NTIA Report 95-319 (2)

Performance Evaluation of Data Communication Services: NTIA Implementation of American National Standard X3.141 Volume 2. Experiment Design

Martin J. Miles

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

Larry Irving, Assistant Secretary
for Communications and Information

August 1995

CONTENTS

CONTENTS (Cont'd)

FIGURES

FIGURES (Cont'd)

TABLES

Page

PERFORMANCE EVALUATION OF DATA COMMUNICATION SERVICES: NTIA IMPLEMENTATION OF AMERICAN NATIONAL STANDARD X3.141

VOLUME 2. EXPERIMENT DESIGN

Martin J. Miles[1]

The six volumes of this report are:

Volume 1. Overview
Volume 2. Experiment Design
Volume 3. Data Extraction
Volume 4. Data Reduction
Volume 5. Data Analysis
Volume 6. Data Display

This volume shows how to design an experiment to evaluate the performance of a data communication service as specified by ANS X3.141. It reviews some statistical concepts required for experiment design and analysis, including dependence between consecutive trials. It discusses objectives from both the user's and the vendor's point of view. It describes the criteria for selecting performance parameters and their appropriate analysis. It discusses the conditions that define the population of each function of a data communication service. It shows how to assign performance values (some of which cause excessive delays to be considered failures). It shows how to conduct a preliminary test to characterize the service. This test provides sample sizes required to achieve specified precisions, and it estimates the test duration for each. It discusses possible designs for the experiment and how to determine the sample size if a preliminary characterization test is either not conducted or if the objective of the experiment is to determine the acceptability of a performance parameter value.

Key words: American National Standards; analysis of variance; confidence level; data communication services; dependent trials; factors; performance measurements; performance parameters; precision; sample size

1. INTRODUCTION

The first part of this introduction discusses the contents of this volume, and the remainder is a review of the statistical concepts required for experiment design and subsequent analysis (Volume 5). Specifically, these concepts concern estimation, confidence intervals, sample size, precision, hypotheses, and tests of significance. The statistical theory contained in this six volume report is summarized in a book authored by M. J. Miles and E. L. Crow (to be published).

[1]The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80303.

1.1 Contents of this Volume

Section 2 discusses some likely objectives of the experiment and the appropriate analysis for each. The two principals most likely to be interested in the experiment are the user and the vendor, and the scope and precision of the experiment should be matched with the economic effects for either. ANS X3.141 recommends one (or more) of four types of analysis of a performance parameter: estimation, acceptance tests, comparison tests, and tests of effects of existing conditions. Figure 1 contains two structured design diagrams that show the relationship among the major activities of experiment design. Specifically, Figure 1a shows the relationship for acceptance and maintenance, and Figure 1b shows the relationship for other objectives. The figure also lists the number of the section in which each activity is discussed.

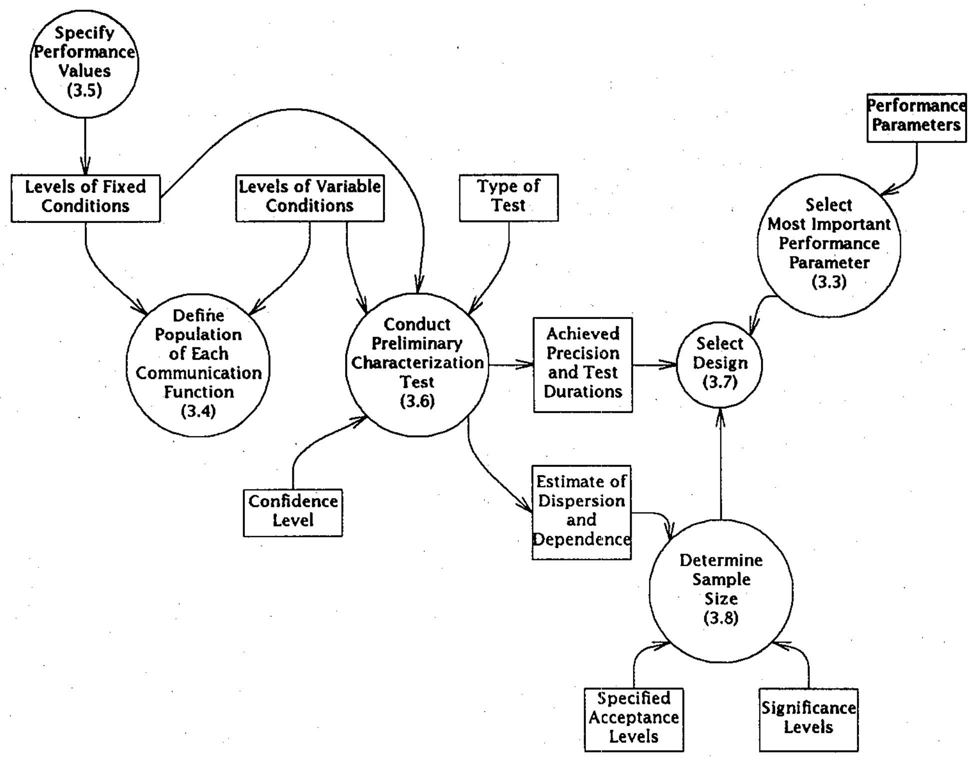
Two types of data communication tests can be conducted: tests of access-disengagement parameters and tests of user information transfer parameters. Section 3 shows how to select the most important performance parameter for each type of test. The first step in selecting performance parameters is to determine which data communication function is most important to your objective. The next step is to determine which performance parameter within that function is most important.

Section 4 shows how to define the population of each data communication function. The trials of each data communication function come from a population whose characteristics are the performance parameters. The populations are defined by conditions existing during the experiment (both fixed and variable conditions).[2] The fixed conditions are those for which there are either no options (such as a single feature supplied by a vendor) or a single level that the experimenter will choose (such as block size). The variable conditions provide more than one level and, hence, an opportunity to design the experiment; <u>the experiment is conducted over selected combinations of levels of the variable conditions</u>.

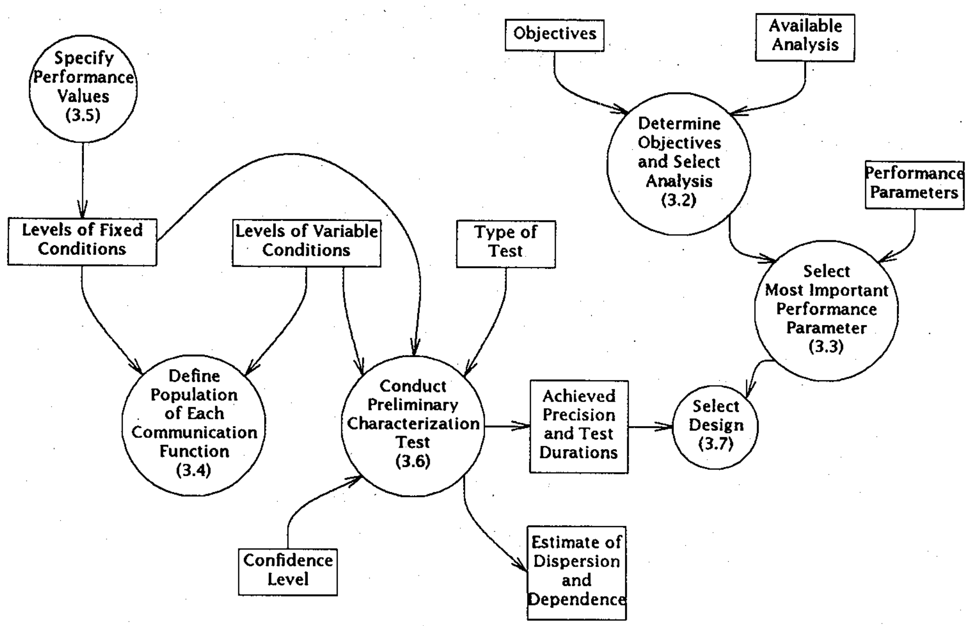
Section 5 discusses how to specify performance values for some parameters. When a delay from a trial is sufficiently long, the trial must be considered a

---

[2]The word "factor" is statistical parlance to describe a variable condition with more than one level (or value). In this report a variable condition is called a factor only after it has been shown to significantly influence the trials of a test (i.e., shown to be a factor).



a. Acceptance and Maintenance



b. Objectives Other Than Acceptance and Maintenance

Figure 1. Structured design diagrams of experiment design.

failure. For example, after a certain period of time an access attempt is no longer an observation of the performance parameter, Access Time, but, perhaps, an observation of the performance parameter, Access Denial Probability. The performance of the parameter, Transfer Denial Probability, is somewhat difficult to define because it is defined in terms of the performance of four other parameters.

Section 6 shows how to conduct a preliminary characterization test and determine the test durations for various precisions. Since the values of the performance parameters are not known - possibly even within an order of magnitude - it is prudent to conduct a preliminary test to characterize the service. For each performance parameter, the preliminary characterization test estimates

- the value,
- the dispersion and dependence,
- the sample size that achieves each of five selected precisions, and
- the corresponding test duration.

All experiments must be conducted within a given budget and/or duration. Section 7 discusses several designs that might be used. The favored design is the full factorial design because it includes all combinations of all levels of the variable conditions. However, this design is sometimes too complex.

Section 8 shows how to determine the sample size when a preliminary characterization test is either impractical or does not provide it (as when the objectives are acceptance or maintenance).

Appendix A presents the mathematical formulas that determine the sample size required to obtain a specified precision. These formulas account for dependence between consecutive trials (i.e., first-order Markov dependence). Program **star** determines the sample size for single tests (and analyzes either single or multiple tests).

Appendix B contains the flowcharts for **star** and for each subroutine required to determine the sample size.

Appendix C shows how to determine the sample size by using **star** interactively.

Appendix D discusses the shell scripts and programs that implement the preliminary characterization test.

As might be expected, many decisions must be made during the design phase of the experiment. To stress the importance of experiment design, Table 1 lists the initial decisions required of the experimenter in this phase.

1.2 Statistical Concepts

This section introduces statistical concepts such as populations, their characteristics, precision, confidence levels, confidence limits, etc. A review of them should answer some questions concerning experiment design and the subsequent analyses. Even though the NTIA implementation accounts for possible dependence between consecutive trials, the following discussion assumes independence.

1.2.1 Density Functions and Their Parameters

Drawing conclusions about the general (i.e., the population) from knowledge of the specific (i.e., the observed sample) is called inductive inference. Even though this procedure results in uncertain success, statistical methodology allows us to measure the uncertainty and reduce it to a known, tolerable level.[3] The population is the totality of elements that have one or more characteristics to be estimated. If the number of elements is finite, the population is said to be real. If the number of elements is infinite, the population is said to be hypothetical. For example, access attempts in an unending series of data communication sessions are elements from a hypothetical population. The number of elements is said to be countable if they can correspond one-to-one with the natural numbers; this would be the case for access attempts.

The number of elements observed from the population is called the sample size. Each element in the sample is called a trial or an observation, and the value of the subject characteristic is called the outcome. If one is to infer

---

[3]In contrast, deductive inference always yields a correct conclusion because the conclusion results from a chain of proved conclusions (e.g., mathematical theorems are proved by deductive inference).

Table 1. Decisions Required to Design an Experiment of Data Communication Services

• Determine the objective of the experiment.*
• Select the analysis to be performed.
• Select the most important performance parameter from the access-disengagement function.
• Select the most important performance parameter from the user information transfer function.
• Identify the fixed conditions that exist during the experiment.
• Identify the variable conditions and their levels that exist during the experiment.**
• Specify the performance values.
• Select the single most representative combination of levels of variable conditions for each type of test.
• Select the block size and number of blocks.
• Specify the intertrial delay for each type of test.
• Select the confidence level.
• Select combinations of levels of the variable conditions.
• Specify the desired precision.
• Estimate the dispersion of the most important time parameters.***
• Estimate the autocorrelation of lag 1 of the most important time parameters.
• Estimate the conditional probability (of a failure given that a failure occurred on the previous trial) of the most important failure probability parameter.

*Even though ANS X3.141 states that determining the objective is part of the experiment design phase, it seems that it would be known before the experimenter decides to implement the standard.

**The word "factor" is statistical parlance to describe a variable condition having more than one level (or value). In this report a variable condition is called a factor only after it has been shown to significantly influence the trials of a test.

***This and the following two estimates are obtained if the recommended preliminary characterization test is conducted.

something about the population from a sample, care must be taken that the following two conditions are met:

- Intended Population. Elements in the sample must come from the intended (target) population. (Otherwise the conclusions drawn from the sample are probably representative of another population).

- Random Sample. The sample must be a random sample (i.e., a sample from a hypothetical population whose outcomes are independent).

For real (finite) populations, a random sample can be obtained only if the elements are replaced after each sampling. For hypothetical populations, randomness of the sample is not influenced by replacement.

A random variable, X, is a function that assigns a real value to each element in the set of all possible random samples (i.e., the sample space). Almost always, random variables are either of the discrete type or of the continuous type:[4]

A. Random Variables of the Discrete Type

Suppose the number of values is either finite or countable, and the random variable can assume the value x_i ($i = 1, 2, \ldots$) with probability $P(X = x_i) = p_i$. The random variable, X, is of the discrete type if

$$p_i \geq 0 \qquad \text{for all } i,$$

and

$$\begin{cases} \sum_{i=1}^{n} p_i = 1 & \text{(for a finite number of possible values of } x_i) \\ \sum_{i=1}^{\infty} p_i = 1 & \text{(for a countable number of possible values of } x_i). \end{cases}$$

The probability function is

$$P(X = x_i) = p_i ,$$

[4]Very rarely, they have both discrete and continuous components.

and the distribution function is

$$F(x) = P(X \le x) = \sum_{x_i \le x} p_i$$

(i.e., for all x for which $x_i \le x$).

B. Random Variables of the Continuous Type

A random variable, X, is of the continuous type if there exists a function, f, such that

$$f(x) \ge 0 \qquad \text{for all real } x ,$$

and

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du.$$

The density function is f(x), and the distribution function is F(x). The values obtained from a sample of a continuous random variable have a pattern called the sample density. To determine the sample density from such a sample, several steps are required:

1. Order the trial values.
2. Determine the range of trial values.
3. Partition the range into appropriate intervals (as indicated by the number of trials and range of values).
4. Record the number of values occurring in each interval.
5. Determine the relative frequency of occurrence of trials in each interval (by dividing the number of values in each interval by the total number of values).

The shape of the sample density can then be seen by plotting the relative frequencies for each interval. This plot is called a histogram.

Example: The following sixteen values have been sampled from a population: 5.1, 7.9, 3.2, 0.9, 1.2, 2.1, 3.3, 6.7, 6.5, 5.1, 6.2, 5.4, 3.4, 6.3, 6.9, 8.2. Determine the sample density.

Solution:

1. Order the values: 0.9, 1.2, 2.1, 3.2, 3.3, 3.4, 5.1, 5.1, 5.4, 6.2, 6.3, 6.5, 6.7, 6.9, 7.9, and 8.2.

2. The values range from 0.9 to 8.2.

3. Appropriate intervals seem to be [0,2), [2,4), [4,6), [6,8,) and [8,10).

4. and 5. These two steps are demonstrated by Table 2.

Table 2. Example of a Sample Density

INTERVAL	STEP 4 NUMBER OF VALUES	STEP 5 RELATIVE FREQUENCY
[0,2)	2	0.1250
[2.4)	4	0.2500
[4.6)	3	0.1875
[6,8)	6	0.3750
[8,10)	1	0.0625
-	16	1.0000

Since sample density functions are composed of relative frequencies, they are non-negative functions that sum to 1. Figure 2 is the histogram corresponding to Table 2. If the sample had been much larger, the range could have been divided into many smaller intervals. Then if the sample were a random sample, the sample density would have more nearly approximated the population density.

The population (true) value of a parameter can be estimated by a function of the values obtained from a random sample. These functions are called statistics. The most common statistic is the sample mean. The sample mean of the previous example is

$$\overline{x} = \frac{1}{16}\ (5.1 + 7.9 + \ldots + 8.2) = 4.9.$$

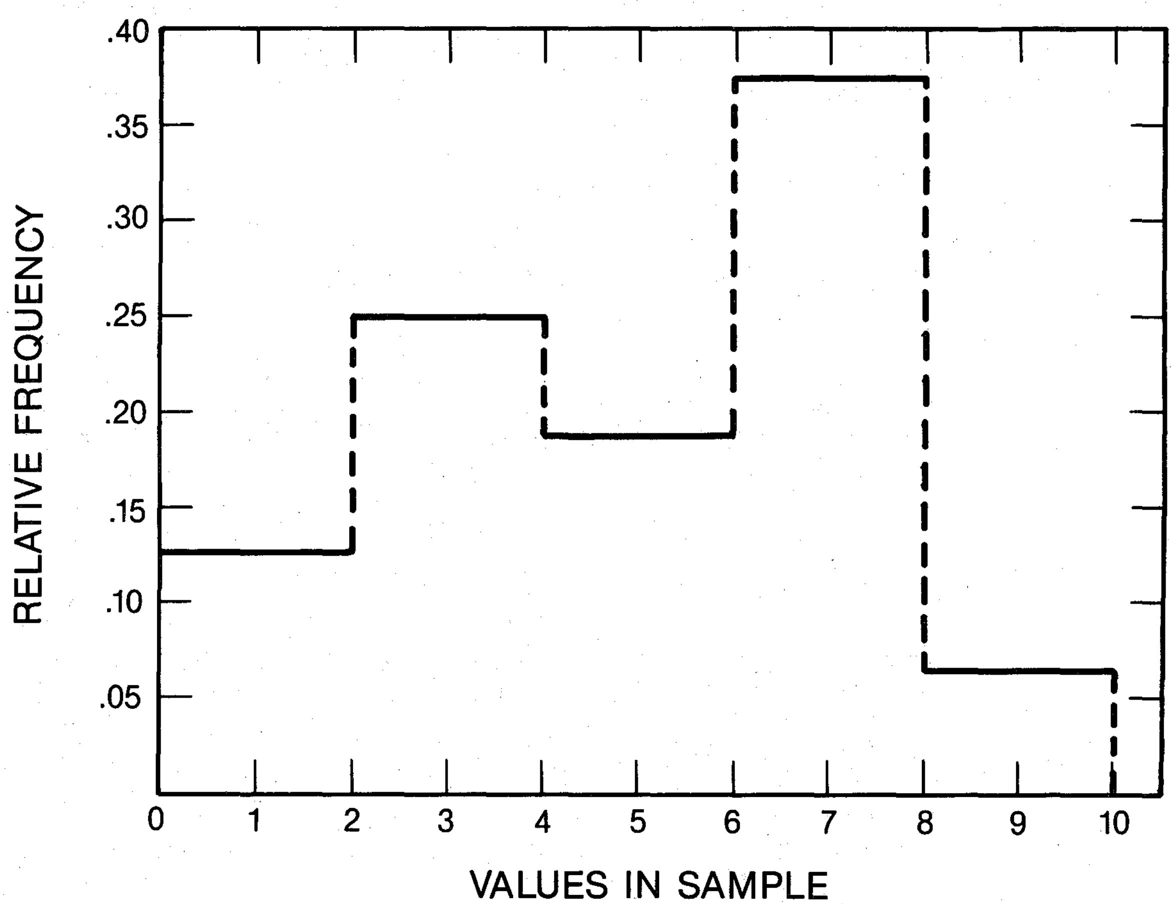


Figure 2. Example of a sample density.

When the mean is computed from the entire population, the statistic is called the expected value of x or, simply, the population mean (usually denoted by μ). Another important statistic is the sample variance. It estimates the dispersion of the population values about the mean. The population variance is usually denoted by σ^2; the square root of the variance is called the standard deviation and is denoted by σ.

Figure 3 shows the normal density function and the gamma density function (both for a random variable of the continuous type). Figure 3 also shows the binomial probability function and the Poisson probability function (both for a random variable of the discrete type). The figure lists distribution parameters, means, and variances (expressed in terms of these parameters). The normal density is an example of a two-parameter density (whose parameters also happen to be the population mean and variance).

1.2.2 Estimating the Population Mean

The random sample can indicate something about the value of the population mean.[5] This can be demonstrated from the following two important observations of sampling:

- Sample Size. Suppose we want to obtain a sample for which the sample mean deviates from the population mean by less than a given amount. We can determine the sample size required to bring the probability of this as close to 1 as is desired. Moreover, the required sample size is essentially independent of the shape of the population density function (but dependent on the population variance).

- Sampling Density. Suppose random samples, each of size n, are repeatedly drawn from a population. After each draw, the sample mean is determined. The set of sample means so obtained have their own sample density (called a sampling density). In fact, regardless of the density of the parent (original) population, the sample mean has approximately the normal density: Its mean equals the parent population mean, but its variance is only 1/n as large as the parent population variance. Hence, as, n, the sample size, becomes larger, the variance of the sample mean becomes proportionally smaller. Thus the sample means are closely clustered about the population mean when n is large. Figure 4 shows a parent population which is a gamma density (with mean 4 and variance 2). The sample mean, obtained from samples drawn from this population, is shown as having a density that is very close to the normal density for samples of size 10, 20, and 30.

These two observations of sampling show that knowledge of the sample mean allows us to infer knowledge of the population mean, and the inference is less uncertain

[5]Provided only that the population has a mean and a finite variance.

