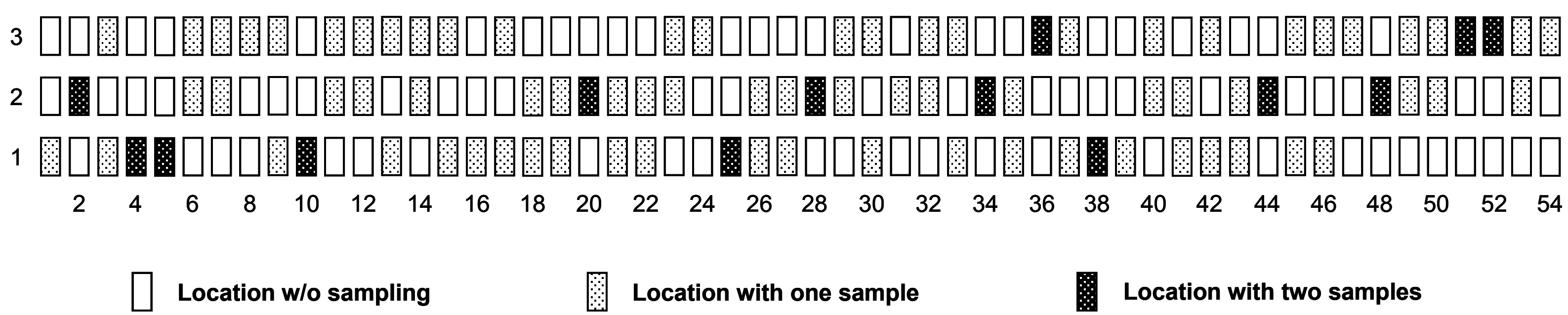


Figure C.8: Sampling scheme for the Clipper Valley Wash Bridge (Lt.) (*x54y3n100*).

C.3 Acceptance Criteria for a QA Process

Once the sampling scheme is determined, the acceptance criteria for a QA process is needed to ensure that the acceptance level is obtained.

Figure C.9 presents the binomial distributions (in terms of count of failures) with various sample sizes (*n* = 30, 50, and 100) and various proportions (0.95, 0.9, 0.8, and 0.7). As can be seen, the figure indicates that the normal approximation of a binomial distribution seems to be rational and the normal approximation is more apparent as the sample size increases. As mentioned in Section 2.3.2, the use of hypothesis testing of $H_0 : p = 0.95$ and $H_1 : p < 0.95$ to establish the performance specification is more appropriate than use of the other two hypotheses: $H_0 : \mu = 0.95$ and $H_1 : \mu \neq 0.95$ or $H_0 : p = 0.95$ and $H_1 : p > 0.95$. It is recognized that the acceptance criterion is determined by null and alternative hypotheses, power level, proportion, sample size, and α level. Figure C.10 plots the relationship of power versus estimated proportion at various sample sizes (*n* = 30, 50, 100) under the hypothesis testing $H_0 : p = 0.95$ and $H_1 : p < 0.95$ and $\alpha = 0.05$.

To establish the acceptance criterion, first the agency has to determine the power level that it is confident enough to correctly reject H_0 if H_0 is not true. *It is recommended that power = 0.8 be specified to establish the acceptance criterion.* Let $\hat{p}$ be the estimated proportion and *Y* the count of failures based on the sampling result of a QA process from the specified bridge sampling scheme. For example, the interpretation of Figure C.10 under the hypothesis testing $H_0 : p = 0.95$ and $H_1 : p < 0.95$ at *n* = 30 and power = 0.8 is that the agency will have at least 0.8 power to reject the null hypothesis $H_0 : p = 0.95$ and favor the alternative hypothesis $H_1 : p < 0.95$ if $\hat{p}$ < 0.826 (i.e., *Y* > 5); otherwise, if $p \geq 0.826$, then the agency will have insufficient power to reject the null hypothesis. Therefore, the acceptance criterion is specified such that if there are more than five failures, then the agency has more power than 0.8 to reject $H_0 : p = 0.95$ and favor $H_1 : p < 0.95$. The acceptance criteria for each bridge are summarized in Table C.1.

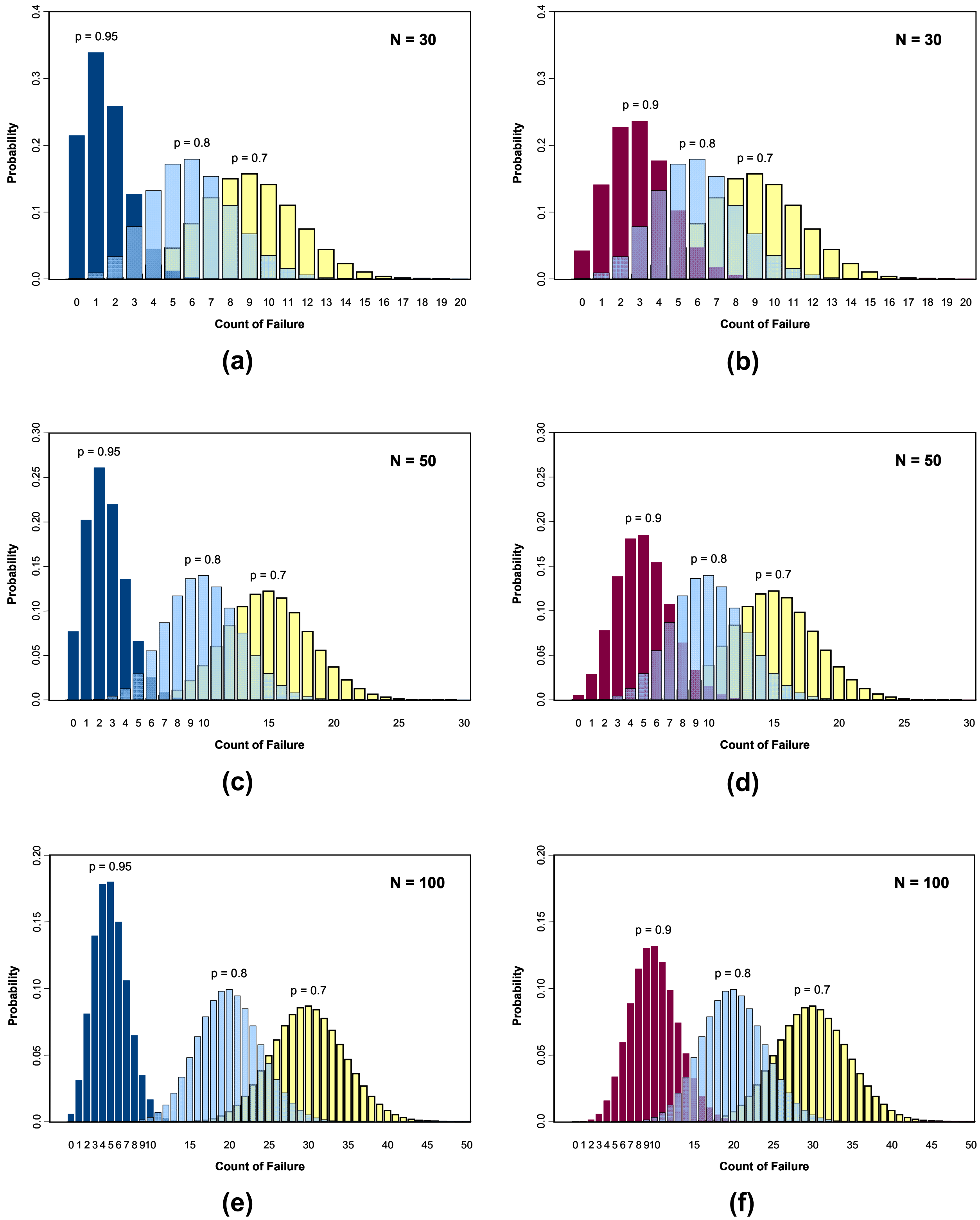


Figure C.9: Binomial distributions with various sample sizes (*n* = 30, 50, and 100) and proportions (0.95, 0.9, 0.8, and 0.7).

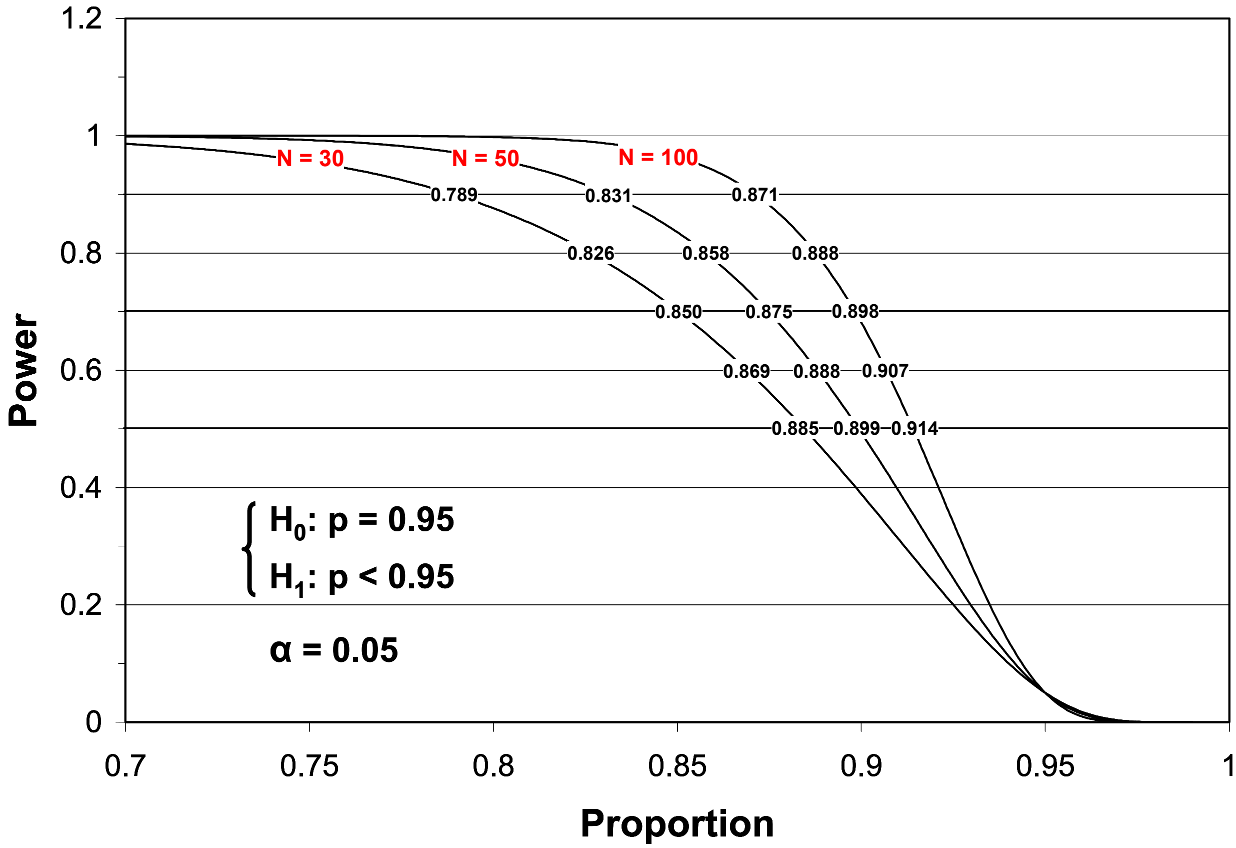


Figure C.10: The relationship of power versus estimated proportion at various sample sizes (n = 30, 50, and 100) under the hypothesis testing $H_0 : p = 0.95$ and $H_1 : p < 0.95$ with α = 0.05.