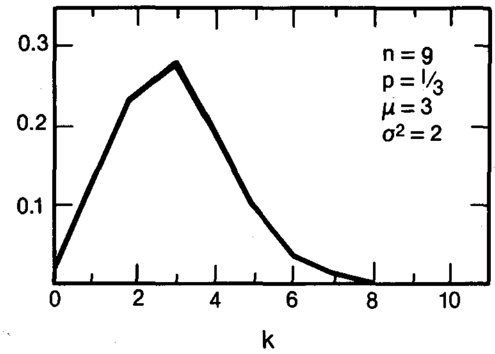
BINOMIAL PROBABILITY FUNCTION

$P(X=k) = \binom{n}{k} p^k(1-p)^{n-k} \quad k=0, 1, \cdots, n$

Parameters: $1 \leqslant n$, $o<p<1$

Mean: $\mu = np$

Variance: $\sigma^2 = np\,(1-p)$



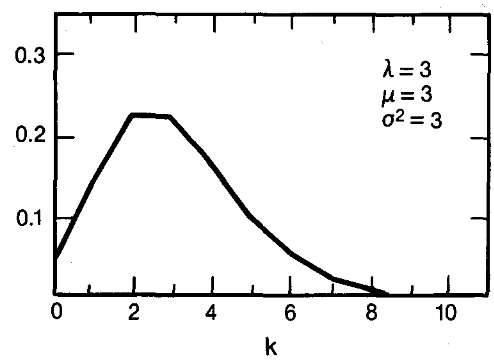
POISSON PROBABILITY FUNCTION

$P(X=k) = \frac{\lambda}{k!} \cdot e^{-\lambda} \qquad k=0, 1, 2, \cdots$

Parameters: $o<\lambda$

Mean: $\mu = \lambda$

Variance: $\sigma^2 = \lambda$



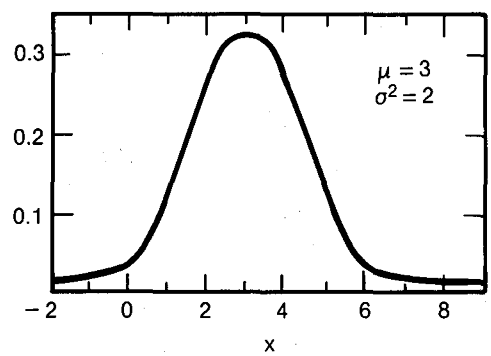
NORMAL DENSITY FUNCTION

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Parameters: μ, σ (positive)

Mean: $\mu = \mu$

Variance: $\sigma^2 = \sigma^2$



GAMMA DENSITY FUNCTION

$$f(x) = \begin{cases} \frac{b^p}{\Gamma(p)} \cdot x^{p-1}e^{-bx} & 0 \leqslant x \\ 0 & \text{Otherwise} \end{cases}$$

Parameters: b, p (both positive)

Mean: $\mu = \frac{p}{b}$

Variance: $\sigma^2 = \frac{p}{b^2}$

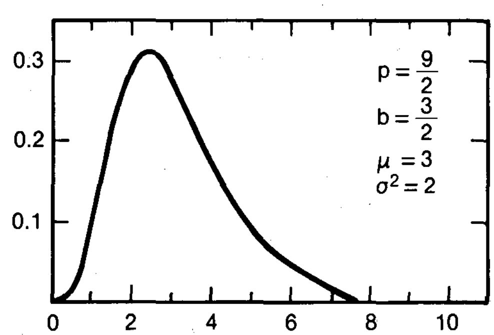


Figure 3. Common probability functions and their density functions.

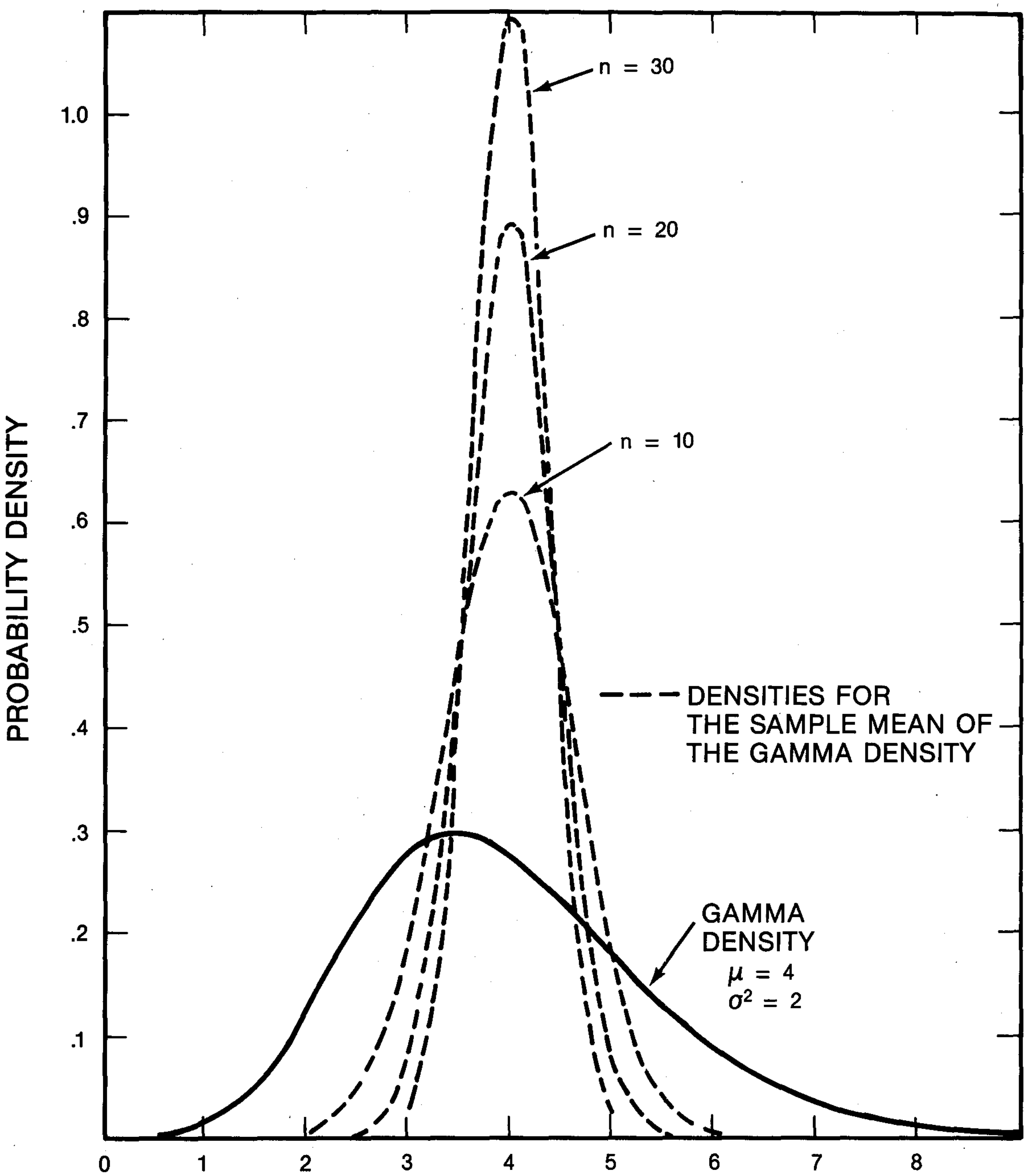


Figure 4. gamma density and its sample mean densities.

when the population variance is small. However, in order to benefit from these observations, it is necessary to quantify them. This is done in the following two sections concerning confidence intervals and precision.

1.2.3 Confidence Intervals

The estimate of the mean should be accompanied by some interval about the estimate together with some assurance that the population mean is in the interval. Such an interval is called a confidence interval. For example, a confidence interval for the population mean is called a 95% confidence interval if we can be 95% confident that it contains the population mean.

In theory, many samples are drawn and a confidence interval is determined from each sample. The end points of each confidence interval depend upon the values in the sample. Then we expect that, say, 95% of all the intervals determined from the samples include the population mean (a single fixed point). However, in practice, a single confidence interval is determined from the values in the sample, and we say that we have a certain confidence that the population mean is in that interval. Confidence intervals can also be defined for parameters other than the population mean (e.g., the population variance).

Although dependence between consecutive trials is taken into account by the NTIA implementation, independence is assumed in this introduction. Suppose random samples are obtained from a population (not necessarily normal) with mean μ and variance σ^2. Suppose also that $\bar{x}$ and s^2 are the estimates of these population values obtained from a sample of size n. The confidence interval can be determined for populations having the normal density and for populations having a non-normal density if the number of samples is large (say, $n \geq 30$).[6] The confidence intervals for these two cases are identical when the population variance is known, but differ when it is not known:

A. Population Variance is Known

From the second observation in Section 1.2.2, $\bar{x}$ has approximately a normal density with mean μ and variance σ^2/n. Hence, the random variable,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}},$$

has the normal density with mean zero and variance one. This normal density is called the standard normal density and is tabulated in most statistics books.

---

[6]Provided the density has a finite variance.

Suppose z_1 and z_2 are the end points of a probability interval for z, and $z_1 < z < z_2$. If we want to determine, say, a 95% confidence interval for μ, we choose z_1 and z_2 so that 95% of the probability of the density (i.e., 95% of the area under the curve) is in the interval. An infinite number of intervals can be defined that include the specified probability. However, the values of z_1 and z_2 that provide the shortest interval are those that are also symmetric about the mean.

It is customary to refer to the confidence level as a $100(1-2\alpha)$% confidence level because the α, so specified, can be used to re-label z_1 and z_2 in terms of the chosen confidence level. For the standard normal density, $z_1 = -z_\alpha$ and $z_2 = z_\alpha$. For example, a 95% confidence level is defined by $\alpha = 0.025$ [i.e., $100(1 - 2 \times 0.025)\% = 100 \times 0.95\% = 95\%$]. The end points, $-z_\alpha = -z_{0.025}$, and $z_\alpha = z_{0.025}$, exclude 2.5% of the probability (i.e., area) from each tail of the density.

From

$$-z_\alpha < z < z_\alpha$$

we obtain

$$-z_\alpha < \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} < z_\alpha ,$$

and then

$$\overline{x} - \frac{\sigma}{\sqrt{n}} \cdot z_\alpha < \mu < \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot z_\alpha .$$

This is the $100(1-2\alpha)$% confidence interval for the population mean when the population variance is known. Notice that the length of the interval

$$\frac{2\sigma}{\sqrt{n}} \cdot z_\alpha$$

is fixed because it does not depend upon the values from the sample; however, the end points will vary because $\overline{x}$ depends upon the values from the sample.

B. Population Variance Is Not Known

It is somewhat unrealistic to suppose that σ is known. (If it were, μ would also probably be known). However, if σ is not known, it is reasonable to estimate it by s (determined from a sample of size n).

However, the random variable,

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}},$$

does not have the standard normal density as does z (particularly for small n). This is because s replaces σ in the expression, and s is a random variable whose value depends upon the random samples. The density for it was determined by Gosset (under the pseudonym Student). The parameter, n, provides only n-1 degrees of freedom (choices) from which to determine s because the values in the sample have the constraint that they must also determine $\overline{x}$ (an independent linear relation). Hence, t is said to have Student's t density with n-1 degrees of freedom. The end-point is labelled, $t_{n-1,\alpha}$. This density has the same mean as does the standard normal density (i.e., 0), but because s is only an estimate of σ, there is more uncertainty, and its variance is larger. Again, as n increases, the uncertainty decreases and the variance approaches that of the standard normal (i.e., 1).[7]

For the 100(1-2α)% confidence level and n-1 degrees of freedom,

$$-t_{n-1,\alpha} < t < t_{n-1,\alpha} .$$

From this inequality and the above equation we obtain:

$$\overline{x} - \frac{s}{\sqrt{n}} \cdot t_{n-1,\alpha} < \mu < \overline{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1,\alpha} .$$

---

[7] $t_{n-1,\alpha} > z_\alpha$ for all n, and $\lim_{n\to\infty}\left(\frac{t_{n-1,\alpha}}{z_\alpha}\right) = 1$. For $\alpha = .025$ and $n = 10$, $t_{9,.025} = 2.262$, so the ratio is 1.154.

The length of the interval,

$$\frac{2s}{\sqrt{n}} t_{n-1,\alpha}$$

is dependent on s, and the end points are dependent on both $\bar{x}$ and s; hence, both the length and the end points vary with the values from the sample. The formulas for the confidence limits are summarized in Table 3.

1.2.4 Sample Size and Precision

To determine the confidence limits for the population mean, we can specify any two of three quantities, and the remaining quantity can be determined from formulas. These three quantities are

- the confidence level,
- the sample size, and
- the length of the confidence interval.

However, it is customary to specify the confidence level. Then either the sample size or the length of the confidence interval must be specified. The sample size is specified when the budget (i.e., sampling time) is important, and the length of the confidence interval is specified when the precision of the estimate is important.

When precision is important, the required sample size is determined by specifying either the absolute precision or the relative precision:[8]

A. Absolute Precision

If the population mean is roughly known, say to an order of magnitude, the half-length of the confidence interval can be stated as the maximum acceptable error with which the sample mean estimates the population mean. This error corresponds to the absolute precision. Suppose m_L and m_U are lower and upper

[8]Mathematical statisticians distinguish between precision and accuracy. Precision describes the closeness of trial values to each other (as measured by the standard error, Equation A-7 in Appendix A). Accuracy describes the closeness of an estimate to the population value. Precision and accuracy should be related; however, a high degree of precision can coexist with poor accuracy if the sample is not random or is not drawn from the target population (i.e., biased).

Table 3. Confidence Limits for the Population Mean Depending on Either the Density of the Parent Population or the Number of Random Samples and Knowledge of the Population Variance

	DENSITY OF THE PARENT POPULATION		
	NORMAL	OTHER	
		$n \geq 30$	$n < 30$
σ KNOWN	$\bar{x} \pm z_\alpha \bullet \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm z_\alpha \bullet \frac{\sigma}{\sqrt{n}}$	(not definable)
σ UNKNOWN	$\bar{x} \pm t_{n-1,\alpha} \bullet \frac{s}{\sqrt{n}}$	$\bar{x} \pm z_\alpha \bullet \frac{s}{\sqrt{n}}$	(not definable)

confidence limits for μ, and x is the sample mean. The absolute error, a, is then

$$a = \frac{m_U - m_L}{2} .$$

In the case of a sample from a normal population (or a non-normal population where $n \geq 30$) in which the population standard deviation is known, the length of the confidence interval is

$$m_U - m_L = \frac{2\sigma z_\alpha}{\sqrt{n}} .$$

Then

$$a = \frac{\sigma z_\alpha}{\sqrt{n}} ,$$

or

$$n = \left(\frac{\sigma z_\alpha}{a} \right)^2 .$$

In this case, specifying the length of the confidence interval directly determines the sample size. If the sample is not a random sample, some dependence exists between pairs of observations, and the formula is more complicated (but accounted for in the computer program **star**).

B. Relative Precision

If the population mean is not roughly known, it is prudent to require the length of the confidence interval to be proportional to the mean. The half-length, specified in this way, is called the relative error. The percent of relative error, r, is then

$$r = \left(\frac{m_U - m_L}{2\bar{x}} 100\% \right).$$

The sample size is determined as it is for absolute precision. If a sample is from a normal population (or a non-normal population where $n \geq 30$) in which the population variance is known,

$$n = \left(\frac{\sigma z_\alpha}{r\bar{x}} 100 \right)^2 .$$

1.2.5 Hypotheses and Tests of Significance

This report discusses four recommended analyses of performance parameters. They are

- estimation of a performance parameter,
- tests to determine if a performance parameter is acceptable,
- tests to compare a performance parameter from two services, and
- tests to determine whether a variable condition affects a performance parameter (i.e., whether it is a factor).

Whereas the previous subsections discuss estimation, this one discusses the other three proposed analyses - all of which involve hypotheses and their tests (i.e., hypothesis tests).

Assume that a relationship exists between (among) the parameters of the distributions; this assumed relationship is called a hypothesis. The hypothesis to be tested (for acceptance or rejection) is called the null hypothesis. In our case, the parameter of the distribution is the mean for a time parameter or the proportion for a failure probability parameter.

Rejection of a hypothesis that is true and, therefore, should have been accepted is called a Type I Error. The maximum acceptable probability that a Type I Error will occur is called the level of significance and is denoted by α. Conversely, acceptance of a hypothesis that is false and, therefore, should have been rejected is called a Type II Error, and its probability is denoted by β.

The test of the null hypothesis requires

- defining a statistic,
- determining the density function of the statistic, and
- specifying the significance level of the significance test.

Figure 5 is a schematic diagram of the density function of a statistic; it also shows the percentage point, θ_0, corresponding to the significance level α. If the value of the statistic is to the left of θ_0, the hypothesis,

$$H_0 \colon \theta = \theta_0 ,$$

can be accepted at the 100α% level. For example, if $\alpha = 0.05$ and the hypothesis is true, there is only a 5% chance that the true value of the statistic is to the right of θ_0 (causing a Type I Error).

When the alternate hypothesis states

$$H_1 \colon \theta = \theta_1 > \theta_0 ,$$

the null hypothesis should be rejected only if the value of the statistic is greater than θ_0 (a one-tail test). Figure 5 is an example of this case. Similarly, when the alternate hypothesis states

$$H_1 \colon \theta = \theta_1 < \theta_0 ,$$

the null hypothesis should be rejected only if the value of the statistic is less than θ_0.

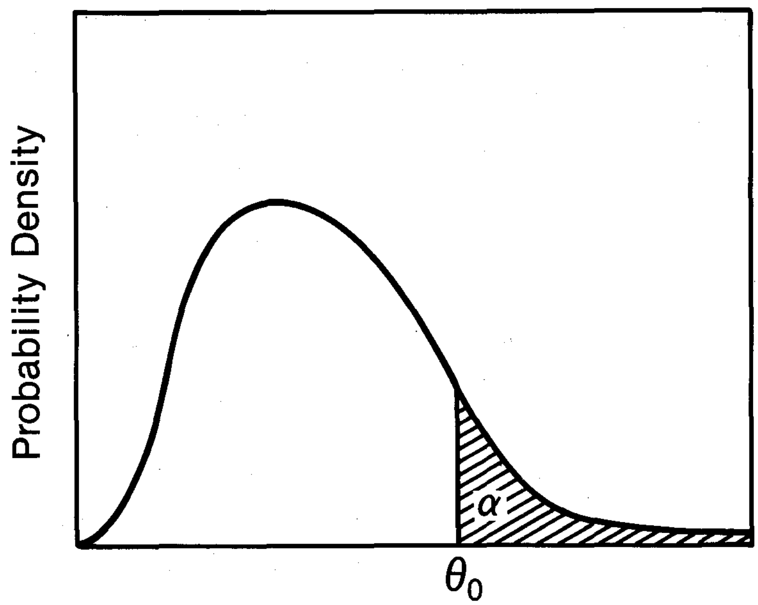


Figure 5. Schematic diagram of a distribution when the null hypothesis is true and the percentage point corresponding to the level of significance.

Usually when the alternate hypothesis states

$$H_0 : \theta = \theta_0 ,$$

a two-tail test is indicated because departure from equality in either direction is of interest. However, there are exceptions: When this alternate hypothesis is tested using the χ^2 statistic, the test is a one-tail test. The χ^2 statistic occurs often since it is the sum of the squares of normally distributed random variables.

2. DETERMINE THE OBJECTIVE OF THE EXPERIMENT AND SELECT ITS ANALYSIS

ANS X3.141 recommends one (or more) of four analyses of performance parameters. These analyses are

- estimation of a performance parameter,
- tests to determine if the value of a performance parameter is acceptable,
- tests to compare the value of a performance parameter from two services, and
- tests to determine whether a variable condition affects the value of a performance parameter (i.e., whether it is a factor).

The selection of these analyses depends upon the objective of the experiment. The two principals with a direct interest in the experiment are often the users and the vendors. Some common objectives for them are

- acceptance,
- characterization,
- design,
- maintenance,
- management,
- optimization, and
- selection.

The analyses are discussed in the following subsection, the objectives are discussed in Section 2.2, and the analyses that could be appropriate for each objective are discussed in Section 2.3.

2.1 Recommended Analyses

Although a great variety of analyses are possible, ANS X3.141 recommends one (or more) of four methods. These methods are discussed thoroughly in Volume 5, but they are introduced here to indicate the implications and requirements of the design.

2.1.1 Estimation Tests

A performance parameter and its confidence limits are estimated from either a single test (conducted under a single combination of levels of variable conditions) or pooled data from multiple tests (conducted under multiple combinations of levels of variable conditions). By virtue of the larger number of trials, pooled data might provide a better estimate.

2.1.2 Acceptance Test

The acceptance test determines whether the value of a performance parameter equals or exceeds an acceptable (threshold) value.[9] Because of sampling error, an interval of uncertainty exists about the acceptable value. The precision of the test is defined in terms of the width of this interval and the probability of making an incorrect decision when the parameter value is outside that interval. This test uses data from single tests.

2.1.3 Comparison Test

The null hypothesis states that the performance parameter means (or proportions) are equal. The comparison test is a test of the null hypothesis and determines whether the value of a performance parameter from one service is significantly different from another. This test is a special case of the following test.

2.1.4 Test to Determine if a Variable Condition is a Factor

Recall that a population is defined by a number of fixed conditions and the levels of a number of variable conditions. Suppose samples are obtained from more than one combination of levels of variable conditions. In this case, the null hypothesis states that the means (or proportions) of the multiple populations are equal. Depending upon the design of the experiment, an F statistic (for means) or a χ^2 statistic (for proportions) is defined to test this hypothesis. The statistic is compared with the appropriate distribution at a given percentage point. Depending on the comparisons, variable conditions are determined to be (or to not be) factors in the experiment.

---

[9]The value of the parameter that is referred to here is the population (true) value, not its estimate from a sample.

2.2 Common Experiment Objectives

Following is a brief discussion of seven common experiment objectives of performance evaluation for a data communication service.

2.2.1 Acceptance

An acceptance test is conducted to determine whether the (true) value of a performance parameter equals or exceeds an acceptable value. An experiment for acceptance could be required by the prospective user and, perhaps, by the vendor. It is performed "out of service" for a finite duration to determine whether long-term (true) values of a performance parameter lies between specified values. The experiment should be conducted only under the conditions and levels expected to be typical. That is, the user's facilities and activities are regarded as fixed. The values from an acceptance experiment should come from a single test conducted under a single combination of typical levels of variable conditions (Crow et al., 1960).

2.2.2 Characterization

A characterization experiment is conducted to estimate a performance parameter and its confidence limits. Results from a characterization test can be used by a vendor to advertise and price the system/service and to define guarantees and warranties. The data can come from a single test or from multiple tests. In the latter case, the data are pooled to obtain an "overall" estimate.

2.2.3 Maintenance

A maintenance experiment is usually conducted by the vendor under a single combination of levels of conditions - those typical of the user's facilities and activities. The need to repair can be determined by an acceptance test. The acceptance test tests the equality of the (true) mean and a specified value (perhaps a value guaranteed by the vendor as a result of characterization tests). There are three types of maintenance tests:

A. Trouble Response

These tests are conducted to verify the need for maintenance in response to a user's trouble report. They may be conducted when the system is either in or out of service.

B. **Repair Verification**

These tests are conducted to verify that a performance parameter is within prescribed limits after repair. They can be conducted when the system is either operational or not.

C. **Service Monitoring**

These tests are conducted to verify that a performance parameter is within prescribed limits when monitored. They are in-service tests and are conducted either continuously or periodically.

2.2.4 Design and Management

Design and management are objectives that could embody many activities. Depending upon these, the experiment to manage systems/services can probably be treated by one of the analyses mentioned above.

2.2.5 Optimization

Experiments for optimization can be conducted by either the vendor or the user. This experiment seeks to identify the effects of levels of variable conditions upon the performance parameter. For example, results of these experiments might indicate the block length that would optimize User Information Bit Transfer Rate or the effect of line speed on this performance parameter.

2.2.6 Selection

The selection experiment is usually conducted by the user to select between two systems/services, depending upon the value of one or more performance parameters. Tests in a selection experiment are conducted under a certain combination of levels of variable conditions for each system/service. The hypothesis of equal means (or proportions) of the parameters is tested.

2.3 Select the Analyses to Meet the Objectives

The objectives and analyses discussed in this section are not exhaustive; the objectives are common objectives, and the analyses are plausible analyses. Table 4 matches some common experiment objects and four plausible analyses. ("Factor" means the analysis determines if a variable condition is a factor). The "√" indicates that the analysis is suitable for that objective.

Select the analysis; it must be selected before the number of levels of the variable conditions can be selected. (Also see Section 4.2.)

Table 4. **Common Experiment Objectives and Plausible Analyses**

	PLAUSIBLE ANALYSES			
EXPERIMENT OBJECTIVE	**Estimation**	Acceptance	Comparison	Factor
Acceptance/Maintenance		√		
Characterization	√			
Design/Management	√	√	√	√
Optimization				√
Selection			√	

3. SELECT THE MOST IMPORTANT PERFORMANCE PARAMETERS

NTIA software supports two types of tests, depending upon the data communication function: access-disengagement tests and user information transfer tests. The experimenter should select the most important performance parameter for each type of test. Each test should then achieve the precision specified for that performance parameter. Other performance parameters will be estimated with a precision that is inversely proportional to the dispersion of their populations - a precision that may be greater or less than that specified for the most important performance parameter.

Testing for acceptance, characterization, and selection may require analysis of a large number of performance parameters. On the other hand, testing for maintenance, design, optimization, and management may require analysis of only a few.

The ANS X3.102 performance parameters are listed in Table 5 according to the data communication functions they define. In Table 5a, the parameters are arranged according to common performance criteria: speed, accuracy, and dependability. In Table 5b, they are arranged according to type of random variable. The latter criteria is necessary to determine the sample size required to achieve a specified precision and subsequent analysis of the sample.

Each user information transfer test produces at most one trial of User Information Bit Transfer Rate and User Fraction of Input/Output Time. Therefore, they should not be selected as the most important parameter (i.e., precision can not be obtained from a single value).

Table 6 lists estimates of means and their 95% confidence limits for the performance parameters that were measured from a public data network (labelled networks "A") by NTIA in 1984 (Spies et al., 1988). These estimates were obtained from pooled data from many tests over a variety of levels; they are shown simply to provide examples of the NTIA implementation. A detailed description of the results can be found in Spies et al. (1988). The most important performance parameters must be selected before selecting the levels of the variable conditions (see Section 4.2) and the precision (see Section 6).

Select the most important performance parameter.

Table 5. ANS X3.102 Performance Parameters

a. Organization by primary communication function and performance criterion

<table>
<tr><td colspan="3" rowspan="2"></td><td colspan="3">PERFORMANCE CRITERIA</td></tr>
<tr><td>SPEED</td><td>ACCURACY</td><td>DEPENDABILITY</td></tr>
<tr><td rowspan="7">PRIMARY COMMUNICATION FUNCTIONS</td><td colspan="2">ACCESS</td><td>• ACCESS TIME
• USER FRACTION OF ACCESS TIME</td><td>• INCORRECT ACCESS PROBABILITY</td><td>• ACCESS DENIAL PROBABILITY
• ACCESS OUTAGE PROBABILITY</td></tr>
<tr><td rowspan="4">USER INFORMATION TRANSFER</td><td>BIT TRANSFER</td><td>———</td><td>• BIT ERROR PROBABILITY
• BIT MISDELIVERY PROBABILITY
• EXTRA BIT PROBABILITY</td><td>• BIT LOSS PROBABILITY</td></tr>
<tr><td>BLOCK TRANSFER</td><td>• BLOCK TRANSFER TIME
• USER FRACTION OF BLOCK TRANSFER TIME</td><td>• BLOCK ERROR PROBABILITY
• BLOCK MISDELIVERY PROBABILITY
• EXTRA BLOCK PROBABILITY</td><td>• BLOCK LOSS PROBABILITY</td></tr>
<tr><td>TRANSFER SAMPLE TRANSFER: TRANSFER AVAILABLITY</td><td>———</td><td colspan="2">• TRANSFER DENIAL PROBABILITY</td></tr>
<tr><td>TRANSFER SAMPLE TRANSFER: THROUGHPUT</td><td>• USER INFORMATION BIT TRANSFER RATE
• USER FRACTION OF INPUT/OUTPUT TIME</td><td>———</td><td>———</td></tr>
<tr><td rowspan="2">DISENGAGEMENT</td><td>SOURCE DISENGAGEMENT</td><td>• SOURCE DISENGAGEMENT TIME
• USER FRACTION OF SOURCE DISENGAGEMENT TIME</td><td colspan="2">• SOURCE DISENGAGEMENT DENIAL PROBABILITY</td></tr>
<tr><td>DESTINATION DISENGAGEMENT</td><td>• DESTINATION DISENGAGEMENT TIME
• USER FRACTION OF DESTINATION DISENGAGEMENT TIME</td><td colspan="2">• DESTINATION DISENGAGEMENT DENIAL PROBABILITY</td></tr>
</table>

b. Organization by primary communication function and random variable

<table>
<tr><td colspan="3" rowspan="2"></td><td colspan="3">RANDOM VARIABLES</td></tr>
<tr><td>DELAY</td><td>RATE</td><td>FAILURE</td></tr>
<tr><td rowspan="7">PRIMARY COMMUNICATION FUNCTIONS</td><td colspan="2">ACCESS</td><td>• ACCESS TIME</td><td>• USER FRACTION OF ACCESS TIME</td><td>• INCORRECT ACCESS
• ACCESS OUTAGE
• ACCESS DENIAL</td></tr>
<tr><td rowspan="4">USER INFORMATION TRANSFER</td><td>BIT TRANSFER</td><td>———</td><td>———</td><td>• BIT ERROR
• BIT MISDELIVERY
• EXTRA BIT
• BIT LOSS</td></tr>
<tr><td>BLOCK TRANSFER</td><td>• BLOCK TRANSFER TIME</td><td>• USER FRACTION OF BLOCK TRANSFER TIME</td><td>• BLOCK ERROR
• BLOCK MISDELIVERY
• EXTRA BLOCK
• BLOCK LOSS</td></tr>
<tr><td>TRANSFER SAMPLE TRANSFER: TRANSFER AVAILABLITY</td><td>———</td><td>———</td><td>• TRANSFER DENIAL</td></tr>
<tr><td>TRANSFER SAMPLE TRANSFER: THROUGHPUT</td><td>———</td><td>• USER INFORMATION BIT TRANSFER RATE
• USER FRACTION OF INPUT/OUTPUT TIME</td><td>———</td></tr>
<tr><td rowspan="2">DISENGAGEMENT</td><td>SOURCE DISENGAGEMENT</td><td>• SOURCE DISENGAGEMENT TIME</td><td>• USER FRACTION OF SOURCE DISENGAGEMENT TIME</td><td>• SOURCE DISENGAGEMENT DENIAL</td></tr>
<tr><td>DESTINATION DISENGAGEMENT</td><td>• DESTINATION DISENGAGEMENT TIME</td><td>• USER FRACTION OF DESTINATION DISENGAGEMENT TIME</td><td>• DESTINATION DISENGAGEMENT DENIAL</td></tr>
</table>

Table 6. Estimates of Performance Parameters from a Public Data Network.

Performance Parameter Summary For PDN A Connections

Performance Parameter	*95% Lower Limit*	*Mean Estimate*	*95% Upper Limit*	*Pooling Disposition***	*Number of Cities*	*Number of Tests*	*Number of Trials*
Access Time (s)	40.7	41.8	42.9	2	3	11	209
User Fraction of Access Time	0.034	0.036	0.037	2	3	11	209
Incorrect Access Probability	0	0	0.062*	1	3	11	220
Access Outage Probability	0	0	0.062*	1	3	11	220
Access Denial Probability	0.018	0.050	0.107	1	3	11	220
Block Transfer Time (s)	3.61	3.79	3.97	2	2	7	559
User Fraction of Block Transfer Time	0.077	0.089	0.102	2	2	7	559
User Fraction of Input/Output Time	0.110	0.214	0.317	3	2	7	7
User Information Bit Transfer Rate (bps)	421	814	1207	3	2	7	7
Bit Error Probability	6.0×10^{-7}	7.0×10^{-6}	3.0×10^{-5}	1	2	7	573440
Bit Misdelivery Probability	-	-	-	-	-	-	-
Extra Bit Probability	0	0	3.0×10^{-5}*	1	2	7	573440
Bit Loss Probability	0	0	3.0×10^{-5}*	1	2	7	573440
Block Error Probability	0	2.0×10^{-3}	3.0×10^{-2}	1	2	7	560
Block Misdelivery Probability	-	-	-	-	-	-	-
Extra Block Probability	0	0	3.0×10^{-2}*	1	2	7	560
Block Loss Probability	0	0	3.0×10^{-2}*	1	2	7	560
Transfer Denial Probability	0	0	5.0×10^{-2}*	1	2	7	273
Source Disengagement Time (s)	14.3	15.1	15.8	2	3	11	194
User Fraction of Source Disengagement Time	0.058	0.061	0.065	2	3	11	194
Source Disengagement Denial Probability	0.042	0.072	0.116	1	3	11	209
Destination Disengagement Time (s)	4.9	5.2	5.4	1	3	11	207
User Fraction of Destination Disengagement Time	0.121	0.128	0.134	1	3	11	207
Destination Disengagement Denial Probability	0.008	0.018	0.033	2	3	11	209

* Conditional Probability Assumed to be 0.8

\- Not Measured

** 1 means no significant difference among tests or cities, so all trials pooled.
2 means no significant difference among cities, so all test means pooled.
3 means significant difference among cities, so no pooling; only 1 or 2 degrees of freedom for confidence limit.

4. DEFINE THE POPULATION OF EACH COMMUNICATION FUNCTION

A data communication session has four primary functions: Access, User Information Transfer, Source Disengagement, and Destination Disengagement. The population is the set of all possible attempts to complete each function (e.g., each access attempt is an element of the hypothetical population of access attempts).

The performance parameters of each function are the characteristics of the population. For example, the performance parameter, Access Denial Probability, is a characteristic of the hypothetical population of access attempts. Each characteristic has a distribution, and a goal of the experiment is to estimate the mean of each distribution with a specified precision and confidence level.

The distributions are defined by a set of existing conditions. Some conditions have levels that are quantitative (e.g., Block Size) and some have levels that are qualitative (e.g., Day of Week).

It is convenient to divide the set of conditions into those that are fixed and those that are variable - as far as the experiment is concerned. Fixed conditions are those for which there is either a single level with no options (such as a feature supplied by a vendor) or a single level that the experimenter will choose (such as a Block Size). Variable conditions have more than one level and, hence, provide an opportunity to design an experiment; the experiment is conducted over all or selected combinations of levels of the variable conditions.

4.1 Fixed Conditions

Most conditions of a data communication performance experiment are fixed. Even though some conditions can change, they probably will not - at least not often. The following fixed conditions apply to tests of any data communication function:

- Interfaces. Common interfaces are those between the application program and the system and between layers of network services.
- Measurement Points. Measurement points are located where performance-significant events occur.
- Reference Events. The events (reference events) that correspond to system-independent interface events must be specified. A diagram of the flow of signals across interfaces

(i.e., a session profile) should be constructed. (The session profile is discussed in Section 5 of Volume 3.)

- End Users. Source and destination end users are application programs provided by NTIA (as opposed to human operators). They perform end user activities and record the nature and times of interface signals in each computer.

4.1.1 Access-Disengagement Tests

The following fixed conditions apply to access-disengagement tests:

- Specified Performance Values. Performance values for some performance parameters must be specified in the file **spi.acd**. For example, if an observation of a delay parameter exceeds three times its specified value, the delay is classified as a failure. (Although this is the appropriate place to discuss the creation of **spi.acd**, the discussion is extensive enough to warrant a separate section, Section 5.)

- Number of Access Attempts. For access-disengagement tests, the number of access attempts is the sample size required to achieve a specified precision for the most important access-disengagement performance parameter (determined in Section 6 or 8). The maximum number of access attempts has arbitrarily been set at 40.

- Number of Blocks Transferred. For access-disengagement tests, the number of blocks transferred is assumed to be one.

- Block Size. The selected block size will probably be the minimum, 64 bytes, because the block size is irrelevant to access-disengagement tests.

4.1.2 User Information Transfer Tests

- Specified Performance Values. Performance values for some performance parameters must be specified in file **spi.xfr**. For example, if an observation of a delay parameter exceeds three times its specified value, the delay is classified as a failure. Also performance parameter values that indicate bit failures must be specified. These parameters are window sizes and bit shifts. (Although this is the appropriate place to discuss the creation of **spi.xfr**, the discussion is extensive enough to warrant a separate section, Section 5.)

- Number of Access Attempts. For user-information transfer tests, the number of access attempts is assumed to be one.

- Number of Blocks Transferred. For user information transfer tests, the number of blocks is the sample size required to achieve a specified precision for the most important block performance parameter (determined in Sections 6 or 8). However, the number of blocks may vary with block size and is limited to 320 since the minimum block size is 64 bytes and the maximum number of bytes that can be transferred is 20,480: (number of blocks) x (number of bytes/block) ≤ 20,480 bytes

- Software Flow Control. A test is conducted by executing a run command. Those run commands having a -f option cause software flow control (Section 11 of Volume 3).

4.2 Variable Conditions

Variable conditions have values or levels that will vary during the experiment. It is the variable conditions that influence the design of the experiment. A variable condition might be a factor for one performance parameter but not for another. Moreover, a performance parameter may be affected by more than one factor. When feasible, it is desirable to replicate tests under each combination of levels of variable conditions to possibly reveal unknown factors. Such a design is called a (full) factorial design.

Variable conditions can be divided into two types of conditions, primary and extraneous.

Primary variable conditions are those whose effects are of interest; they are thought to have a direct effect upon the value of a performance parameter. Primary variable conditions can be quantitative (as is the transmission rate) or qualitative (as is the error control that is provided or not provided). Moreover, qualitative variable conditions can be of two types: those having a few, known levels (called "fixed effects" or Type I variables) or those having randomly selected levels (called "random effects" or Type II variables).

Extraneous variable conditions are those whose effects are not of interest; they are considered to be "nuisance" variable conditions. Time of Day, and Day of Week are examples of (controllable) extraneous variable conditions.[10]

The following variable conditions can affect both types of tests:

- Source Site. Source site can be a city, building, office, computer, etc.

---

[10]Of course, time cannot affect an experiment, but events occurring during time may; listing Time of Day and Day of Week as conditions is a tacit admission that unknown and unknowable events may occur during time, and they could affect the experiment.

- Network. In the NTIA implementation, a network level exists whenever a level of an implicit variable condition changes. For example, line speed is an implicit variable condition, and whenever a different line speed is used, a different level of network exists. Other variable conditions, that may or may not affect the experiment, can define a different network; these conditions could include operating procedures.

- Time of Day. Time (period) of day is a variable condition whose levels are computed by a shell script from a computer clock. There are six periods, each having four hours. For example, the period from 0400 - 0800 is labelled as period 2.

- Day of Week. Day of week is a variable condition whose levels are computed by a shell script from calendar data.

- Destination Site. Destination site can be a city, building, office, computer, etc.

4.2.1 Access-Disengagement Tests

The following variable conditions may affect access-disengagement tests.

- Interaccess Delay. For access-disengagement tests, the trials are access attempts, and the intertrial delay could be positive to attenuate dependence between trials.[11]

- Two Optional Conditions. Two additional variable conditions are available for each type of test. For access-disengagement tests they are labelled O_7 and O_8 in this report.

For access-disengagement tests, select levels of Source Site, Network, Day of Week, Time of Day, Interaccess Delay, Destination Site, O_7, and O_8.

4.2.2 User Information Transfer Tests

The following variable conditions may affect user information transfer tests.

- Interblock Delay. For user information transfer tests, the trials are transferred blocks and the interblock delay may be

[11]Roughly speaking, the delay between consecutive sessions begins with the Disengagement Confirmation reference event and ends with the Access Request reference event of the following session.

zero (because it may be provided by the network), but could be positive to attenuate dependence between blocks.

- Block Size. Block size can be any number of bytes between 64 and 512. For user information transfer tests, the block size is constrained by the number of blocks transferred: The maximum number of bytes that can be transferred is 20,480: (number of blocks) x (number of bytes/block) ≤ 20,480 bytes.

- Two Optional Conditions. Two additional variable conditions are available for each type of test. For user information transfer tests they are labelled U_8 and U_9 in this report.

For user information transfer tests, select levels of Source Site, Network, Day of the Week, Time of Day, Interblock Delay, Block Size, Destination Site, U_8, and U_9 .

4.3 Enter Levels of the Fixed and Variable Conditions

Levels of some variable conditions will be entered as arguments for the shell script **runxt** when the test is conducted, and others will be entered in two files in the home directory: **default** and **netcodes**.

The **default** file exists in the home directory; it must be edited for each experiment. The **default** file contains seven lines of information. On the right of each is the full description. Figure 6 is an example **default** file.

Edit the first three characters of each applicable line of the **default** file.

The file **netcodes** contains

- the three-character code **bbb** for each Source Site and its name,

- the three-character code **ddd** for each Destination Site and its name, and

- the four-character code **aaaa** for each Network and its name.

```
May 30 11:47 1990  default Page 1

      bol    source site
      den    destination site
      512    block size (bytes or chars)
      ---    --xfr info-----
      30     number of blocks
      0      interblock delay (sec)
      ---    --ovh info-----
      25     number of accesses
      55     interaccess delay (sec)

ALL lines below this are comments:

NOTE:  Update appropriate preface file(s) to reflect source/dest. site
changes

Maximum sizes permitted as shown -

Maximum number of bytes or characters per test = 20480
Maximum number of blocks per test              =   320
Maximum number of accesses per test            =    40

Block sizes used     Maximum number of blocks or accesses:
        512                   40
        173                  118
        128                  160
         80                  256
         64                  320
```

Figure 6. Example of a default file.

Any number of sites and networks can be listed and in any order. The identification in this file will be used as legends on the graphs that are produced by the data display portion of the software. The name should be restricted to 14 characters because additional characters will be truncated by the data display software. Although this file has nothing to do with experiment design, it must be created before a test is processed (and it is probably more convenient to create it now than later). Table 7 is a list of the contents of the **netcodes** file.

Create the **netcodes** file.

Table 7. **netcodes** File

LINE	CONTENTS	CHARACTER FIELD	CONTENTS	CHARACTER FIELD
1	bbb	1-3	Source Site	5-19
...	...	...	...	...
...	ddd	1-3	Destination	5-19
...	...	...	...	...
...	aaaa	1-4	Network	5-19
...	...	...	...	...

4.3.1 Access-Disengagement Tests

A. Fixed Conditions

- Specified Performance Values. These values are entered in the **spi.acd** file. (See Section 5.)
- Number of Access Attempts. The number of access attempts is the sample size. There can be between 1 and 40 access attempts. It is entered in the **default** file.
- Number of Blocks Transferred. This number is assumed to be one; it is entered in the **default** file.
- Block Size. The selected block size will probably be the minimum, 64 bytes, because the block size is irrelevant to access-disengagement tests. It is entered in the **default** file.

B. Variable Conditions

- Source and Destination Site. Even though the source site and destination site may vary, they probably will remain fixed for a large part of the experiment and are the first two lines of the **default** file. Each of these variables is identified in the **default** file by three characters.
- Network. Network is represented by four characters in the command line of **runxt**.

- Interaccess Delay. The interaccess delay can be between 0 and 99 s; the longer the delay, the more the interaccess dependence will tend to be attenuated. It is entered in the **default** file.

- Two Optional Conditions. Enter the level of O_7 and the level of O_8 on the command line of **runxt**.