Table C.1: Acceptance Criteria at Various Power Levels for Each Bridge

Bridge Name	No. of Locations	No. of Dowels	Sample Size	Sampling Strategy	Power Level	Acceptance Criterion	
						Proportion	Count of failures
Van Winkle Wash Bridge (Rt.) (with polyester overlay)	72	534	50	Figure C.1	0.5	P ≥ 0.899	Y ≤ 5
					0.6	P ≥ 0.888	Y ≤ 5
					0.7	P ≥ 0.875	Y ≤ 6
					0.8	P ≥ 0.858	Y ≤ 7
					0.9	P ≥ 0.831	Y ≤ 8
Van Winkle Wash Bridge (Lt.) (with polyester overlay)	72	534	50	Figure C.2	0.5	P ≥ 0.899	Y ≤ 5
					0.6	P ≥0.888	Y ≤ 5
					0.7	P ≥ 0.875	Y ≤ 6
					0.8	P ≥ 0.858	Y ≤ 7
					0.9	P ≥ 0.831	Y ≤ 8
Haller Wash Bridge (Rt.) (without polyester overlay)	42	282	30	Figure C.3	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	P ≥ 0.826	Y ≤ 5
					0.9	P ≥ 0.789	Y ≤ 6
Haller Wash Bridge (Lt.) (without polyester overlay)	42	282	30	Figure C.4	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	P ≥ 0.826	Y ≤ 5
					0.9	P ≥ 0.789	Y ≤ 6
Rojo Wash Bridge (Rt.) (with polyester overlay)	45	318	30	Figure C.5	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	P ≥ 0.826	Y ≤ 5
					0.9	P ≥ 0.789	Y ≤ 6
Rojo Wash Bridge (Lt.) (with polyester overlay)	45	318	30	Figure C.6	0.5	P ≥ 0.885	Y ≤ 3
					0.6	P ≥ 0.869	Y ≤ 3
					0.7	P ≥ 0.850	Y ≤ 4
					0.8	P ≥ 0.826	Y ≤ 5
					0.9	P ≥ 0.789	Y ≤ 6
Clipper Valley Wash Bridge (Rt.) (with polyester overlay)	162	993	100	Figure C.7	0.5	P ≥ 0.914	Y ≤ 8
					0.6	P ≥ 0.907	Y ≤ 9
					0.7	P ≥ 0.898	Y ≤ 10
					0.8	P ≥ 0.888	Y ≤ 11
					0.9	P ≥ 0.871	Y ≤ 12
Clipper Valley Wash Bridge (Lt.) (with polyester overlay)	162	993	100	Figure C.8	0.5	P ≥ 0.914	Y ≤ 8
					0.6	P ≥ 0.907	Y ≤ 9
					0.7	P ≥ 0.898	Y ≤ 10
					0.8	P ≥ 0.888	Y ≤ 11
					0.9	P ≥ 0.871	Y ≤ 12

APPENDIX D: CALTRANS BRIDGE SHEAR RETROFIT PROJECTS

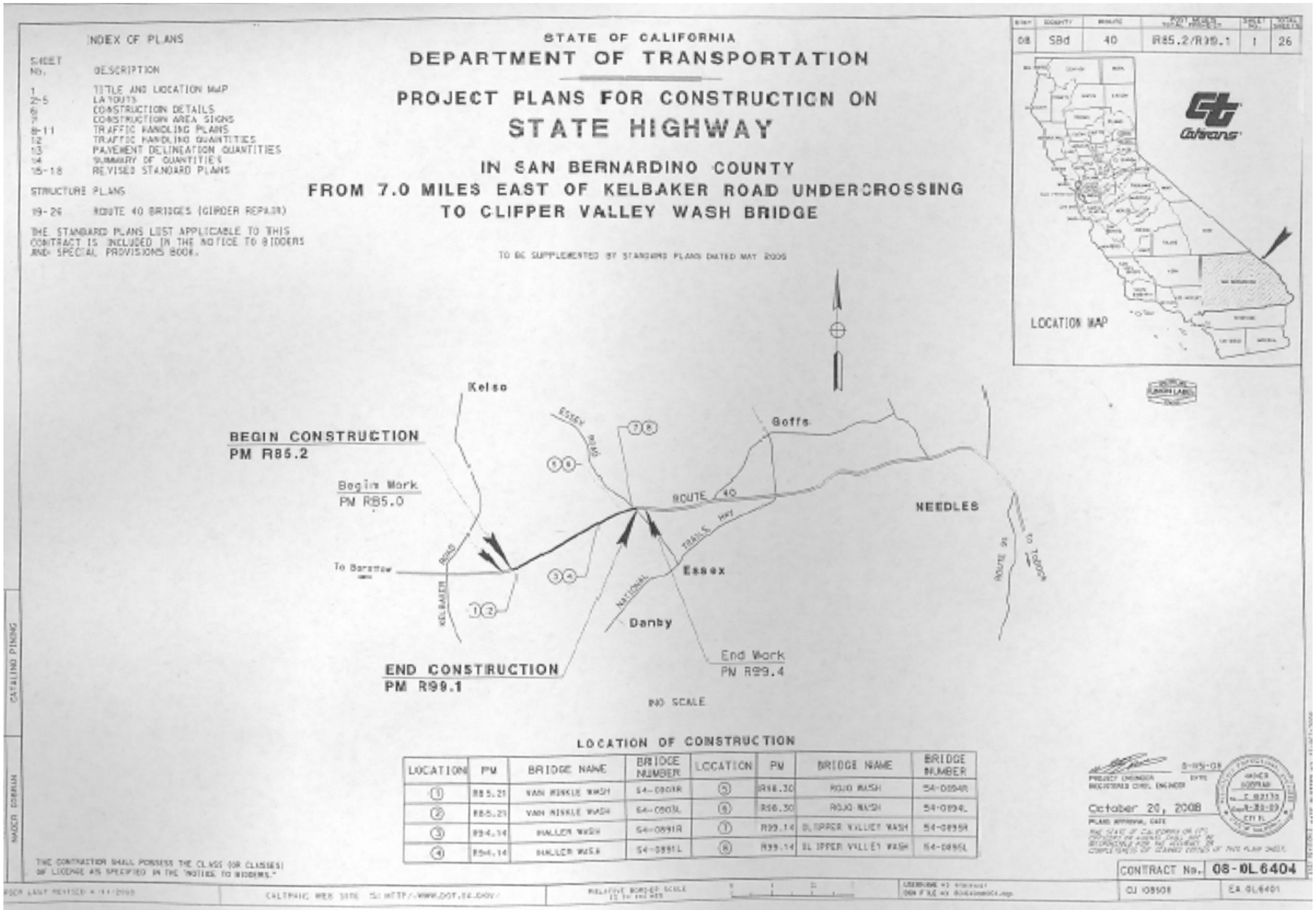


Figure D.1: Project plans for the Caltrans bridge shear retrofit projects in San Bernardino County from 7.0 miles east of the Kelbaker Road undercrossing to the Clipper Valley Wash Bridge.

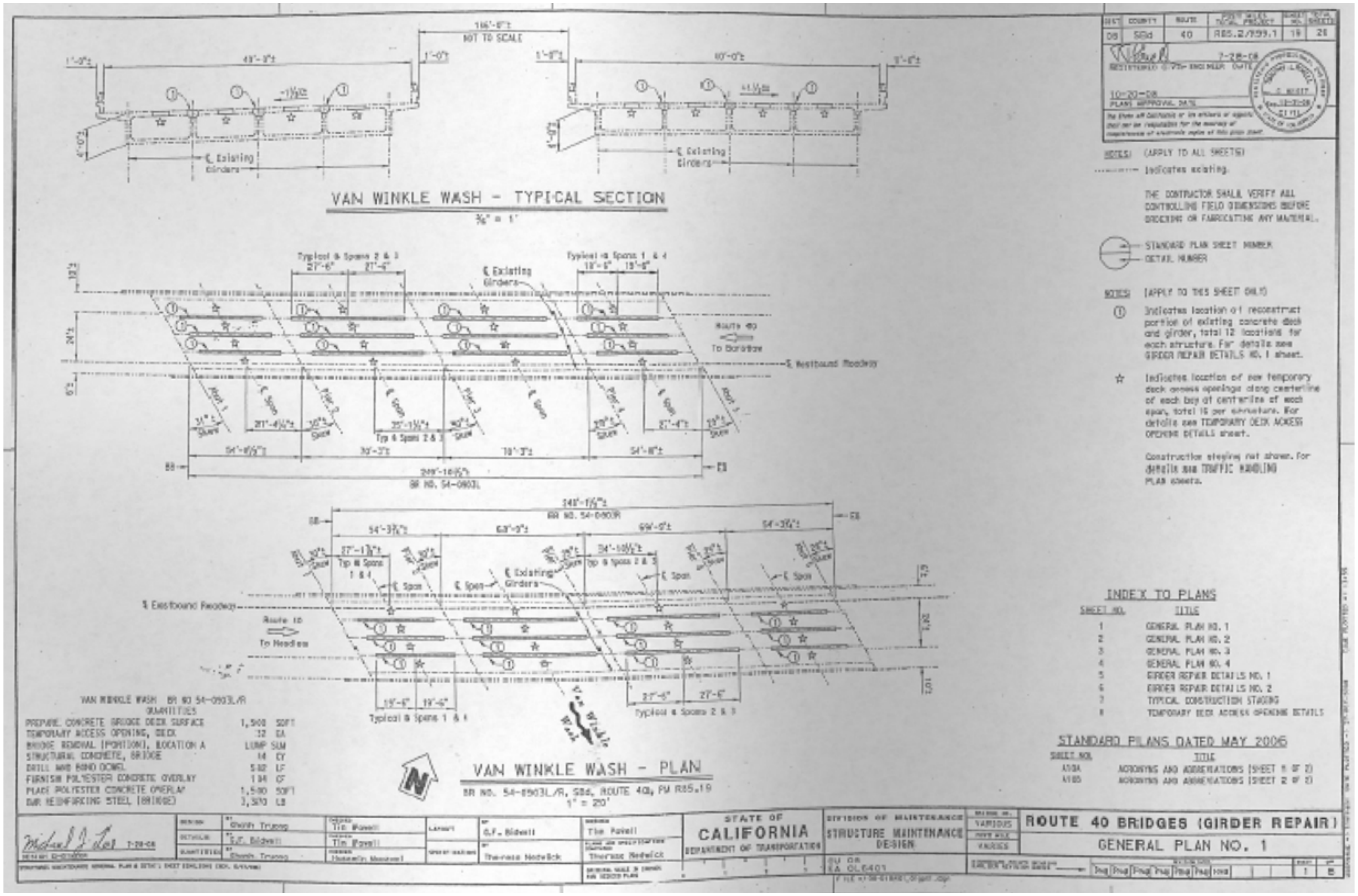


Figure D.2: Typical section and plan for the Van Winkel Wash Bridges (Rt. and Lt.).