Table 8 is a list of fixed and variable conditions for experiment design of access-disengagement tests according to NTIA implementation. The code number associated with the variable conditions is the number that identifies variable conditions for analysis of multiple tests (Volume 5).

4.3.2 User Information Transfer Tests

A. Fixed Conditions

- Specified Performance Values. These values are entered in the **spi.xfr** file. (See Section 5.)

- Number of Access Attempts. This number is assumed to be one; it is entered in the **default** file.

- Number of Blocks Transferred. The maximum number of bytes is 20,480, and the minimum block size is 64 bytes. Hence, the maximum number of blocks is 320. The number of blocks is the sample size. It is entered in the **default** file.

B. Variable Conditions

- Source and Destination Site. Even though the source site and destination site will vary, they remain fixed for a large part of the experiment and are the first two lines of the **default** file. Each of these variables is identified in the **default file** by three characters.

- Network. Network is represented by four characters in the command line of **runxt**.

- Interblock Delay. The interblock delay should probably be set at 0 s since this is the delay that will probably exist during normal operation of the network. Enter the interblock delay in file **default**.

Table 8. Files and Shell Scripts Containing Levels of Conditions for Access-Disengagement Tests

a. Fixed Conditions

FIXED CONDITIONS	WHERE ENTERED
Specified Performance Values	**spi.acd**
Number of Access Attempts	**default**
Number of Blocks Transferred	**default**
Block Size (bytes)	**default**

b. Variable Conditions

VARIABLE CONDITIONS	CODE	WHERE ENTERED
Source Site	1	**default netcodes**
Network	2	**runxt netcodes**
Day of Week	3	(computed by **runxt**)
Time of Day	4	(computed by **runxt**)
Interaccess Delay	5	**default**
Destination Site	6	**default netcodes**
O_7	7	**runxt**
O_8	8	**runxt**

- <u>Block Size</u>. Select the block size. The smallest block size is 64 bytes and the maximum block size is 512 bytes. Enter the block size in file **default**.

- <u>Two Optional Conditions</u>. The level of U_8 and the level of U_9 are entered in the command line of **runxt**.

Table 9 is a list of fixed and variable conditions for experiment design of user information transfer tests according to NTIA implementation. The code number associated with the variable conditions is the number that identifies variable conditions for analysis of multiple tests (Volume 5). Unlike access-disengagement tests, user information transfer tests have Block Size as a variable condition.

Table 9. Files and Shell Scripts Containing Levels of Conditions for User Information Transfer Tests

a. Fixed Conditions

FIXED CONDITIONS	WHERE ENTERED
Specified Performance Values	spi.xfr
Number of Access Attempts	default
Number of Blocks Transferred	default

b. Variable Conditions

VARIABLE CONDITIONS	CODE	WHERE ENTERED
Source Site	1	default netcodes
Network	2	runxt netcodes
Day of Week	3	(computed by runxt)
Time of Day	4	(computed by runxt)
Interblock Delay	5	default
Destination Site	6	default netcodes
Block Size	7	default
U_8	8	runxt
U_9	9	runxt

5. SPECIFY PERFORMANCE VALUES

Performance values must be specified to declare failures (and to assign responsibility for them) and to identify outcomes of bits from user information transfer. A specified performance value can be a nominal value (as determined from a preliminary characterization test), a required value, or any other reasonable value.

A timeout is a factor of three times a specified primary performance parameter. It should be noted that a time limit (called **Timelimit**) is listed in program **converse** to halt the test if a response to a command does not occur or is not detected. To assure that this listed value does not preempt a normal attempt, it should be set larger than the maximum timeout of an access attempt, a block transfer attempt, a source disengagement attempt, and a destination disengagement attempt. (See Section 7 of Volume 3.)

> The specification files, **spi.acd** and **spi.xfr**, should be created while reading this section. They reside in **/usr/net/spi**. All specified times are to be entered as seconds.

5.1 Access-Disengagement Tests

Specified values for access-disengagement tests are listed in the **spi.acd** file. This file consists of five lines of ASCII characters. Table 10 shows the character fields, data edit descriptors, and contents of each line.[12]

The first line contains (left-justified) the words **SPECIFICATIONS INPUT.**

The second line contains an identifier that uniquely describes the test. Either leave it blank or enter a place holder in the first 60 columns; the identifier will be supplied by the identifier in the **SOI** file, and the four digit test number for columns 61-64 will be supplied by a shell script. The first 60

[12] FORTRAN defines three types of edit descriptors: data edit descriptors, control edit descriptors, and string edit descriptors. Data edit descriptors specify the conversion of values to and from the internal representation to the character representation in the formatted record of a file (e.g., A32, A64, **I4**, and **E16.0**). Control edit descriptors specify the spacing and position within a record, new records, interpretation of blanks, and plus sign suppression. String edit descriptors transfer strings of characters represented in format specifications to output records.

columns of the identifier must be identical to the batch identifier field in each overhead file for the test (as obtained from **preface.r** and **preface.x**).

The third line contains the three assessment option codes, either 1 or 0 depending upon whether performance parameters of that data communication function are to be analyzed or not, respectively.[13] For access-disengagement tests, the assessment option line is

1 0 1.

Table 10. **spi.acd** File Designation for Access-Disengagement Tests

LINE	CHARACTER FIELD	DATA EDIT DESCRIPTOR	CONTENTS
1	1-32	A32	File Descriptor: SPECIFICATIONS INPUT
2	1-64	A64	Batch Identifier
3	1- 4 5- 8 9-12	I4 I4 I4	Access Assessment Option: 1 U. I. Transfer Assessment Option: 0 Disengagement Assessment Option: 1
4	1-16 17-32	E16.0 E16.0	Access Time U. F. Access Time
5	1-16 17-32 33-48 49-64	E16.0 E16.0 E16.0 E16.0	Source Disengagement Time U. F. Source Disengagement Time Destination Disengagement Time U. F. Destination Disengagement Time

5.1.1 Access Function Specifications

The fourth line contains the specified values of the Access Time and the User Fraction of Access Time. Access outcomes are determined as in Figure 7. This scheme uses a factor of three times the specified Access Time (as the Access Timeout) and a factor of one times the User Fraction of Access Time.

5.1.2 Disengagement Function Specifications

The fifth line contains the specified values of both disengagement functions. The first half of the fifth line contains the specified values of the

[13]The reduction portion of the software permits these combinations of assessment options: (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1), (1,1,0), and (1,1,1). However, the NTIA software permits only (1,0,1) for access-disengagement tests and (0,1,0) for user information transfer tests.

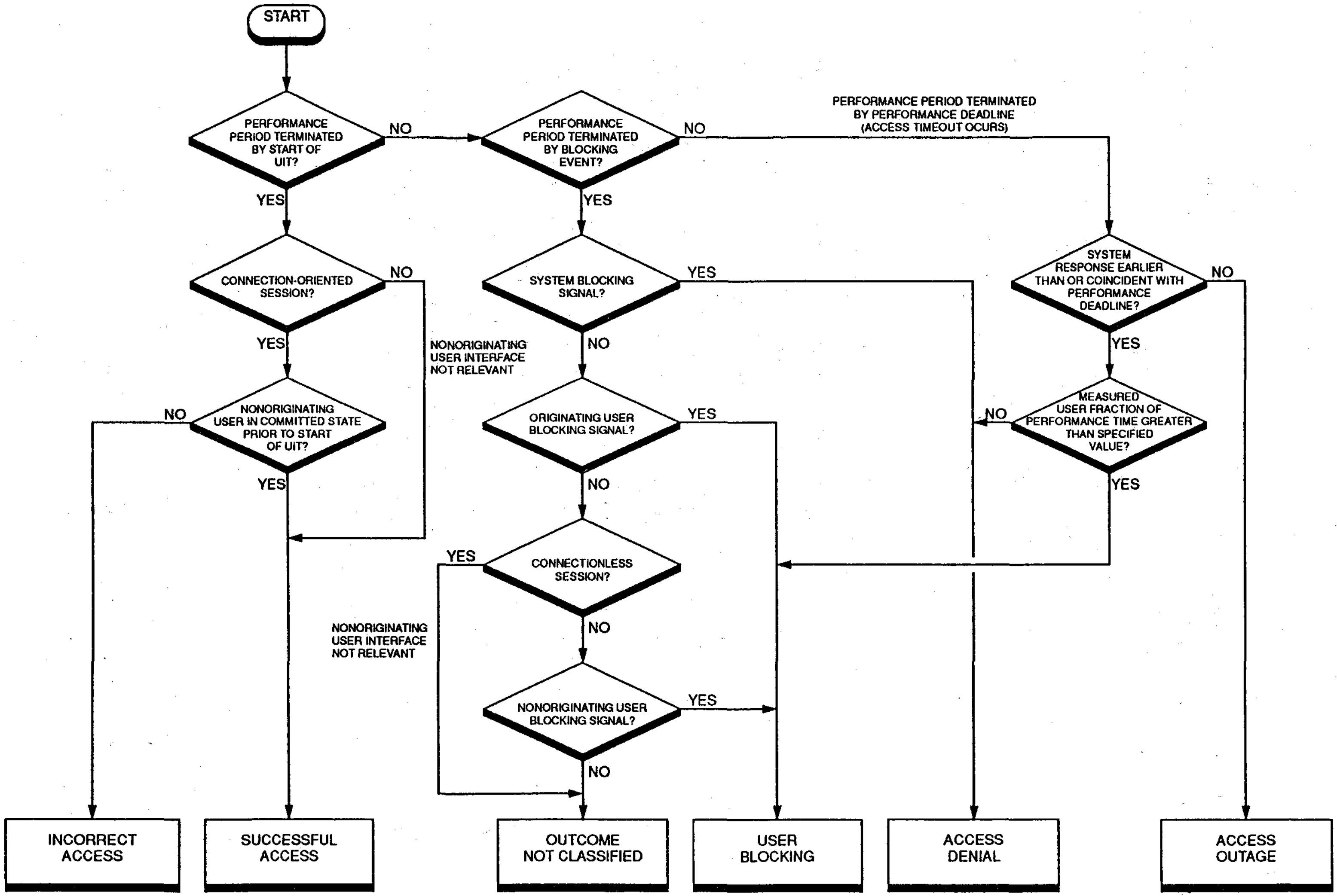


Figure 7. Scheme for determining access outcomes.

Source Disengagement Time and the User Fraction of Source Disengagement Time. Source disengagement outcomes are determined as in Figure 8.

The second half of the fifth line contains the specified values of the Destination Disengagement Time and the User Fraction of Destination Disengagement Time. This scheme uses a factor of three times the specified Source Disengagement Time (as the Source Disengagement Timeout) and a factor of one times the User Fraction of Source Disengagement Time. Destination disengagement outcomes are determined as in Figure 8. This scheme uses a factor of three times the specified Destination Disengagement Time (as the Destination Disengagement Timeout) and a factor of one times the User Fraction of Destination Disengagement Time. Figure 9 is a sample **spi.acd** file for an access-disengagement test.

5.2 User Information Transfer Tests

Specified values for user information transfer tests are listed in the **spi.xfr** file. This file consists of six lines of ASCII characters. Table 11 shows the character fields, data edit descriptors, and contents of each line.

The first line contains (left-justified) the words **SPECIFICATIONS FILE.**

The second line contains an identifier that uniquely describes the test. Either leave it blank or enter a place holder in the first 60 columns; the identifier will be supplied by the identifier in the **SUI** file, and the four digit test number for columns 61-64 will be supplied by a shell script. The first 60 columns of the identifier must be identical to the batch identifier field in each user information file for the test (as obtained from **preface.r** and **preface.x**).

The third line contains the three assessment option codes. For user information tests, the assessment option line is

0 1 0.

5.2.1 Block Transfer Function Specifications

The first half of the fourth line contains the specified values of the Block Transfer Time and the User Fraction of Block Transfer Time. Block transfer outcomes are classified as in Figure 10. This scheme uses a factor of three times the specified Block Transfer Time (as the Block Transfer Timeout), and a factor of one times the User Fraction of Block Transfer Time.

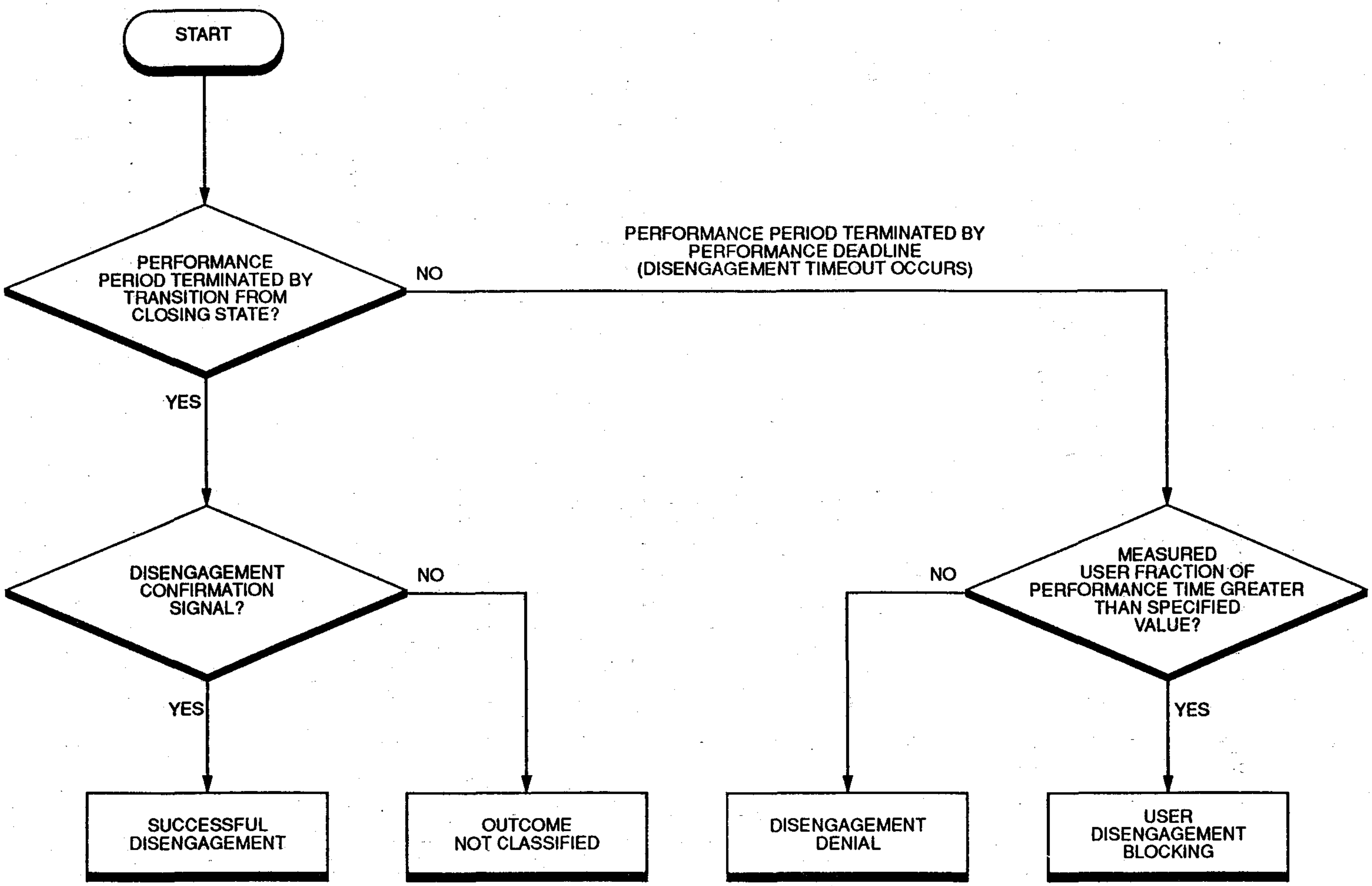


Figure 8. Scheme for determining disengagement outcomes.

```
SPECIFICATIONS INPUT
NTIA-ITS (Boulder)                                                 1352
   1   0   1
         4.5E+01             5.0E-02
         1.4E+01             7.5E-02            4.0E+00      2.0E-01
```

Figure 9. Example of specified performance values for an access-disengagement test.

Table 11. **spi.xfr** File for User Information Transfer Tests

LINE	CHARACTER FIELD	DATA EDIT DESCRIPTOR	CONTENTS
1	1-32	A32	File Descriptor: SPECIFICATIONS INPUT
2	1-64	A64	Batch Identifier
3	1-4 5-8 9-12	I4 I4 I4	Access Assessment Option: 0 U. I. Transfer Assessment Option: 1 Disengagement Assessment Option : 0
4	1-16 17-32 33-48 49-64	E16.0 E16.0 E16.0 E16.0	Block Transfer Time U. F. Block Transfer Time U. I. Bit Transfer Rate U. F. Input/Output Time
5	1-16 17-32 33-48 49-64	E16.0 E16.0 E16.0 E16.0	Bit Error Prob.for Transfer Availability Trial Bit Loss Prob. for Transfer Availability Trial Extra Bit Prob.for Transfer Availability Trial Transfer Sample Size
6	1-8 9-16 17-24 25-32	I8 F8.0 F8.0 F8.0	User Information Window Size (bits) Maximum Bit Shift in Incorrect Bit Identification Algorithm (bits) Maximum Bit Shift in Undelivered Bit Identification Algorithm (bits) Maximum Shift in Extra Bit Identification Algorithm (bits)

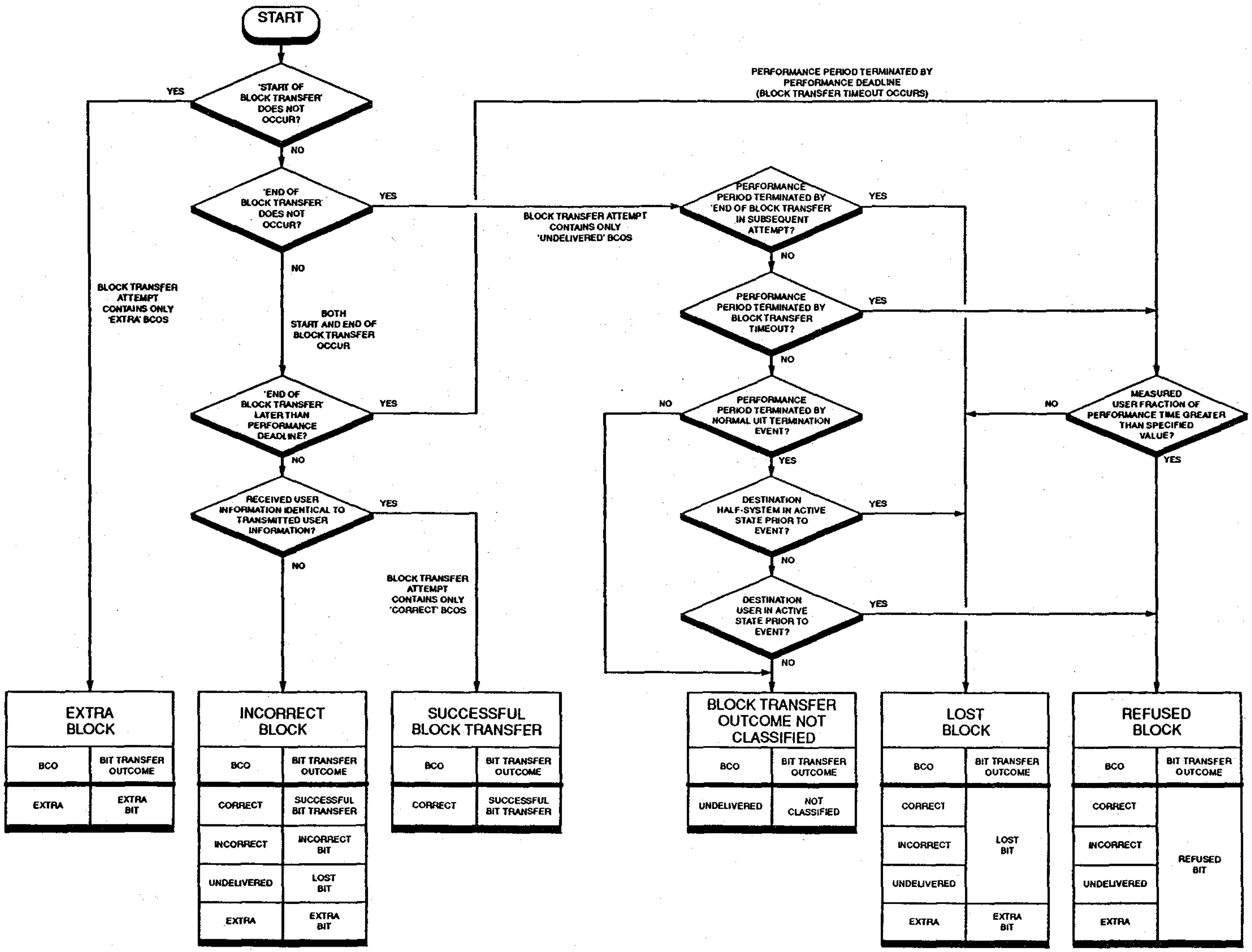


Figure 10. Scheme for determining block transfer and bit transfer outcomes.

5.2.2 Transfer Sample Specifications

All transfer sample specifications are used to evaluate one performance parameter, Transfer Denial Probability.

The last fourth of the fifth line contains the specified minimum number of user information bits for a transfer sample. This number is typically several thousand bits, but it must include at least one block.

A transfer availability trial is the smallest transfer sample. The outcome of a transfer availability trial is evaluated using four supported performance parameters. The supported parameters are

- User Information Bit Transfer Rate,
- Bit Error Probability,
- Bit Loss Probability, and
- Extra Bit Probability.

The threshold values of the three supported probability performance parameters are the fourth roots of their specified values, and the threshold value of the User Information Bit Transfer Rate is one third of its specified value.[14] Table 12 lists the supported performance parameters, the variable indicating their specified performance values, and the corresponding threshold values.

The second half of the fourth line contains the specified values of the User Information Bit Transfer Rate and the User Fraction of Input/Output Time. Even though these two performance parameters will be examined for each transfer availability trial (to evaluate Transfer Denial Probability), the specified values should be long-term values.

The first three-fourths of the fifth line contains the specified Bit Error Probability, Bit Loss Probability, and Extra Bit Probability.

Transfer availability trial outcomes are determined as in Figure 11. This scheme uses threshold values of the four supported performance parameters and a factor of one times the specified value of the User Fraction of Input/Output Time. The outcome of a transfer availability trial is Successful Transfer if the estimates of all <u>four</u> supported parameters are "better" than their threshold

[14]The revised ANS X3.102 replaces fourth root with square root.

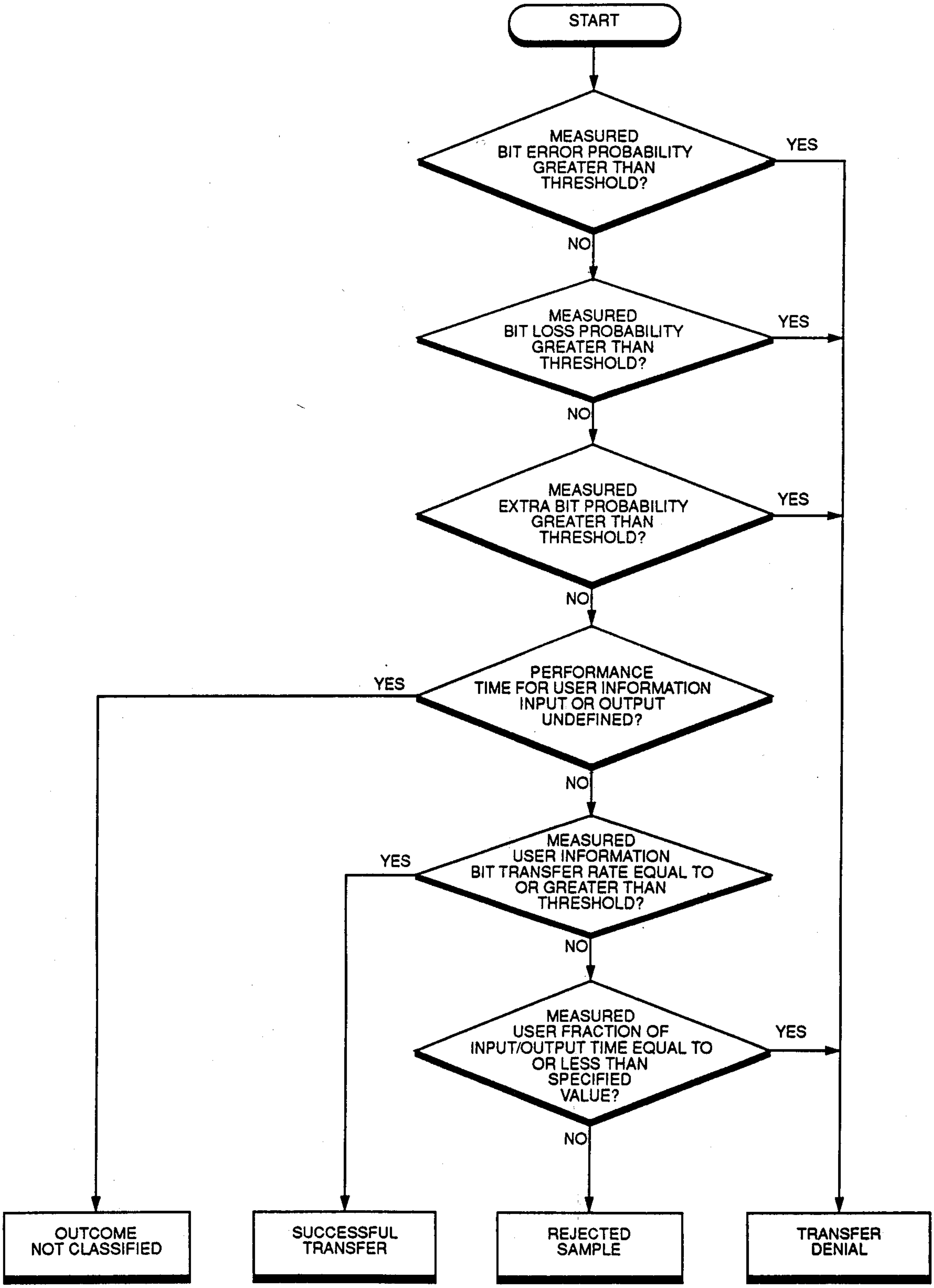


Figure 11. Scheme for determining transfer availibility trial outcomes.

values. Otherwise, the outcome depends upon the User Fraction of Input/Output Time: the outcome is Transfer Denial if the system is responsible and Rejected Sample if the user is responsible.[15]

Table 12. Specified Performance Values and Threshold Values of Supported Performance Parameter

SUPPORTED PERFORMANCE PARAMETERS	SPECIFIED VALUE	THRESHOLD VALUE
Bit Error Probability	P_e	$P_e^{1/4}$
Bit Loss Probability	P_ℓ	$P_\ell^{1/4}$
Extra Bit Probability	P_x	$P_x^{1/4}$
User Information Bit Transfer Rate	r	r/3

5.2.3 Bit Transfer Function Specifications

Since user information bit transfer failures either occur or do not occur, we need not specify performance values to declare them as failures. However, we must specify values to identify them if they occur. Before discussing these specified values, we will discuss the outcomes of bit transfer attempts, called bit comparison outcomes (BCOs), and assumptions about their occurrence.

A. Bit Transfer Attempts and BCOs

A bit transfer attempt is associated with

- each pair of corresponding source and destination bits,
- each undelivered source bit, and
- each extra destination bit.

Each bit transfer attempt results in one of the following four BCOs:

- Correct BCO. This is a bit transfer attempt associated with a corresponding pair of source and destination bits having the same binary value.
- Incorrect BCO. This is a bit transfer attempt associated with a corresponding pair of source and destination bits having a different binary value.

---

[15]The user is considered to be responsible if the Input/Output Time exceeds the specified value of User Fraction of Input/Output Time.

- Undelivered BCO. This is a bit transfer attempt associated with an undelivered source bit.

- Extra BCO. This is a bit transfer attempt associated with an extra destination bit.

B. Assumptions about the Occurrence of BCOs

The following assumptions are made about the occurrence of BCOs:

- Model. The occurrence of BCOs is described by a probability model.

- Failures. The probability is very small that any of the three failure BCOs will occur.

- Sequence. A sequence of each type of BCO either is a string or a cluster:

  - String. A string is a sequence of bits in which all bits in the sequence have the same BCO. It is assumed that a string consists of correct BCOs, undelivered BCOs, or extra BCOs because, for all three BCOs, the cause of the failures (or absence of causes in the case of correct BCOs) probably persists.

  - Cluster. A cluster is a sequence of bits in which a fraction of bits in the sequence have a differing BCO (i.e., a fraction that is probably greater than the failure probability that is expected over the experiment).[16] It is assumed that a cluster consists of incorrect BCOs (the only BCO defined by differing source and destination bit values) because the cause of the failures is probably more transitory than that causing (or failing to cause) the other three BCOs.

---

[16]Clustering of bits tends to occur when the conditional probability of a bit failure, given a failure in the previous bit, exceeds the probability of a bit failure (over the experiment). In the NTIA implementation, dependence is measured between consecutive bits by the autocorrelation of lag 1.

C. Identification of BCOs

The procedure for identifying BCOs is called bit correlation. This procedure is discussed briefly here and fully in Section 4.2.1.a of Volume 4.

If no bit failures are identified, source words are compared with destination words (using an algorithm called the fast correlation algorithm).

If a bit failure is identified, source bits are compared with destination bits (using an algorithm called the basic correlation algorithm). A sequence of source bits and a sequence of destination bits are stored in arrays (of equal size) called the source user information window and the destination user information window, respectively. Since the occurrence of a bit failure causes the sequences to be stored, the initial bits in each sequence do not match. The two sequences are shifted past these two windows in unison, and all bits are compared until

- all bit in each window match,
- a specified number of bits has been shifted, or
- no bits remain to be compared.

Identification depends upon the specified size of the windows and the specified size of the shifts.

1. User Information Window Size

The two windows must be the same size, and they are set at 16 bits.

> Enter the user information window size in first one-fourth of the sixth line of the **spi.xfr** file. If the window size is changed from 16 bits, also change the array size in two variables: IDIWIN and **ISIWIN**. They are located in common blocks **DUIWIN** and **SUIWIN**, respectively. **DUIWIN** is contained in subroutines **BITCOR**, **DWLOAD**, **DSHIFT**, and **COMPAR**. **SUIWIN** is contained in subroutines **BITCOR**, **SWLOAD**, **SSHIFT**, and **COMPAR**. If the window size exceeds 32, a dimension statement in **BITCOR** must be increased.

2. Maximum Bit Shift

Selection of the maximum bit shift depends upon the speed and the degree of certainty with which the experimenter wishes to identify failures. As is usual, one is obtained at the sacrifice of the other:

- Speed. The basic correlation algorithm is slow if the maximum bit shift is large.
- Accuracy. The basic correlation algorithm identifies a fraction of failures that roughly depends upon the size of the maximum bit shift (unless the string or cluster exceeds it, in which case no failures are identified):
  - To identify no failures, set the maximum bit shift equal to zero.
  - To identify probably all failures, set the maximum bit shift equal to the number of bits in several blocks,
  - To identify all failures, set the maximum bit shift equal to the number of bits transferred (or at least equal to the largest failure string or cluster - but this number is unknowable).

Specify the maximum bit shift for each of the three failure BCOs in the last three one-fourth portions of the sixth line of the **spi.xfr** file.

- For incorrect BCOs, this is arbitrarily set at 256.
- For undelivered BCOs and extra BCOs, this could be set at the number of bits in several blocks "to detect probably all failures."

Figure 12 is an example of the specifications file for a user information transfer test (i.e., the second entry in the third line is 1).

```
SPECIFICATIONS INPUT
NTIA-ITS (BOULDER)                                                      1352
   0   1   0
         3.0E+00              5.0E-01                1.0E+03      5.0E-01
         1.0E-08              1.0E-08                1.0E-08      30000.
      16      256.     8192.     8192.
```

Figure 12. Example of specified performance values for a user information transfer test.

6. CONDUCT A PRELIMINARY CHARACTERIZATION TEST AND DETERMINE THE TEST DURATION

Before testing, the performance parameters are probably known only within an order of magnitude or so. The experiment can be conducted with this limited knowledge; however, it is strongly recommended that a preliminary "typical" test be conducted to roughly characterize the system or service. The knowledge obtained will allow the experiment to be designed more efficiently:

- Sample Size. The (minimum) sample size required to achieve a specified precision can be determined more accurately because the dispersion and dependence of trial values will have been estimated.[17]

- Test Duration. The time required to conduct a test having the specified precision can be determined. Hence, the number of tests and the selected levels of variable conditions can be allocated judiciously. This allocation is the subject of Section 7.

- Meaningful Values. Meaningful specified performance values can be assigned, particularly for the four supported performance parameters required for Transfer Denial Probability (Section 5).

Figure 13 is a structured design diagram depicting the preliminary test and the determination of test durations for five selected precisions.

Refer to Volume 3 (Data Extraction) and conduct a preliminary test.

The last section of Volume 3 (Section 12) shows how to process regular tests (i.e., tests other than a preliminary characterization test); the remainder of this section is a version of that Section 12, but modified for the preliminary characterization test.

---

[17]If the number of failures is zero or one, the upper confidence limit can be determined only if the number of trials exceeds a certain size (a size that depends upon the conditional probability of a failure given that a failure occurred in the previous trial); this minimum number of trials is listed in Table 3 of Volume 5.

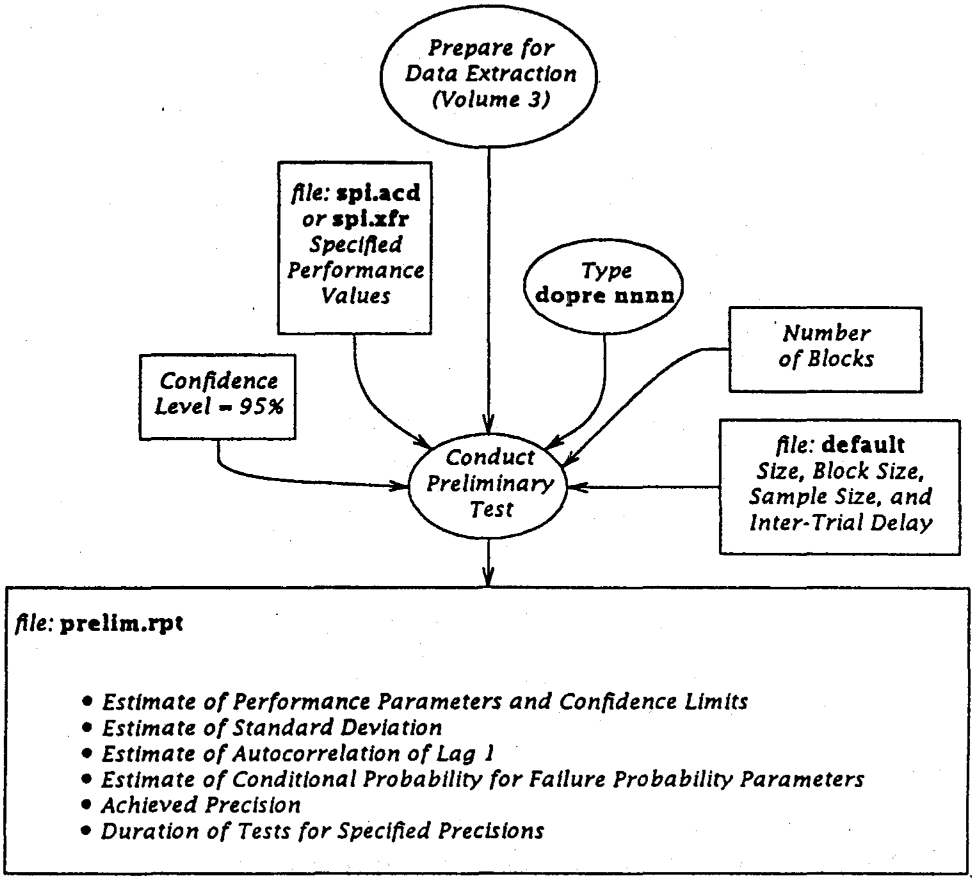


Figure 13. Structured design diagram of the preliminary characterization test and resultant test durations.

6.1 Access-Disengagement Tests

6.1.1 Process the Preliminary Characterization Test

The test can now be processed by the shell script **dopre**.

Type

dopre nnnn

where **nnnn** is the number assigned to the preliminary test.

The test results are contained in **usr/data/nnnn.pre**. Figure 14 is the first part of this file. For each performance parameter, it lists the number of trials and estimates of the mean, standard deviation, and autocorrelation of lag 1.[18] It also lists the precision achieved at the 95% confidence level - which depends upon all listed estimated values except that of the mean.

6.1.2 Select the Precision

The standard deviation and autocorrelation of lag 1 are used to determine the sample size required to obtain a specified absolute precision for time parameters. Similarly, the probability of a failure given that a failure occurred in the previous trial is used to determine the number of failures required to obtain a specified relative precision for failure probability parameters.[19]

Using the values obtained from the preliminary characterization test, the required sample size and the time required to complete a test of this size are estimated for all access-disengagement performance parameters for various precisions at the 95% confidence level. This information is the last part of file **nnnn.pre** and is listed in Figure 15.

---

[18]If nominal values are used as specified values (in the previous section), these estimates of the time parameters can replace them.

[19]Program star asks for the maximum of these three values. This value would provide a conservatively large estimate of the sample size. However, since a preliminary test has been conducted, the estimates of these values should be used - not their maximum.

PRELIMINARY CHARACTERIZATION TEST

Access-Disengagement Parameters

Test Number 2115

Confidence Level = 95%
Time between Access Attempts = 55 s

Time Parameters

Performance Parameter	Number of Trials	Estimate of Mean	Estimate of St. Dev.	Estimate of Autocorr. of Lag 1	Achieved Absolute Precision
Access Time	20	4.268e+01	1.344e+00	-1.920e-01	5.180e-01
U. F. of Access time	20	3.084e-03	5.232e-04	9.055e-04	1.025e-03
Source Diseng. Time	20	1.886e+00	2.644e-01	-6.494e-02	1.169e-01
U. F. of S. Diseng. Time	20	4.139e-02	1.232e-03	-8.952e-02	2.414e-03
Destination Diseng. Time	20	3.810e-01	1.821e-02	-1.359e-02	8.410e-03
U. F. of D. Diseng. Time	20	1.605e-01	1.745e-03	3.267e-02	3.420e-03

Failure Probability Parameters

Performance Parameter	Number of Failures	Estimate of Proportion	Estimate of Autocorr. of Lag 1	Achieved Relative Precision
Incorrect Access	0	0.000e+00	0.000e+00	indet.
Access Outage	0	0.000e+00	0.000e+00	indet.
Access Denial	0	0.000e+00	0.000e+00	indet.
Source Diseng. Denial	0	0.000e+00	0.000e+00	indet.
Dest. Diseng. Denial	0	0.000e+00	0.000e+00	indet.

Figure 14. Results of an access-disengagement preliminary characterization test.

REQUIRED SAMPLE SIZE AND TEST DURATION

Access-Disengagement Parameters

Confidence Level = 95%
Time between Access Attempts = 55 s

Access Time

Specified Absolute Precision	Req'd no. of delays	Test Duration
1.036e-01	439	12.54h
2.072e-01	110	3.14h
3.108e-01	49	83.98m
4.144e-01	28	47.99m
5.180e-01	18	30.85m

Source Disengagement Time

Specified Absolute Precision	Req'd no. of delays	Test Duration
2.337e-02	440	12.57h
4.674e-02	110	3.14h
7.011e-02	50	85.70m
9.348e-02	29	49.71m
1.169e-01	18	30.85m

Destination Disengagement Time

Specified Absolute Precision	Req'd no. of delays	Test Duration
1.682e-03	439	12.54h
3.364e-03	110	3.14h
5.046e-03	50	85.70m
6.728e-03	29	49.71m
8.410e-03	19	32.57m

Figure 15. (Part 1). **Specified precisions and test durations for access-disengagement performance parameters.**

Incorrect Access

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Access Outage

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Access Denial

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Source Disengagement Denial

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Destination Disengagement Denial

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Figure 15. (Part 2). **Specified precisions and test durations for access-disengagement performance parameters.**

6.2 User Information Transfer Tests

6.2.1 Process the Preliminary Characterization Test

The test can now be processed by the shell script **dopre**.

Type,

dopre nnnn

where **nnnn** is the number assigned to the preliminary test.

The results are contained in **usr/data/2p/nnnn.pre**. Figure 16 is the first part of this file. For each user information transfer performance parameter, it lists the number of trials (i.e., blocks) and estimates of the mean, standard deviation, and autocorrelation of lag 1. It also lists the precision achieved at the 95% confidence level.

This information is used (as described in Section 5) to estimate the performance parameters (called supported performance parameters) that are required to estimate Transfer Denial Probability. This performance parameter is estimated in Figure 16.[20]

6.2.2 Select the Precision

The estimates of the standard deviation and the autocorrelation of lag 1 are used to determine the sample size required to obtain a specified absolute precision for time parameters (from program **star**). Similarly the probability of a failure, given that a failure occurred in the previous trial, is used to determine the number of failures and to obtain a specified relative precision for failure probability parameters. Test durations are listed in Figure 17.

User Information Bit Transfer Rate and User Fraction of Input/Output Time have one trial per test. Hence, there is no absolute precision. Unfortunately, to characterize these performance parameters, several preliminary tests are required. If it is decided that these performance parameters should be characterized, use the analysis described in Section 1.2.2 of Volume 5. Specifically, use the **rate** shell script described in part B of that section. It

[20]If nominal values are used as specified values (in the previous section), these values of time parameters and bit failure probabilities can replace those values.

PRELIMINARY CHARACTERIZATION TEST

User Information Transfer Parameters

Test Number 2187

Confidence Level = 95%
Time between Block Transfer Attempts = 0 s

Specified Performance Values
Bit Error: 1.000e-08
Bit Loss: 1.000e-08
Extra Bit: 1.000e-08

Time Parameters

Performance Parameter	Number of Trials	Estimate of: Mean	Estimate of: St. Dev.	Estimate of: Autocorr. of Lag 1	Achieved Absolute Precision
Block Transfer Time	33	2.099e+00	9.303e-01	7.524e-01	8.778e-01
U. F. of Blk. Trans. Time	33	2.021e-02	1.267e-03	1.498e-01	2.483e-03
U. I. Bit Transfer Rate	1	5.981e+03	N/A	N/A	N/A
U. F. Input/Output Time	1	6.955e-02	N/A	N/A	N/A

Failure Probability Parameters

Performance Parameter	Number of Failures	Estimate of: Proportion	Estimate of: Autocorr. of Lag 1	Achieved Relative Precision
Bit Error	0	0.000e+00	0.000e+00	indet.
Extra Bit	0	0.000e+00	0.000e+00	indet.
Bit Loss	24975	1.524e-01	1.000e+00	218%
Block Error	2	5.714e-02	-6.061e-02	156%
Extra Block	0	0.000e+00	0.000e+00	indet.
Block Loss	5	1.250e-01	7.949e-01	237%
Transfer Denial	2	5.000e-01	3.333e-01	91%

Figure 16. Results of a user information transfer preliminary characterization test.