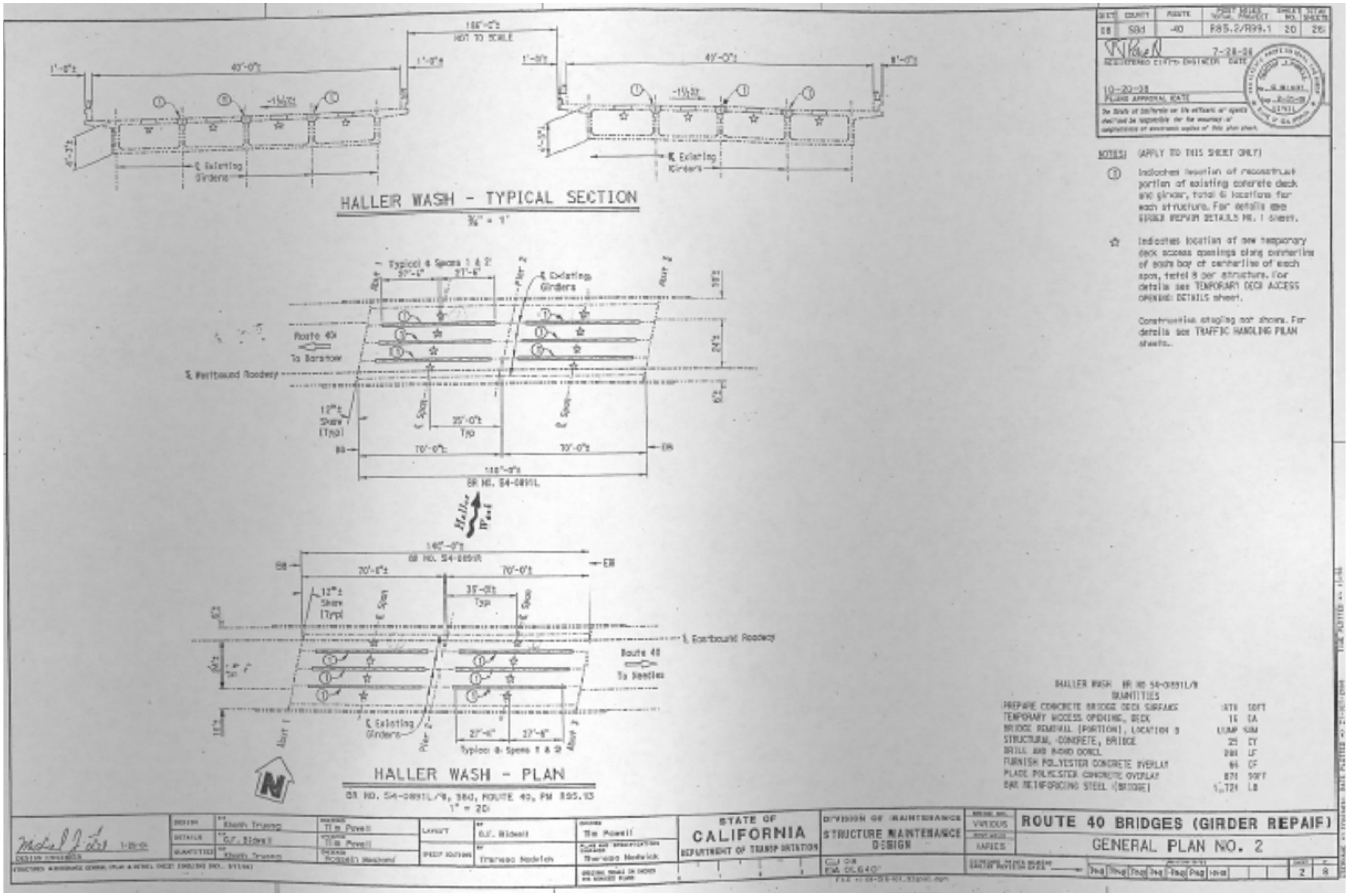


Figure D.3: Typical section and plan for the Haller Wash Bridges (Rt. and Lt.).

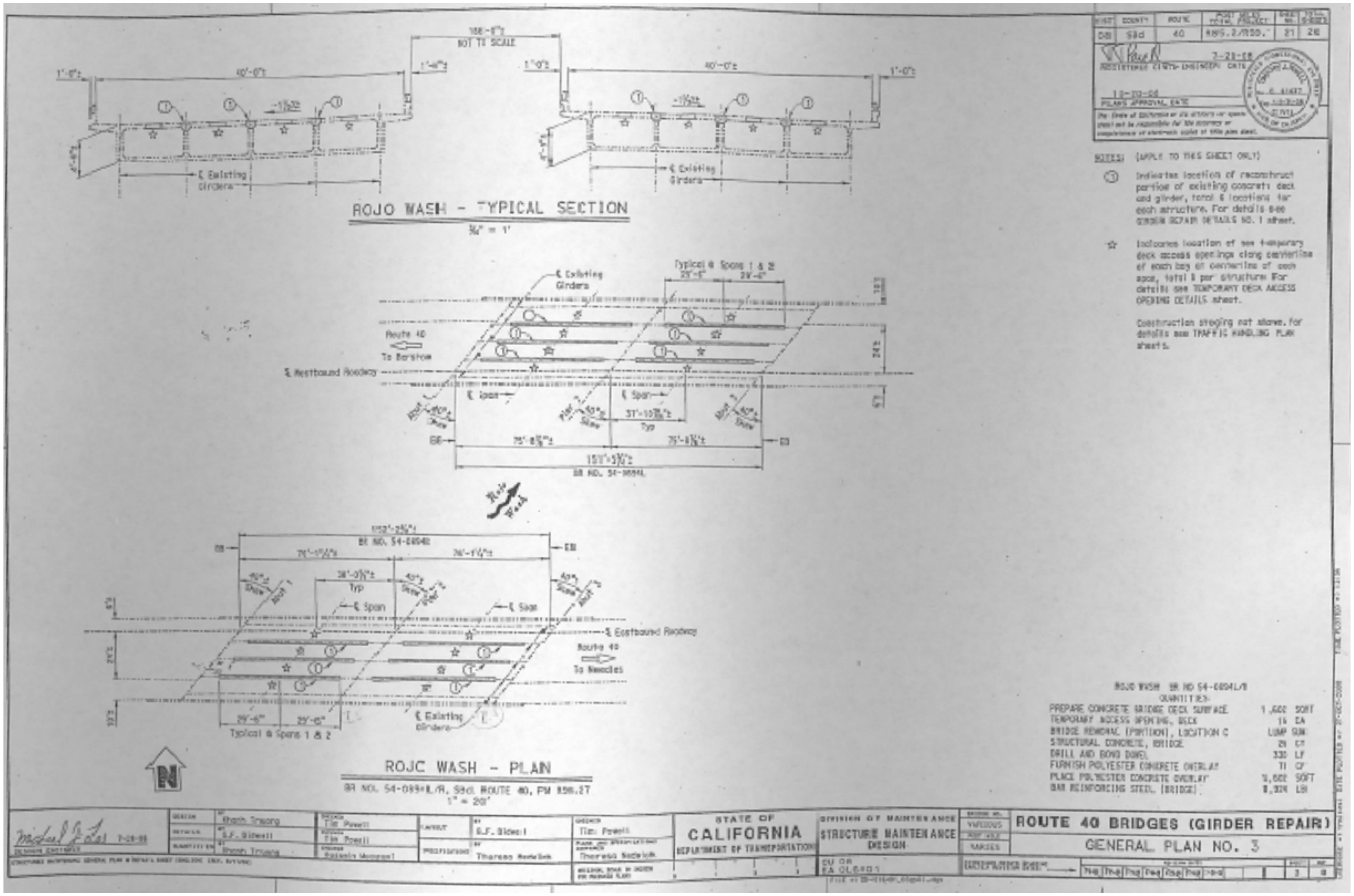


Figure D.4: Typical section and plan for the Rojo Wash Bridges (Rt. and Lt.).