REQUIRED SAMPLE SIZE AND TEST DURATION

User Information Transfer Parameters

Confidence Level = 95%
Time between Block Transfer Attempts = 0 s

Block Transfer Time

Specified Absolute Precision	Req'd no. of delays	Test Duration
1.756e-01	765	14.01m
3.511e-01	145	3.52m
5.267e-01	85	1.56m
7.022e-01	50	54.96s
8.778e-01	36	39.57s

Bit Error

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

^2

Extra Bit

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Bit Loss

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	6449683	42309650	6.31h
50%	1067797	7004703	62.66m
100%	297941	1954480	17.48m
200%	102912	675098	6.04m
500%	42875	281258	2.52m

Figure 17. (Part 1). Specified precisions and test durations for access-disengagement performance parameters.

Block Error

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	97	1697	31.09m
50%	17	297	5.44m
100%	5	87	1.59m
200%	2	34	37.37s
500%	1	17	18.69s

Extra Block

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	871	unknown	unknown
50%	145	unknown	unknown
100%	41	unknown	unknown
200%	14	unknown	unknown
500%	6	unknown	unknown

Block Loss

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	981	7848	2.40h
50%	163	1304	23.89m
100%	46	368	6.74m
200%	16	128	2.34m
500%	7	56	61.56s

Transfer Denial

Transfer sample is 8 blocks

Specified Relative Precision	Req'd No. of failures	Req'd Number of Trials	Test Duration
20%	484	968	2.36h
50%	81	162	23.74m
100%	23	46	6.74m
200%	8	16	2.34m
500%	4	8	70.35s

Figure 17. (Part 2). **Specified precisions and test durations for access-disengagement performance parameters.**

will provide an estimate of each performance parameter and its confidence limits. The achieved absolute precision can easily be determined as one half of the difference between the upper and lower confidence limits (Section 1.2.4).

Using the values obtained from the preliminary test, the sample size and the time required to complete a test of this size are estimated for all user information transfer performance parameters (except for User Information Bit Transfer Rate and User Fraction of Input/Output Time); they are determined for various precisions at the 95% confidence level.[21]

Select either an absolute or a relative precision, depending upon whether the most important parameter is a time or a failure probability parameter, respectively. Enter the corresponding number of blocks and block size (i.e., sample size) and the interblock delay in file **default**.

The duration of tests for User Information Bit Transfer Rate and User Fraction of Input/Output Time can be determined in the following way, using User Information Bit Transfer Rate as the criterion (because it is the primary performance parameter):

- Sample Size. Determine the sample size (i.e., number of tests) required to achieve the specified precision by entering the following values in the interactive program, **star** (Section 1.1.2 of Volume 5).

- Absolute Precision. Specify an absolute precision.

- Sample Standard Deviation. Estimate the standard deviation from the sample (i.e., from the trials from each test).

- Independent Trials. Assume the autocorrelation of lag 1 to be zero; this is probably true since each trial comes from a different test.

- Duration of the Tests. The duration of the tests is the product of the sample size and the duration of an average test. (The beginning and ending times for each test are listed in file **pretimes.xfr**.)

[21]Estimates of all bit and block failure probabilities are listed in this table, although it is likely that some values are too small to be estimated accurately by a sample of this size.

7. SELECT THE DESIGN

The test durations required to achieve a set of precisions can be determined by the preliminary characterization test, and some acceptable designs are discussed in this section. Then, the experimenter must select a precision and a design which, combined, meet time/budget requirements - and achieve the selected objective.

Two or more primary variable conditions may interact (i.e., they may be dependent). Therefore, it is desirable to observe a performance parameter for all combinations of levels of primary variable conditions.

Extraneous variable conditions should be allowed to assume their natural levels so that the conclusions will be representative.

It is essential that levels of primary and extraneous variable conditions not vary together (i.e., their levels should not be correlated). Their effects should be neutralized by applying general principals such as

- randomization which tends to remove bias,
- replication to provide precision,
- blocking which tends to reduce variability, and
- balance which tends to provide equal precision.

When levels cannot be controlled, the order of testing should be randomized using random permutation tables (Moses and Oakford, 1963). These tables and some of the following designs have been applied to the design of experiments to quantify the performance of satellite networks (Cass and Miles, 1990).

Following is a brief description of four commonly used designs:

- Factorial Design. A (fully) factorial design requires that a test be conducted at every combination of levels of the primary variable conditions. For example, if four primary variable conditions have 4, 3, 3, and 2 levels, there will be 4 x 3 x 3 x 2 = 72 tests. Factorial design can be considered if there is more than one primary variable condition.

- Fractional Factorial Design. Fractional factorial designs are factorial designs having some missing combinations of levels of primary variable conditions. This design might be considered if it is impractical to implement the more desirable factorial design. The missing combinations of levels should be selected judiciously. Selection might be easier if some pairs of primary variable conditions are thought to be independent. In this case, the value at different combinations of levels should be random. Fractional factorial designs should be favored over factorial designs for preliminary experiments because they are easier to implement and the results of such experiments cannot be easily predicted (i.e., redesign is likely).

- Blocking Designs. It is important that levels of primary and extraneous variable conditions not vary together (i.e., as if they are correlated). For example, it is not wise to test one level of a condition always on Monday and another level always on Wednesday. In this case, the effects of Day of the Week and that condition could not be separated: They would be confounded. To separate the effects of primary and extraneous variable conditions, their levels can be combined either randomly or systematically. For each level of an extraneous variable condition (called a block in the blocking design), a test is conducted at the combinations of levels of the primary variable condition. Moreover, it is desirable to achieve balance by including all combinations of levels of each primary variable condition the same number of times within each block. If they are selected randomly, the design is called a randomized block. If they are selected systematically, the design might be an incomplete block design, a Latin square, a Greco-Latin square, or others. At each level of an extraneous variable condition, obtain an observation for all combinations of levels of the primary variable conditions. This allows the effect of the extraneous variable conditions to be estimated and removed. Then, the decision concerning statistical significance of the primary variable condition is more accurate.

- Response Surface Design. If levels of the primary variable conditions are quantitative, orthogonal central composite designs can be used, and a regression surface can be fitted.

Select a design and the combination of levels of the variable conditions for each type of test.

8. DETERMINE THE SAMPLE SIZE

For most objectives, the sample size required to achieve a specified precision is determined from a preliminary characterization test (Section 6). The shell script **dopre** passes data from a preliminary test to program **star**. This shell script determines the required sample size for most performance parameters and a selected set of precisions at the 95% confidence level.[22]

The following section shows how to determine the sample size for the experiment objectives of acceptance and maintenance and for tests for which a preliminary characterization test is not available. Section 8.2 shows how to determine the sample size for other experiment objectives.

8.1 Determine the Sample Size for Acceptance and Maintenance Objectives

A hypothesis test, called an acceptance test, can determine whether the mean of a performance parameter equals a threshold (acceptable) value. Hence, an acceptance test is appropriate to determine the sample size for two experiment objectives:

- Acceptance. An acceptance test can determine if the mean of a performance parameter is acceptable.
- Maintenance. An acceptance test can determine if the system requires maintenance to return the mean of a performance parameter to an acceptable value.

The methods described in Section 1.2.5 are used in acceptance testing:

- Threshold. The threshold (acceptable) value is specified. This is a value that can be accepted with indifference.
- Interval about Threshold. Because a sample has a finite number of trials, an interval of uncertainty exists about the threshold level. This interval is defined by two values, one that is considered to be totally satisfactory and one that is considered to be totally unsatisfactory. The narrower the interval of uncertainty, the greater the precision.

---

[22]Since each test of User Information Bit Transfer Rate and User Fraction of Input/Output Time results in a single trial, the sample size need not be determined for these two parameters.

In the case of failure probability parameters, if a test results in zero or one failure, the sample must be a certain size in order to avoid computer arithmetic fault in the algorithm to estimate the upper confidence limit. See Table 3 of Volume 5.

- <u>Null Hypothesis</u>. The null hypothesis states that the population value of the performance parameter is equal to the totally satisfactory value.[23] Because we are interested in whether the performance parameter value is better than the totally satisfactory value, this hypothesis is tested by a one-sided test.

Acceptance tests involve two precision objectives:

- <u>Incorrect Rejection</u>. The probability of incorrectly rejecting a performance value which is totally satisfactory is to be α = 5% or less (a probability called the significance level). This type of error is called a Type I error. The 5% significance level is traditionally used, but it could be, say, 1% if the loss incurred from committing this error would be large. Although it is the subject of analysis (Volume 5), it seems appropriate to state here, that the null hypothesis would be accepted at the α significance level if all or part of the $100(1 - 2\alpha)$% confidence interval of the parameter estimate lies in the totally satisfactory region, and rejected otherwise. Since NTIA analysis uses 90% or 95% confidence limits, α should be 5% or 2.5%, respectively.[24]

- <u>Incorrect Acceptance</u>. The probability of incorrectly accepting the performance parameter value when its value is totally unsatisfactory is β. This type of error is called a Type II error. This probability is achieved by selecting a sufficiently large sample size – the subject of this section.

The probability of acceptance is some function of the performance parameter value, called the operating characteristic (OC). The concepts of acceptance testing are depicted by the schematic OC curve in Figure 18. In this figure,

- the probability of accepting the hypothesis when performance is totally satisfactory is $1 - \alpha$,

- the probability of accepting the hypothesis when performance is at the threshold is 0.5, and

- the probability of (incorrectly) accepting the hypothesis when performance is totally unsatisfactory is β,

---

[23]It should be understood that a performance parameter value better than the totally satisfactory value is even more acceptable, so the composite hypothesis ("less than or equal to") is not stated.

[24]The one-sided 100α% significance level corresponds to a $100(1 - 2\alpha)$% confidence interval (i.e., two-sided).

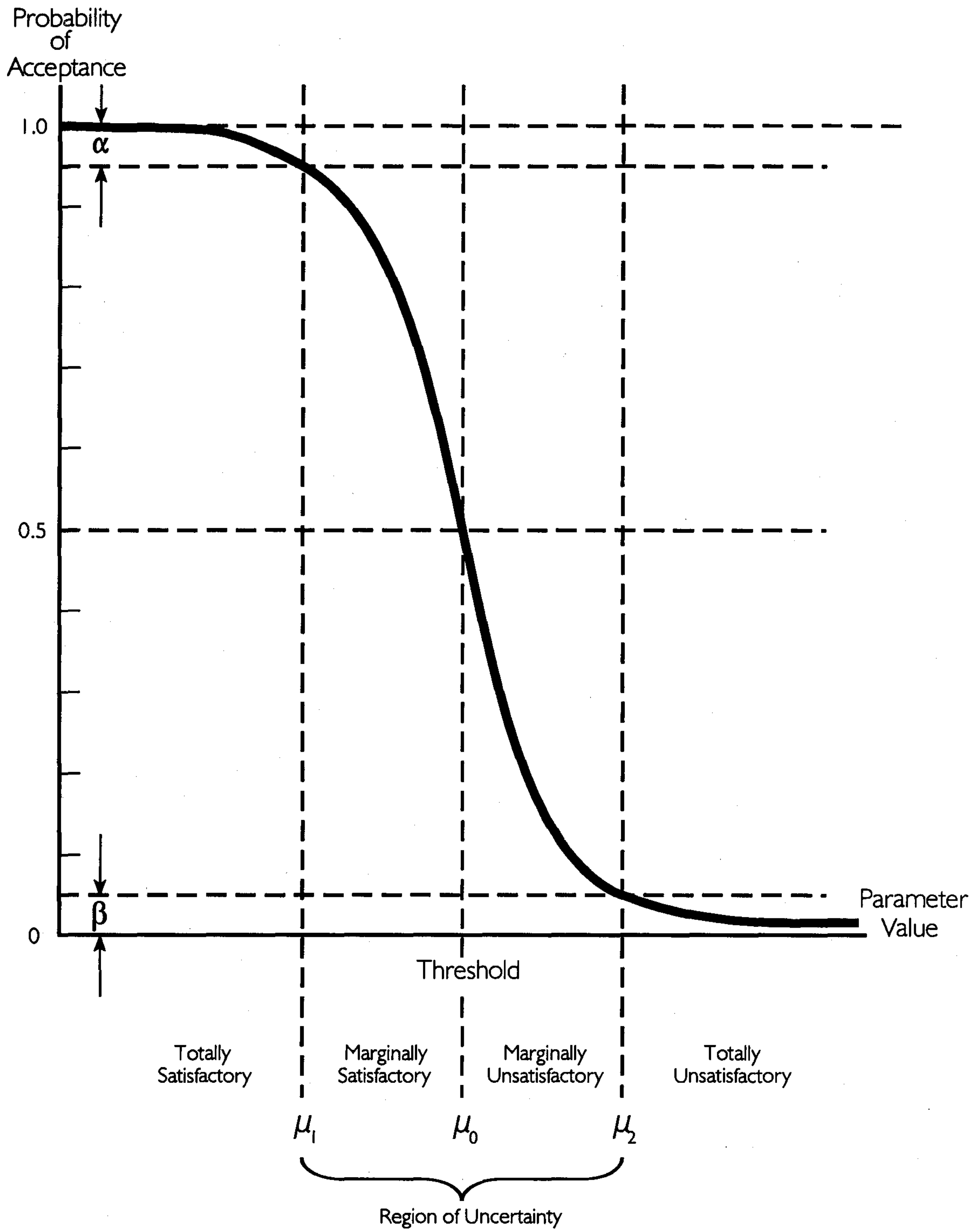


Figure 18. Schematic operating characteristic curve of a sampling plan for acceptance tests.

The sample size required to avoid the Type II error will be affected by the intertrial dependence - which is usually affected by the intertrial delay: The closer the trials are in time, the greater is the dependence (as measured by, r_1, the estimate of the autocorrelation of lag 1).

The acceptance test is the simplest for continuous variables such as time. Variations of it, such as testing the homogeneity of means of several samples by an F test, exist (Bowker and Lieberman, 1955; Odeh and Fox, 1975). However, these variations do not consider dependent trials.

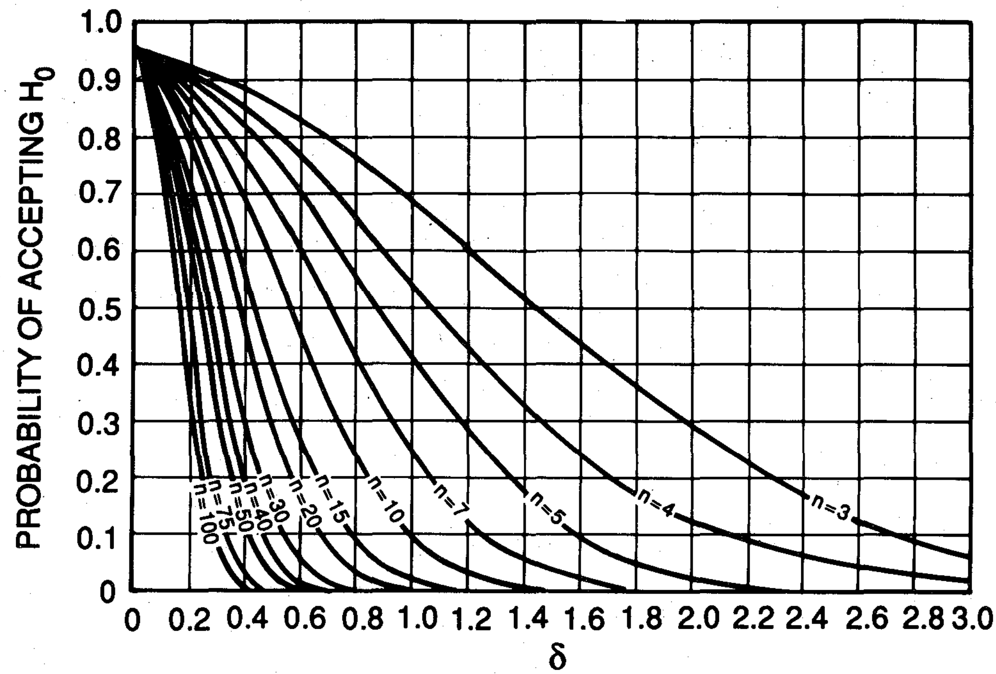
8.1.1 Time Parameters

Suppose that a mean delay of μ_0 would be barely acceptable (i.e., the threshold value). The true delay cannot be known with certainty from a finite sample. In other words, we cannot achieve the ideal OC (i.e., ideal being the probability of a Type II error is zero) with the probability of acceptance of unity for $\mu < \mu_0$ and of zero for $\mu > \mu_0$. That is, there is an interval of uncertainty about this probability. The boundaries separating the interval of uncertainty from the interval of totally satisfactory performance is μ_1, and the interval of totally unsatisfactory performance is μ_2. The interval to the left of this interval is the totally satisfactory interval, and the null hypothesis states that the (population) performance parameter value lies in this interval:

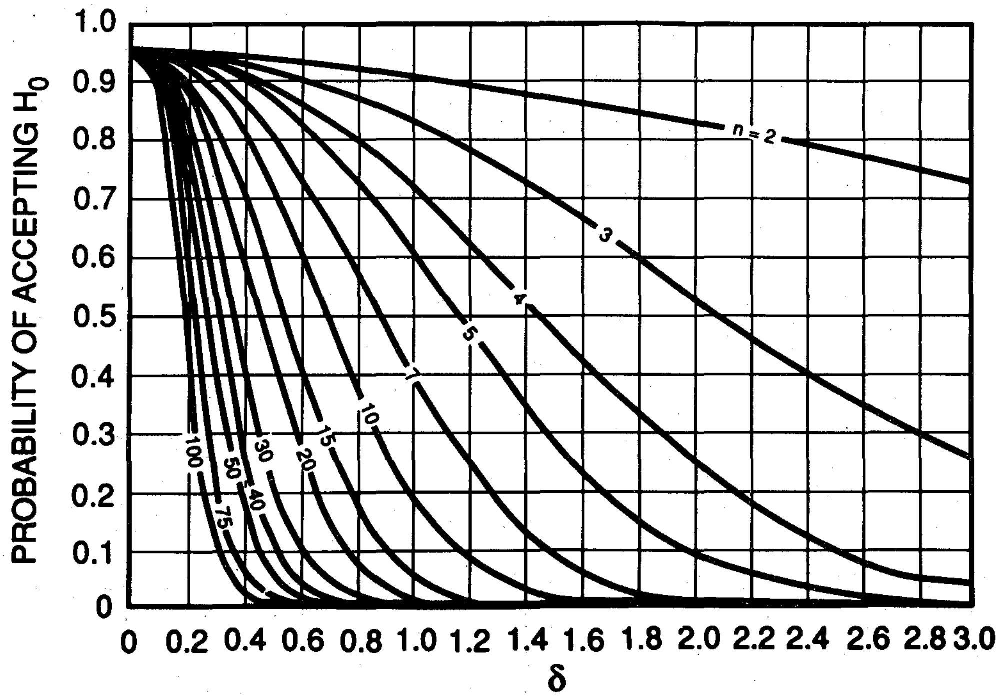
$$H_0: \mu = \mu_1 .$$

Delays are assumed to be normally distributed (at least approximately) with unknown standard deviation, and they are initially assumed to be independent.

If the standard deviation σ of the population is known, the sample size n can be determined. In practice, σ would probably not be known (even after a preliminary characterization test, since s is only an estimate of σ), but an approximation to it will suffice to approximate the required sample size. Figure 19 is a set of operating characteristic curves of the one-sided Student t test, (revised from Bowker & Lieberman, 1955). Figure 19a is a set of curves from the α = 5% significance level, and Figure 19b is a set of curves from the α = 2.5%



a. OC curves of the one-sided Student t test at the α=5% significance level.



b. OC curves of the one-sided Student t test at the α=2.5% significance level.

Figure 19. Sample size determination curves for an acceptance test of time parameters.

significance level.[25] In both cases, the ordinate is β, and the abscissa is

$$\delta = (\mu_2 - \mu_1)/\sigma \ .$$

To determine the sample size, n,

- select a probability β of (incorrectly) accepting a totally unsatisfactory delay, and
- determine δ by selecting μ_1 and μ_2 and estimating σ.

Dependence can approximately be accounted for if the required sample size is multiplied by

$$(1 + r_1)/(1 - r_1)$$

where r_1 is the estimate of the autocorrelation of lag 1 between successive delays.

Determine the sample size from the appropriate curve and modify it for dependence.

Example (delay): Suppose a proposed data communication service is required to have an Access Time of not more than 45 s. Furthermore, it is considered that 40 s would be totally satisfactory and 50 s would be totally unsatisfactory. It is assumed that an individual Access Time measurement has approximately a normal distribution with standard deviation $\sigma = 8$ s, and the autocorrelation of lag 1 is estimated to be 0.5. Determine the sample size required to obtain a probability of accepting a totally satisfactory time of $1 - \alpha = 0.95$ and a probability of accepting a totally unsatisfactory time of $\beta = 0.10$.

Solution: In this case, $\mu_1 = 40$ s, $\mu_2 = 50$ s, $\sigma = 8$ s, and $\delta = (50 - 40)/8 = 1.25$. Enter Figure 19a at the ordinate $\beta = 0.10$, and follow horizontally to the OC that crosses at $\delta = 1.25$. This curve is labelled n = 7 (or, perhaps, n = 8 to be conservative). That is, the required sample size is

[25] Formulas for these curves are complicated and involve an integral that cannot be solved exactly (Kotz and Johnson, 1985).

$n = 8$. Dependence can approximately be accounted for if the n obtained from Figure 19a is multiplied by

$$(1 + r_1)/(1 - r_1)$$

In this example, $r_1 = 0.5$, so that n is 24 rather than 8.