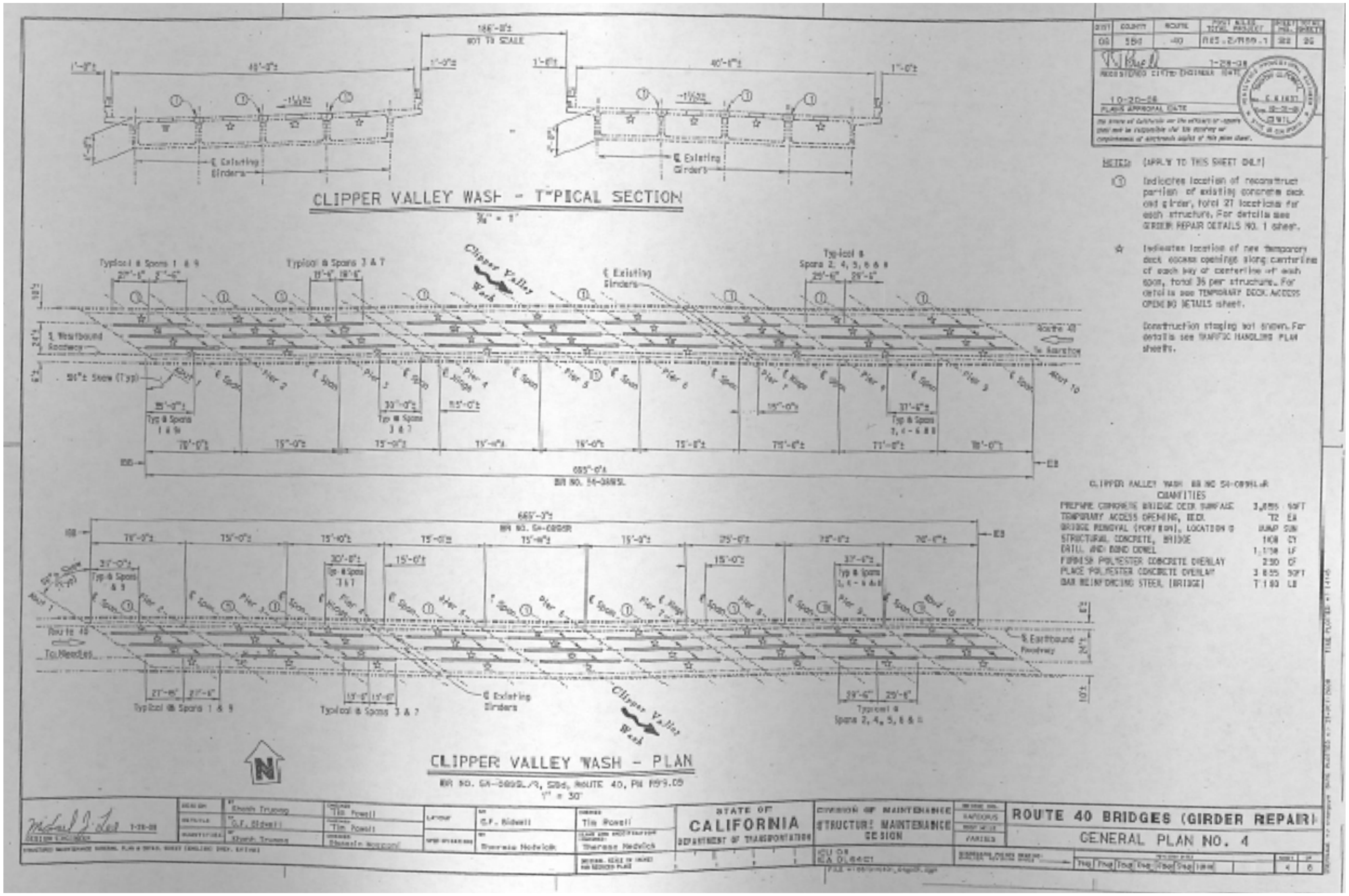


Figure D.5: Typical section and plan for the Clipper Valley Wash Bridges (Rt. and Lt.).

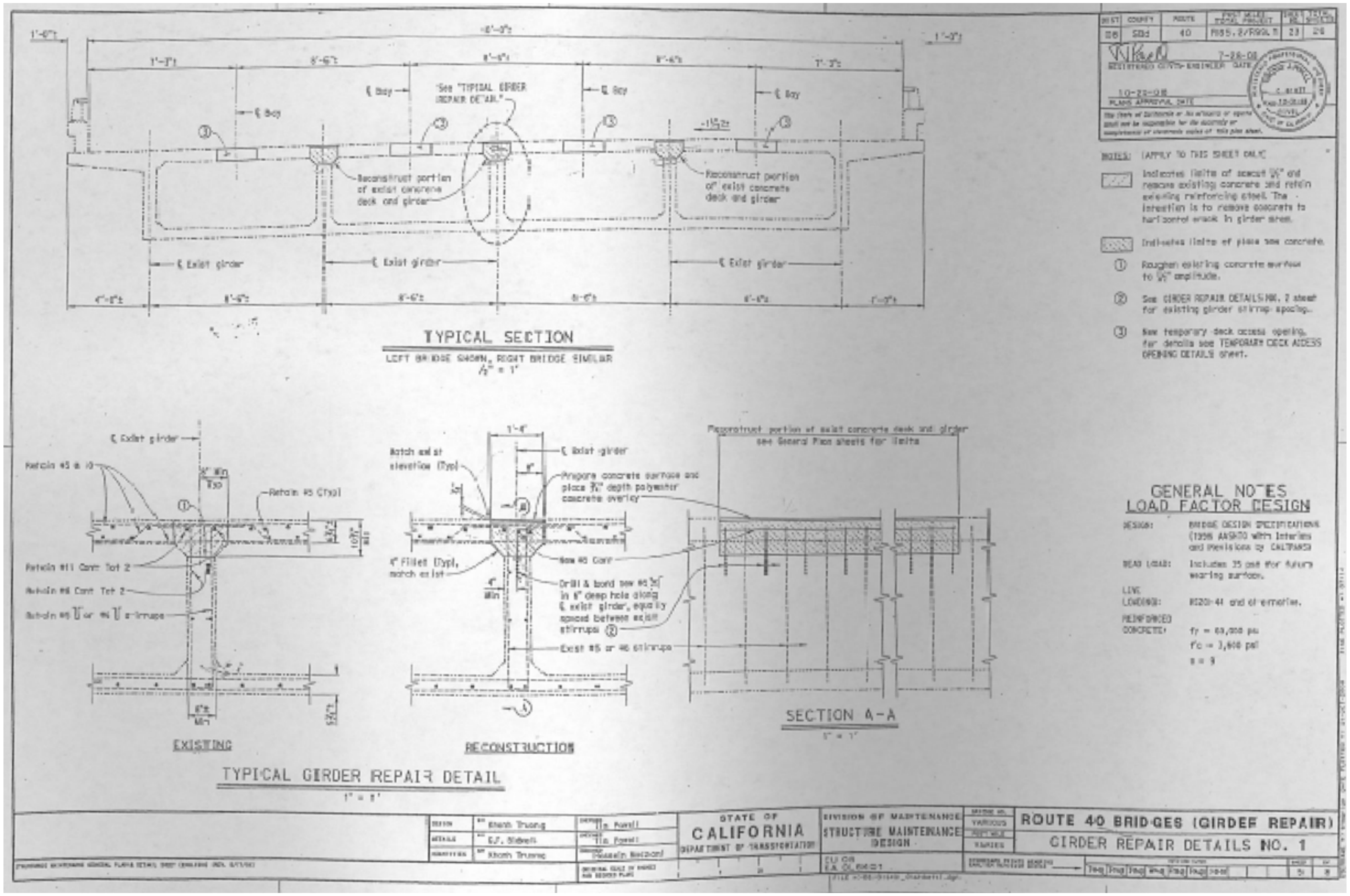


Figure D.6: Girder repair details (No. 1).

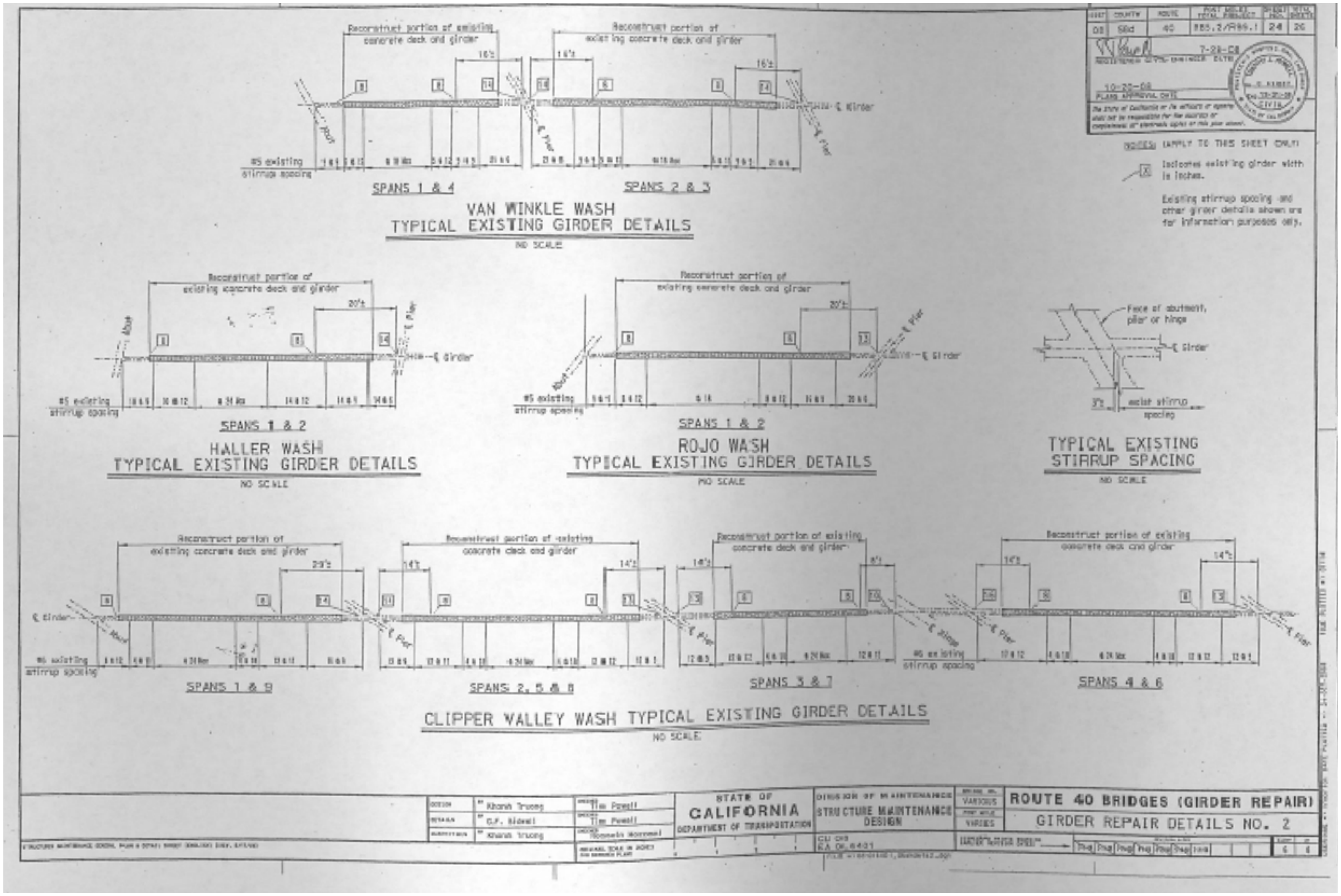


Figure D.7: Girder repair details (No. 2).