Since $\alpha = 5\%$ was selected, acceptance will be determined by $(1 - 2\alpha)100\% = 90\%$ confidence limits. This example is analyzed in the example of Section 3.1 of Volume 5.

Example (rate): A network is considered acceptable if the long-term throughput is 3 Mbps. A throughput of 3.6 Mbps (20% more) is considered to be totally satisfactory and a throughput of 2.4 Mbps (20% less) is considered to be totally unsatisfactory. From experience, the standard deviation of the throughput is 1.5 Mbps. Since there is one throughput trial per test, the trials are considered to be independent. Determine the sample size (number of tests) required to achieve the probability of accepting a totally satisfactory throughput of $1 - \alpha = 95\%$, and the probability of incorrectly accepting a totally unsatisfactory throughput of $\beta = 5\%$.

Solution: In this example, $\mu_1 = 3.6$ Mbps, $\mu_2 = 2.4$ Mbps, and $\sigma = 1.5$ Mbps. Then $\delta = (3.6 - 2.4)/1.5 = 0.8$. The ordinate of Figure 19a is $\beta = 0.05$ and the abscissa is $\delta = 0.8$. Approximately $n = 20$ tests are required to achieve the two probabilities. Since the trials are thought to be independent, $r_1 = 0$, and the required number of tests remains 20.

Since $\alpha = 5\%$ was selected, acceptance will be determined by $(1 - 2\alpha)100\% = 90\%$ confidence limits. This example is analyzed in the example of Section 3.1 of Volume 5.

8.1.2 Failure Probability Parameters

Specify p_0, the (threshold) failure probability that will be tolerated with indifference (i.e., probability of acceptance = 0.50). The true failure probability, p, cannot be known with certainty from a finite sample. In other words, we cannot achieve the ideal OC with the probability of acceptance of unity for $p < p_0$ and of zero for $p > p_0$. That is, there is a interval of uncertainty about this probability; the narrower the interval of uncertainty, the greater the precision. The boundaries separating the interval of uncertainty from the totally satisfactory performance and totally unsatisfactory performance are p_1 and p_2, respectively. The interval to the left of this interval is the totally satisfactory interval, and the null hypothesis states that the (population) performance parameter value lies in this interval:

$$H_0: p = p_1$$

The sample size is determined by specifying the two probabilities of the sampling plan:

- Probability of accepting a totally satisfactory failure probability = $P(p_1) = 1 - \alpha$.
- Probability of accepting a totally unsatisfactory failure probability = $P(p_2) = \beta$.

The sample size can be determined from Table 13a for $\alpha = 0.05$ and from Table 13b for $\alpha = 0.025$. In both tables, $\beta = 0.05$.[26] The operating ratio is defined as $R_o = p_2/p_1$, the relative width of the interval of uncertainty. Table 13 includes values of R_o as small as 2. Locate this ratio in the table, and note the corresponding required number of failures and np_1. Assuming the trials are independent, the sample size is obtained by solving for n. However, to account for dependence, multiply the sample size by the factor

[26]Table 13 is based on the Poisson distribution where μ_1 = np1, $\mu2$ = np2, and s = number of failures:

$$P[S \le s \mid \mu = \mu_1] = \sum_{i=0}^{s} (\mu_1^i e^{-\mu_1}/i!) > = 1 - \alpha,$$

and

$$P[S \le s \mid \mu = \mu_2] = \sum_{i=0}^{s} (\mu_2^i e^{-\mu_2}/i!) \le \beta.$$

For example, let $\alpha = 0.05$, $\beta = 0.05$, p2/p1 = 58, s = 0, $\mu1$ = np1 = 0.051, so $\mu2$ = np2 = 2.958. Then,

$$P[S = 0 \mid \mu = 0.051 = e^{-0.051} = 0.95028 \text{ (approx. 0.95)}.$$
$$P[S = 0 \mid \mu = 2.958 = e^{-2.958} = 0.05192 \text{ (approx. 0.05)}.$$

$$(1 + r_1)/(1 - r_1)$$

where r_1 is the estimate of the autocorrelation of lag 1.

> Determine the required sample size by locating the acceptable number of failures in the table, determine the expected number of trials, and modify it for dependence.

Example (Failure Probability): A proposed data communication service is required to have a Bit Error Probability not greater than $p_0 = 10^{-4}$, a value that is acceptable with indifference. We decide that sufficient assurance is provided if the totally satisfactory and totally unsatisfactory failure probabilities are, respectively,

$$p_1 = 10^{0.5} \times p_0, \text{ and } p_2 = 10^{0.5} \times p_0 \ .$$

Select $\alpha = 0.05$ and $\beta = 0.05$. The trials are dependent with autocorrelation of lag 1 which is estimated to be 0.4. Determine how many failures will permit acceptance of the service and how many failures will cause rejection of the service for this performance parameter.

Solution: now,

- Probability of accepting $p_1 = P(p_1) = P(10^{-0.5} \times p_0)$ $= P(3.162 \times 10^{-1} \times 10^{-4}) = 1 - \alpha = 0.95$

- Probability of accepting $p_2 = P(p_2) = P(10^{+0.5} \times p_0)$ $= P(3.162 \times 10^{-4}) = \beta = 0.05$

Assuming independence, the operating ratio, $R_o = p_2/p_1 = 10$, does not appear in Table 13a, but it can be satisfied conservatively by the operating ratio of 7.698. This operating ratio also corresponds to the required number of failures of 2 and $np_1 = 0.818$.

Table 13a. Constants for a Sampling Plan with OC Curves Through the Points (p_1, α= 0.05) and (p_2, β = 0.05)

OPERATING RATIO $R_o = p_2/p_1$	REQUIRED NUMBER OF FAILURES	np_1
53.615	0	0.052
13.328	1	0.356
7.698	2	0.818
5.673	3	1.267
4.644	4	1.971
4.022	5	2.614
3.604	6	3.286
3.303	7	3.901
3.074	8	4.696
2.895	9	5.426
2.749	10	6.170
2.629	11	6.925
2.528	12	7.690
2.442	13	8.465
2.367	14	9.247
2.301	15	10.036
2.243	16	10.833
2.192	17	11.635
2.145	18	12.442
2.103	19	13.255
2.065	20	14.072
2.030	21	14.894
1.998	22	15.720

Table 13b. Constants for a Sampling Plan with OC Curves Through the Points (p_1, α = 0.025) and (p_2, β = 0.05)

OPERATING RATIO $R_o = p_2/p_1$	REQUIRED NUMBER OF FAILURES	np_1
115.229	0	0.026
19.526	1	0.242
10.172	2	0.619
7.114	3	1.090
5.637	4	1.624
4.775	5	2.702
4.207	6	2.815
3.807	7	3.454
3.507	8	4.116
3.275	9	4.796
3.089	10	5.492
2.936	11	6.201
2.809	12	6.923
2.700	13	7.655
2.607	14	8.396
2.525	15	9.146
2.454	16	9.903
2.390	17	10.669
2.333	18	11.440
2.282	19	12.217
2.235	20	13.000
2.193	21	13.788
2.155	22	14.580
2.119	23	15.378
2.086	24	16.179
2.056	25	16.984
2.027	26	17.794
2.001	27	18.606
1.977	28	19.422

Test until 2 failures are observed. If, in fact, the failure probability is $p_1 = 3.162 \times 10^{-5}$, then the expected number of trials is

$$n = 0.818/(3.162 \times 10^{-5}) = 25,870.$$

That is, the service provided should be accepted if 2 or fewer failures are observed, and rejected if 3 or more failures are observed in 25,870 trials.

To account for dependence, the number of trials and failures should both be multiplied by

$$(1 + r_1)/(1 - r_1) = (1.4/0.6) = 2.3.$$

That is, the expected number of trials is 25,870 × 2.3 = 59,501. The service should be accepted if 4 or fewer failures are observed (because 2 × 2.3 = 4.6), and rejected if 5 or more failures are observed in 59,501 trials.

Since $\alpha = 5\%$ was selected, acceptance will be determined by $(1 - 2\alpha)100\% = 90\%$ confidence limits. This example is analyzed in the example of Section 3.2 of Volume 5.

8.2 Determine the Sample Size for Other Common Objectives

Section 2.2 lists several possible objectives for an experiment. This section shows how to determine the sample size required for a specified precision for all objectives except acceptance and maintenance.

The sample size required to estimate a performance parameter with the specified precision can be determined from the interactive program **star**. This program is contained on diskette 5.

To use **star** interactively, enter

- 0 (to determine the sample size),
- the number corresponding to the type of random variable to be analyzed (delay, rate, or failure probability),
- the number corresponding to the selected confidence level,
- the specified precision (relative or absolute), and
- some information about the dispersion of the population (i.e., the estimate of the maximum standard deviation and the estimate of the maximum autocorrelation of lag 1). The

preliminary characterization test provides this information, but without benefit of this test, these estimates are probably not available.

The program returns

- the sample size required to achieve the specified precision (if that is desired), and
- a code number to be entered for subsequent analysis of the test.

Appendix C shows in detail how the sample size is determined by an operator using **star** interactively.

The precision specified for a test might be reconsidered after the experiment is designed (Section 7). The experiment design might show that either more or fewer tests are required than earlier believed and, consequently, that less or more precision can be obtained within a given time and/or budget.

Enter the sample size and intertrial delay in file **default**.

9. ACKNOWLEDGMENTS

The author wishes to thank Dr. Edwin L. Crow for his suggestions (particularly with Section 8.1) and extensive technical assistance, and Dr. Earl D. Eyman and Catherine L. Edgar for verifying many examples of the software. He also is grateful to Christopher A. Bogart for writing the shell scripts that implement the preliminary characterization test and the subsequent determination of the test duration. The author wishes to also thank Margaret H. Morris for many suggestions and Lorna L. Kent, Robert S. Reichart, and David J. Atkinson for drawing the structured design diagrams. Rob Reichart also was responsible for coordinating all figures and tables.

10. REFERENCES

ANSI (1987), American National Standard for Information Systems - Data communication systems and services - measurement methods for user-oriented performance evaluation, ANSI X3.141-1987 (American National Standards Institute, Inc., New York, NY).

ANSI (1992), American National Standard for Information Systems - Data communication systems and services - user-oriented performance parameters, ANSI X3.102-1992 (American National Standards Institute, Inc., New York, NY).

Bowker, A.H., and G.J. Lieberman (1955), *Handbook of Industrial Statistics* (Prentice-Hall, Inc., Englewood Cliffs, NJ).

Cass, R.D., and M.J. Miles (1990), Design of experiments to quantify communication satellite system performance, NTIA report 90-269, November, 102 pp. (NTIS Order Number PB 91-129635)

Crow, E.L. (1978), Relations between bit and block error probabilities under Markov dependence, Office of Telecommunications Report 78-143 (U.S. Government Printing Office, Washington, DC).

Crow, E.L., F.A. Davis, and M.W. Maxfield (1960), *Statistics Manual* (Dover Publications, Inc., New York, NY).

Davies, O.L. (ed), (1956), *Design and Analysis of Industrial Experiments* (Hafner Publishing Company, New York, NY).

Johnson, N.L., and F.C. Leone (1977), *Statistics and Experimental Design in Engineering and The Physical Sciences*, Vol. II, 2nd Ed. (John Wiley and Sons, New York, NY).

Kotz, S., and N.L. Johnson (1985), *Encyclopedia of Statistical Sciences, Noncentral t-Distribution*, Volume 6, pp 286-290 (John Wiley and Sons, New York, NY).

Miles, M.J., and E.L. Crow, *Statistical Methods for Data Communication Performance Evaluation* (Book to be published).

Moses, L.E., and R.V. Oakford (1963), *Tables of Random Permutations* (Stanford University Press, Stanford CA).

Odeh, R.E., and M. Fox (1975), *Sample Size Choice: Charts for Experiments with Linear Models* (Marcel Dekker, New York, NY).

Spies, K.P., D.R. Wortendyke, E.L. Crow, M.J. Miles, E.A. Quincy, and N.B. Seitz (1988), User-oriented performance evaluation of data communication services: measurement design, conduct, and results, NTIA Report 88-238, August, 294 pp. (NTIS Order Number PB 89-117519/AS).

APPENDIX A: FORMULAS FOR SAMPLE SIZE DETERMINATION

This Appendix contains the mathematical formulas to determine the sample size necessary to estimate a performance parameter with the desired precision for objectives other than acceptance and maintenance.[1] Often the sample size is determined by time or budget. However, it is also often determined by the required precision.

If precision is required, the sample size is determined by either the relative or the absolute precision and the confidence level. In one type of problem, the equations determine the sample size from knowledge of certain maximum values of the population. In a second type, in the absence of such knowledge, they determine a preliminary sample size. Statistical information from the preliminary sample is then used to determine the number of additional observations (if any) necessary to achieve the specified precision.

These formulas are implemented in the FORTRAN program **star**.

A.1 Time Parameters

Even though time parameters consist of both delays and rates, the required sample size is determined from delays (i.e., the time to accomplish functions such as Access, Block Transfer, Source Disengagement, and Destination Disengagement). Delays are non-negative values. Since they are bounded below by zero but unbounded above, they cannot be normally distributed. Data indicate they are asymmetric, but not far from normally distributed (possibly a log-normal or gamma distribution). However, in this report we assume delays are normally distributed.

The sample mean has been selected to estimate the population mean. It is sometimes not used because it can be contaminated by outlying values. This objection has been removed since ANS X3.102 defines a delay to be a failure if it exceeds three times the specified delay (i.e., timeout). The sample mean is an unbiased estimate for any distribution and efficient for the normal and gamma

[1]Only the results of the theory are given here. For a detailed discussion of the theory, see a book authored by M. J. Miles and E. L. Crow (to be published).

distribution.[2] The sample median is another common estimate of the population mean. However, it is inefficient for the normal and gamma distributions. To estimate the population mean of the delays, specify

- the absolute precision with which the estimate must approximate the mean delay, and
- the level of confidence with which the absolute precision is to be achieved.

The required sample size can be determined if we know both the maximum value of the population standard deviation of the delays and the statistical dependence among them.[3] Otherwise a preliminary test of at least ten delays is recommended to estimate the sample standard deviation and the statistical dependence (as measured by the autocorrelation of lag 1).

The required sample size can be determined from one of the following four cases.

A.1.1 The Upper Bound of the Population Standard Deviation of the Delays is Known, and the Delays are Independent

If σ_{max} is the upper bound of the population standard deviation, u_α is the upper 100α% point of the normal density, and a is the absolute precision as specified by the half-length of the $100(1-2\alpha)$% confidence interval, the required number of delays is

$$n = \left(\frac{u_\alpha \sigma_{max}}{a}\right)^2 . \qquad \text{(A-1)}$$

[2]An estimate of a parameter is said to be an unbiased estimate if its expected value equals the parameter.

[3]The upper bound of the standard deviation can be approximated by multiplying a sample estimate of the standard deviation by a factor that depends upon the sample size and a confidence level whose upper confidence limit you believe represents the upper bound. For the upper 95% confidence limit, this factor is listed in Table C-1 of Appendix C.

A.1.2 The Upper Bound of the Population Standard Deviation is Known, and the Delays are Dependent with Known Autocorrelation of Lag 1

If ρ_1 is the autocorrelation of lag 1, the required number of delays is

$$n = \left(\frac{u_\alpha \sigma_{max}}{a}\right)^2 \cdot \left(\frac{1+\rho_1}{1-\rho_1}\right) \tag{A-2}$$

A.1.3 The Upper Bound, σ_{max}, of the Population Standard Deviation is Known, and the Delays are Dependent with Unknown Autocorrelation of Lag 1

There are five steps.

A. Conduct a preliminary test of n′ delays, w_i, (at least 10 delays) in order to estimate the autocorrelation of lag 1.

B. From the preliminary test, compute, $r_1(w)$, the estimate of the autocorrelation of lag 1:

$$r_1(w) = \frac{1}{s^2(n'-1)} \sum_{i=1}^{n'-1} (w_i - \overline{w})\ w_{i+1} - \overline{w}) \tag{A-3}$$

where

$$\overline{w} = \frac{1}{n'} \sum_{i=1}^{n'} w_i , \tag{A-4}$$

and

$$s^2 = \frac{1}{n'-1} \sum_{i=1}^{n'} (w_i - \overline{w})^2 . \tag{A-5}$$

C. Determine, $t_{n'-1,\alpha}$, the upper 100α% point of the Student t distribution corresponding to n′-1 degrees of freedom. (This value is determined in subroutine **studnt**.)

D. Compute

$$A(r_1) = t_{n'-1,\alpha} \cdot \frac{\sigma_{max}}{\sqrt{n'}} \cdot \sqrt{\frac{1+r_1(w)}{1-r_1(w)}} . \tag{A-6}$$

The quantity,

$$s_{\bar{w}} = \frac{s}{\sqrt{n'}} \cdot \sqrt{\frac{1 + r_1(w)}{1 - r_1(w)}}\ , \tag{A-7}$$

is a measure of the uncertainty of $\bar{w}$, called the standard error.

E. Now,

- if $A(r_1) > a$, observe

$$n = n' \left[\frac{A(r_1)}{a}\right]^2 - n' \tag{A-8}$$

additional delays.

- Otherwise no more delays need to be observed.

A.1.4 The Upper Bound of the Population Standard Deviation is not Known

Proceed as in A.1.3, except use s instead of σ_{max} in (A-6).

A.2 Failure Probability Parameters

If trials result in either success or failure and are statistically independent, they have the binomial distribution with failure probability, p. However, successive trials are usually dependent. In this case, one can model the failure probability with a Markov chain of order m (where a large m indicates a high order of dependence between successive trials). For instance, order m = 1 assumes that the occurrence of a failure depends (to some extent) on the occurrence of a failure on only the previous trial.

This report models the trials by a stationary first-order Markov chain. A stationary Markov chain is one that is independent of time. The first-order model results from defining the parameter, λ, the conditional probability of a failure given that a failure occurred in the previous trial.

Suppose x_1, x_2, ..., x_n is a sequence of identically distributed random variables, each of which can assume one of two values (0 for a success and 1 for a failure).

Then

$$p = p[x_i = 1] \text{ for } i = 1, 2, \ldots, n$$

and,

$$\lambda = P[x_i = 1 | x_{i-1} = 1] \text{ for } i = 2, 3, \ldots, n.$$

We see that

- $\lambda > p$ means 1's and 0's tend to cluster,
- $\lambda = p$ means the trials are independent, and
- $\lambda < p$ means 1's and 0's tend to alternate.

Let s be the number of failures and r be the number of pairs of consecutive failures. For a sample of size n,

$$s = \sum_{i=1}^{n} x_i, \text{ and } r = \sum_{i=2}^{n} x_{i-1}x_i \ .$$

To determine the sample size required to estimate the mean of the failure probability parameter with the desired precision, it is necessary to specify

- the relative precision with which the estimate should approximate the failure probability mean, and
- the level of confidence that the relative precision has been achieved.

The method to determine the required sample size depends upon whether or not the maximum value of λ (called λ_{max}) is known.[4]

A.2.1 λ_{max} is Known

The number of failures to be attained before testing is stopped is

$$s_0 = s_{ind} \cdot \frac{1 + \lambda_{max}}{1 - \lambda_{max}} \qquad \text{(A-9)}$$

where s_{ind} is a function of the specified relative precision and confidence level. This function is called **sindf** in subroutine **ssdflr**.

[4]Since testing is halted because a prescribed number of failures (not trials) has been achieved, failures have the negative binomial distribution.

A.2.2 λ_{max} is not Known

Seven steps are required to determine the number of failures.

A. Specify, b, the absolute precision with which λ is to be determined. For example, if λ is needed only to a rough approximation, specify b = 0.5. We are concerned only with λ being too large. For example, we might estimate λ to be 0.2 when it is actually 0.7. On the other hand, if λ is needed quite precisely (or thought to be near zero), specify b = 0.1. The sample size required to estimate λ is inverely proportional to b.

B. Specify the one-sided confidence level for λ, and determine its associated upper 100α percentage point, u_α, of the normal density.[5]

C. The number of failures necessary to estimate λ as close as or closer than b with the desired level of confidence is

$$s' = \left(\frac{u_\alpha}{2b}\right)^2 . \tag{A-10}$$

D. Conduct a preliminary test that generates at least s' failures.

E. If the test resulted in s failures (i.e., $s \geq s'$) and r pairs of consecutive failures, compute

$$\lambda_u = \frac{2s\hat{\lambda} + u_\alpha^2 + [(2s\hat{\lambda} + u_\alpha^2)^2 - 4s\hat{\lambda}^2(s + u_\alpha^2)]^{1/2}}{2(s + u_\alpha^2)} \tag{A-11}$$

where

$$\hat{\lambda} = \frac{r}{s - \frac{s}{n}} . \tag{A-12}$$

F. Determine

$$s_0 = s_{ind} \cdot \frac{1 + \lambda_u}{1 - \lambda_u} \tag{A-13}$$

---

[5]The confidence level of 95% is used in star, so u_α = 1.645.

G. Now,

- If $s_0 > s'$, observe

$$s = s_0 - s' \tag{A-14}$$

additional failures.

- If $s_0 < s'$, no more observations are required.

APPENDIX B: FLOWCHARTS FOR SAMPLE SIZE DETERMINATION

This appendix contains a set of flowcharts. Figure B-1 shows the structure of **star**; however, the analysis of multiple tests is not shown in detail. The flowcharts of each subroutine required to determine the sample size are listed alphabetically.[1]

In these flowcharts, diamonds indicate decisions, rectangles indicate arithmetic operations, and parallelograms indicate input/output.