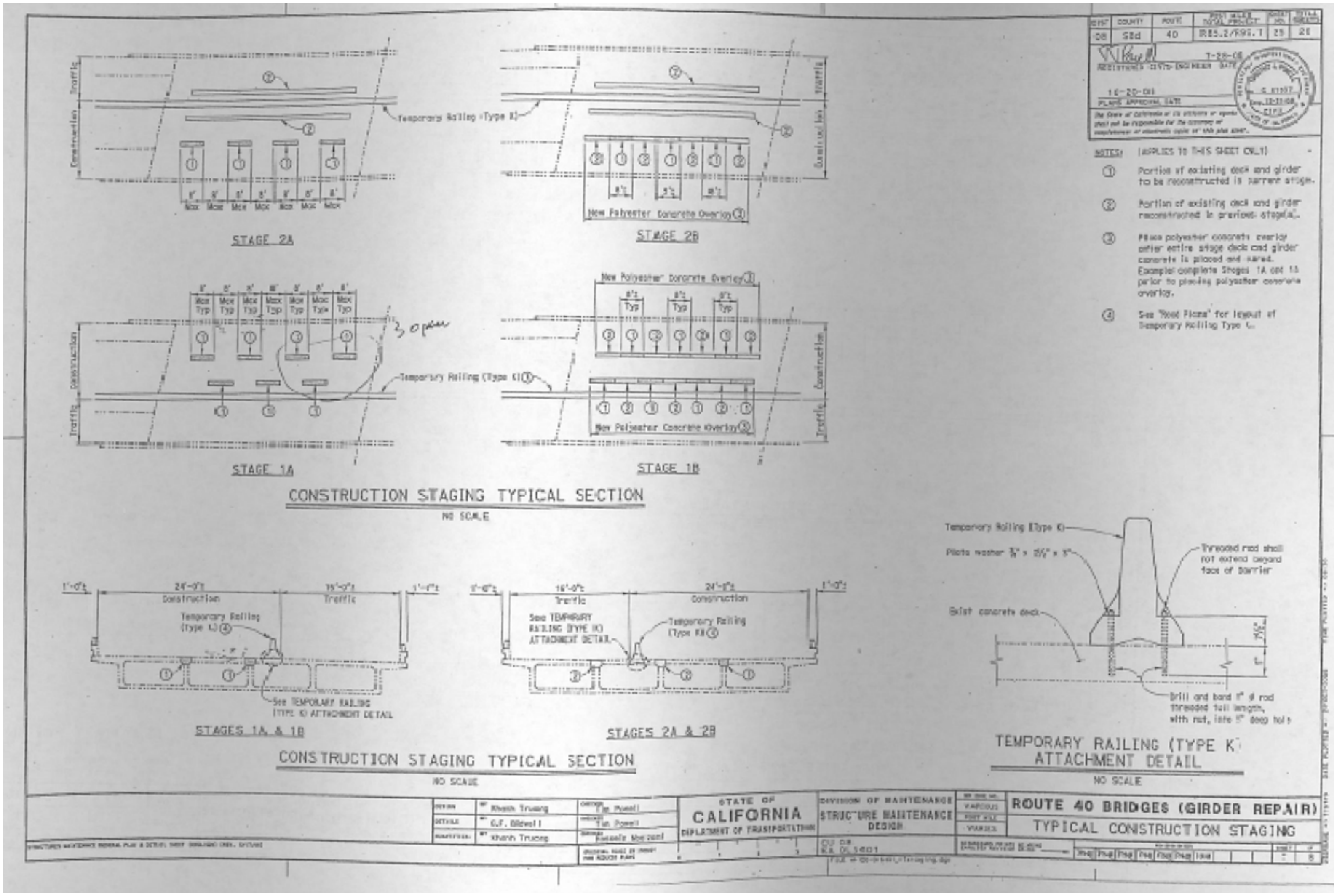


Figure D.8: Typical construction staging.

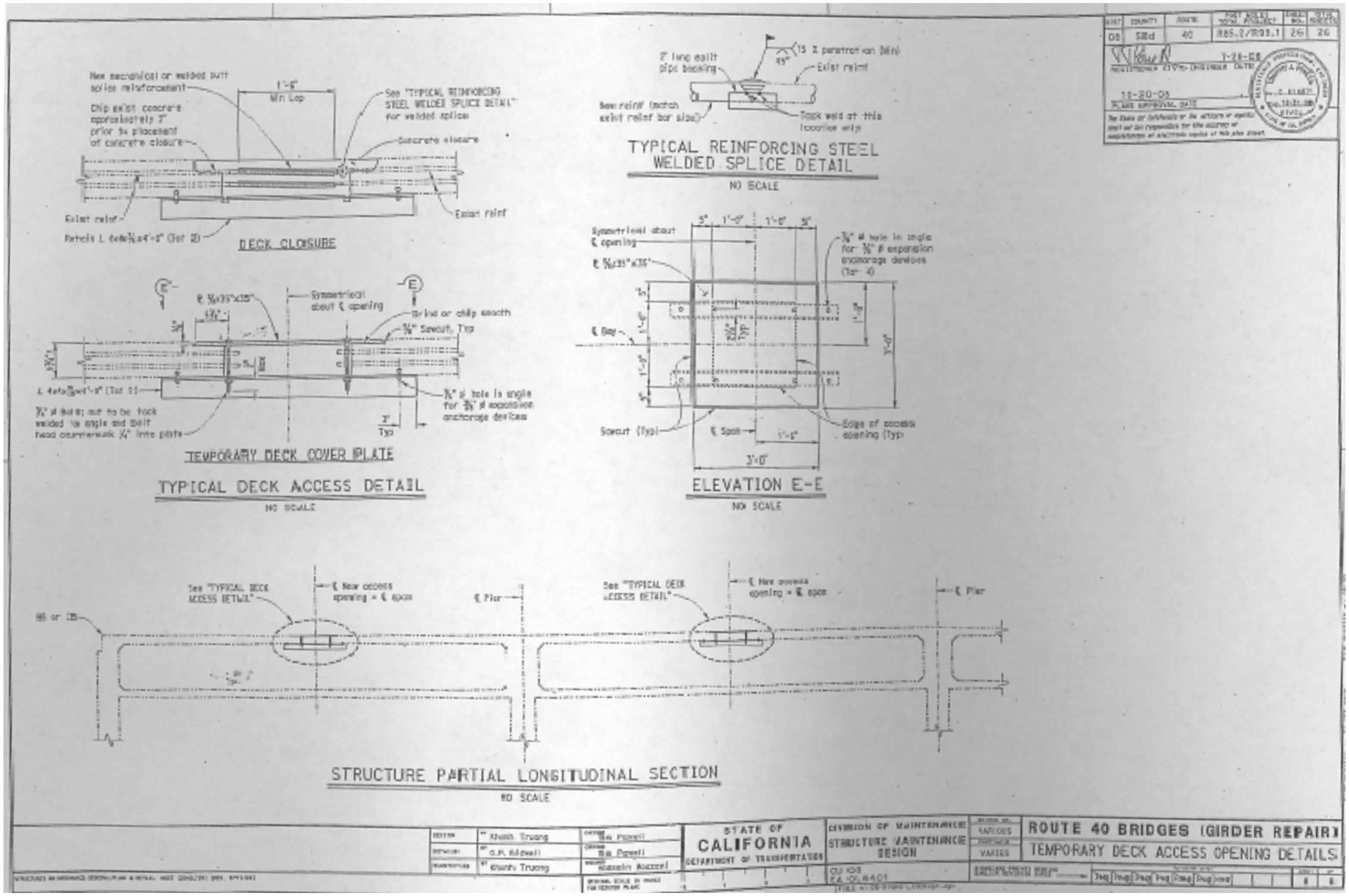


Figure D.9: Temporary deck access opening details.