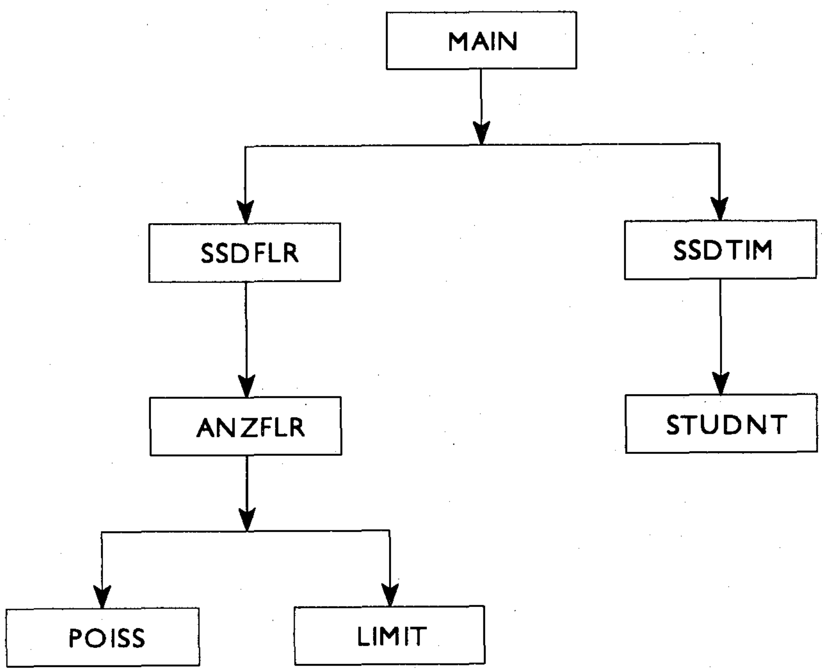


Figure B-1. Flowchart of relationship among subroutines of **star**.

[1]Three subroutines (called **entera**, **enteri**, and **enterx**) that allow entry of data and responses from a keyboard are omitted. Also, if data from time parameters are entered from a file only trivial changes occur in Figure B-1.

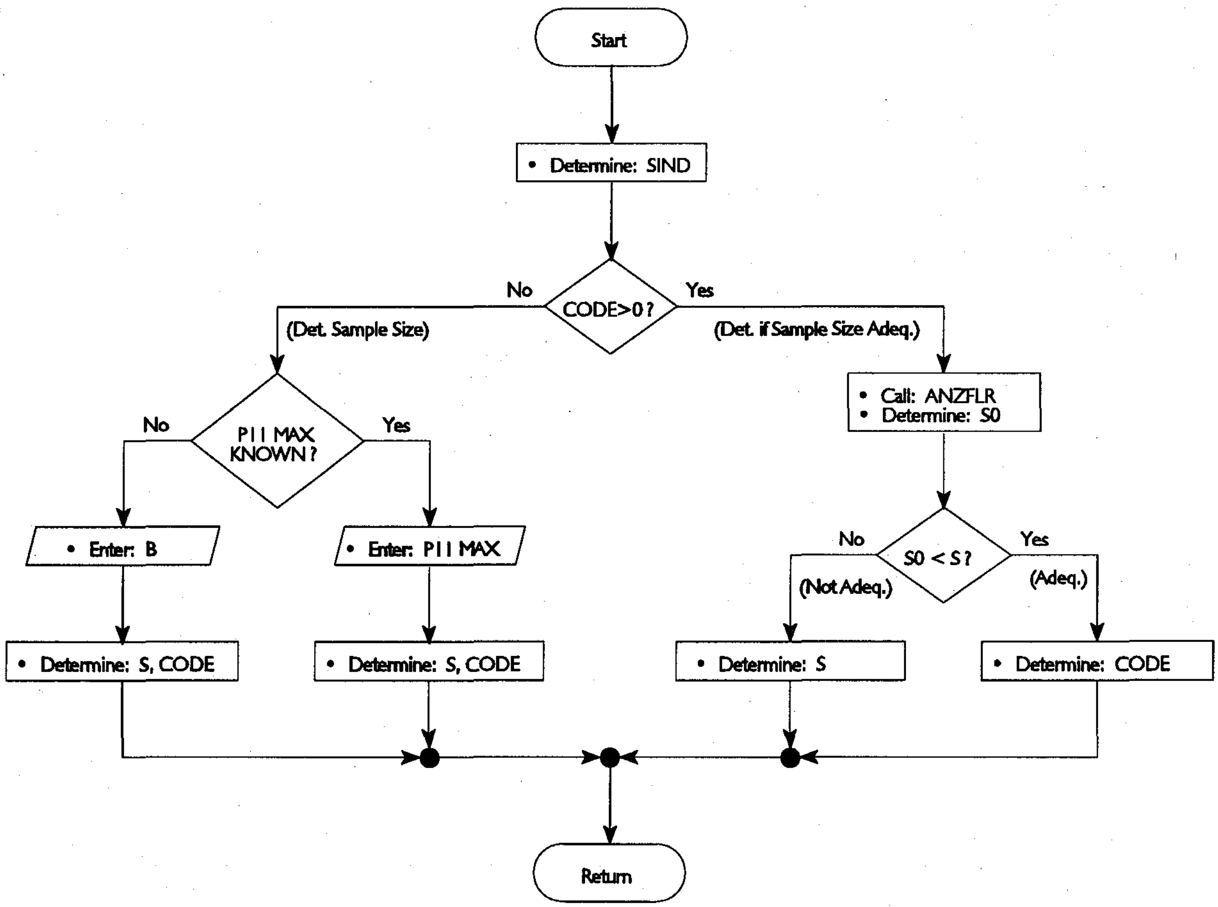


Figure B-2. Flowchart of subroutine ssdflr: sample size determination for failure probability parameters.

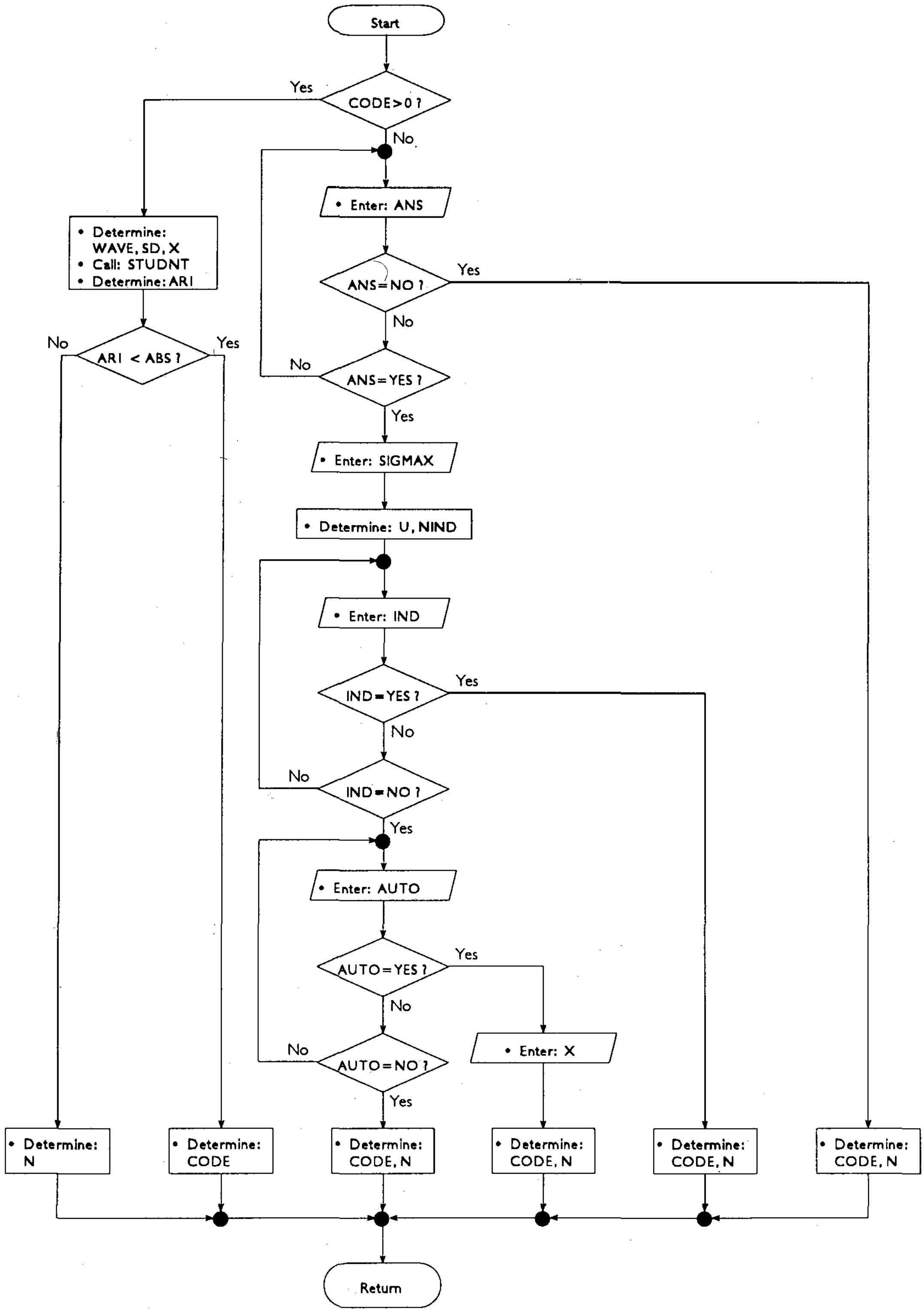


Figure B-3. Flowchart of subroutine **ssdtim**: sample size determination for time parameters.

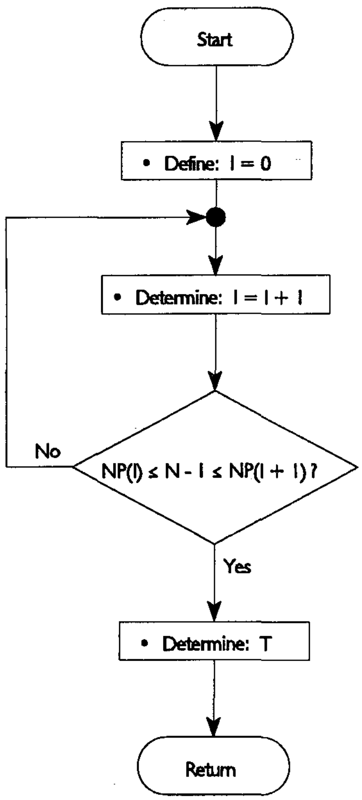


Figure B-4. Flowchart of subroutine **studnt**: percentage point of Student t distribution for specified confidence level.

APPENDIX C: OPERATOR IMPLEMENTATION OF SAMPLE SIZE DETERMINATION

The sample size is determined by the FORTRAN program **star**. Figure C-1 is a diagram of the operator's interaction with program **star** to determine the sample size for a test and, later, to analyze the test. This diagram shows only those decisions required of the operator. The top part shows how a sequence of three to seven decisions results in one of nine tests (labeled A through I). The bottom part shows how each test is analyzed. In three cases (tests B, E, and H), it is possible that the sample size was insufficient, and more data must be obtained. This can happen when the values in the sample varied more than expected or were more dependent than expected. The horizontal bar at the top of some boxes means that the program is accessed there, and the small circles in the analysis portion mean that the program makes a decision based upon the information previously entered by the user.

Figure C-2 is the operator-decision diagram for sample size determination; it is the top part of Figure C-1. As seen in the diagram, nine possible tests result from the sequence of decisions by the operator.[1]

When **star** is accessed, the introductory statements in Figure C-3 are listed.

Since the sample size is to be determined, type the integer zero. Then select the type of performance parameter to be tested. All performance parameters are one of three types: Delays, rates, and failure probabilities. The statements in Figure C-4 show how the type of performance parameter is selected. Although delays and rates are different types of parameters, the sample size for each is determined in the same way; both are referred to as time parameters.

The statements in Figure C-5 instruct the operator to select a confidence level.

The next two sections show how the sample size is determined for time and failure probability performance parameters.

---

[1]Since two confidence levels (90% and 95%) can be selected, there are actually 18 possible tests. However, the tests resulting from this selection differ only in the required sample size, not in the method of analyzing the data.

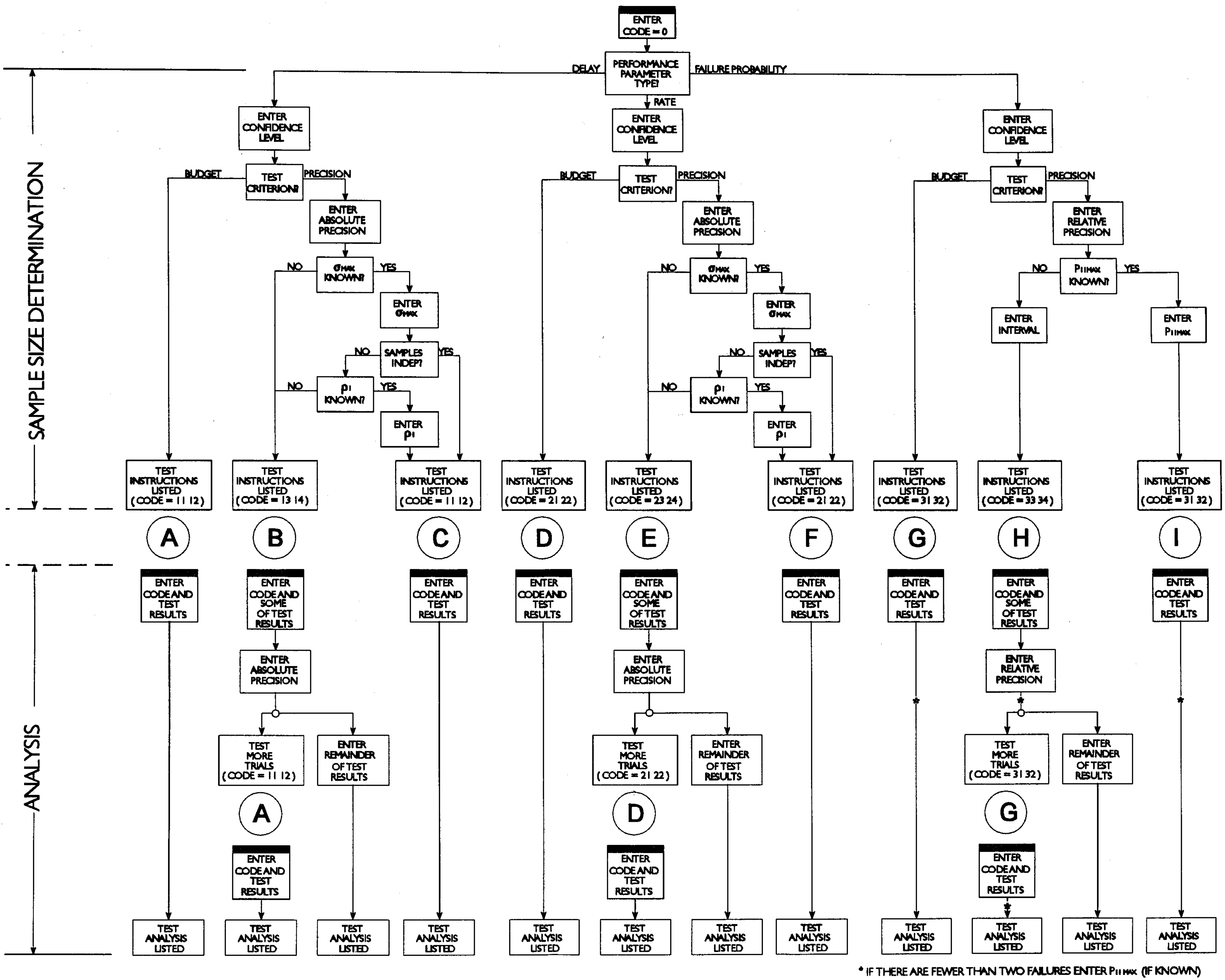


Figure C-1. Operator interaction with **star** for sample size determination and subsequent analysis of a single test.

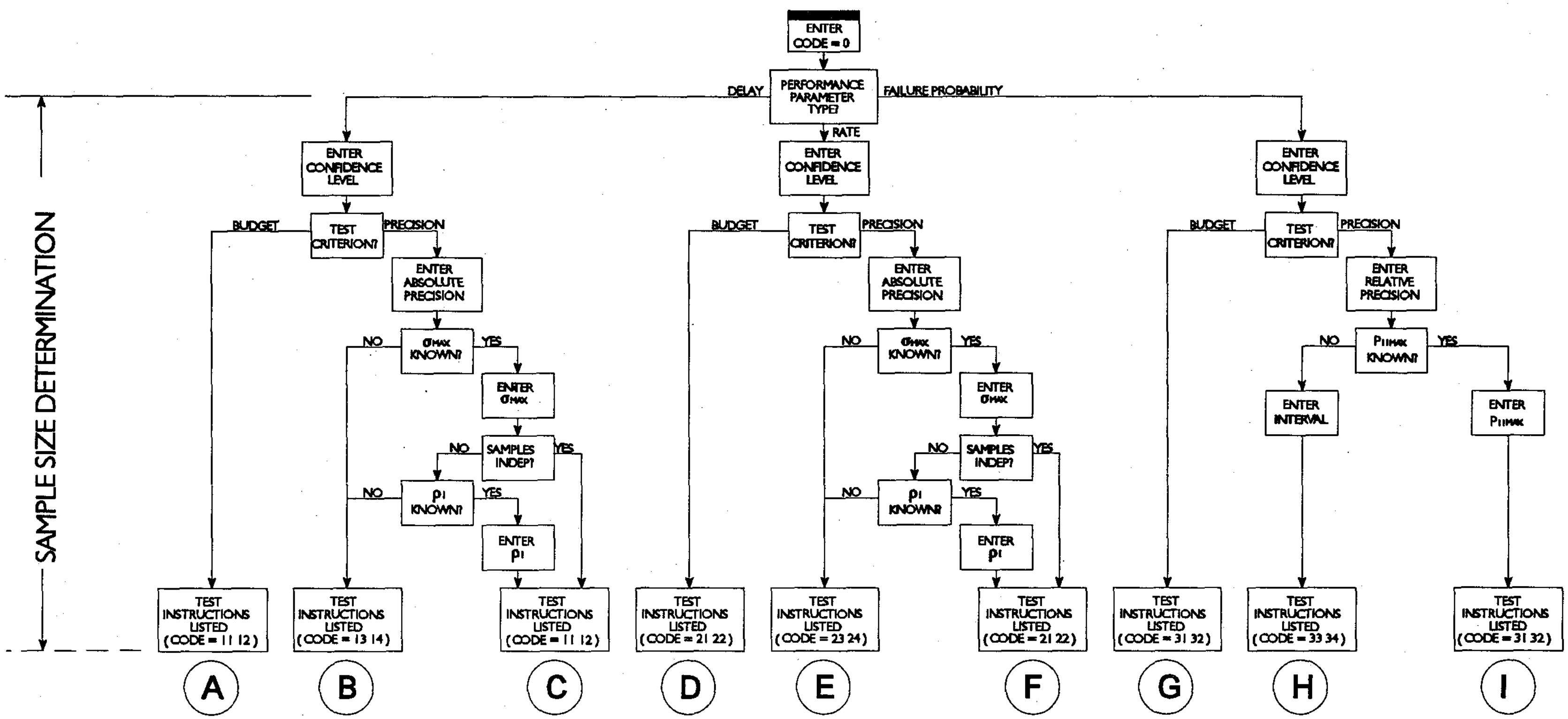


Figure C-2. Operator interaction with **star** for sample size determination.

```
THIS IS THE NTIA EXPERIMENT DESIGN AND DATA ANALYSIS
COMPUTER PROGRAM (STAR).
  IF YOU ARE ACCESSING THIS PROGRAM TO DETERMINE THE SAMPLE
SIZE FOR YOUR TEST, PLEASE TYPE THE INTEGER 0.
  IF YOU ARE ACCESSING THIS PROGRAM TO ANALYZE A SINGLE TEST,
PLEASE TYPE THE CODE NUMBER YOU WERE ASSIGNED WHEN THE
SAMPLE SIZE WAS DETERMINED.
  IF YOU ARE ACCESSING THIS PROGRAM TO ANALYZE MULTIPLE TESTS,
PLEASE TYPE THE INTEGER 40.
```

Figure C-3. Introductory message from star.

```
      YOU CAN TEST THE SYSTEM WITH RESPECT TO
  1.  DELAYS,
  2.  RATES,
OR
  3.  FAILURES
 PLEASE TYPE THE INTEGER LISTED AT THE LEFT OF THE TYPE
OF PERFORMANCE PARAMETER THAT YOU WISH TO ANALYZE.
```

Figure C-4. Performance parameter selection.

```
  THE PERFORMANCE PARAMETER CAN BE ESTIMATED
WITH ONE OF THE FOLLOWING LEVELS OF CONFIDENCE:
  1.  90%
  2.  95%
  PLEASE TYPE THE INTEGER LISTED AT THE LEFT OF THE
CONFIDENCE LEVEL THAT YOU
HAVE SELECTED.
```

Figure C-5. Confidence level selection.

C.1 Time Parameters

The two types of time parameters are the delay and rate parameters. The sample size required for both is determined from the same random variable, the time to accomplish a function. For delays, the time to accomplish is the delay itself. For the single rate parameter the time to accomplish is the Input/Output Time; the User Information Bit Transfer Rate is then the number of Successful Bit Transfer outcomes divided by the Input/Output Time[2]. From Figure C-2, it can be seen that the procedure for delays and rates is identical; hence, it is convenient to discuss the sample size determination for delays and mention rates only parenthetically. The sequence of decisions results in one of three tests: A, B, and C for delays (D, E, and F for rates).

To begin, the test criterion is the operator's decision to test a sample whose size is determined by either budget or precision. If the sample size is determined by budget, no further decision is required, and the computer program lists the test instructions. This test is called test A (test D for rates).

On the other hand, if the sample size is to be large enough to provide a given precision (with a specified confidence level), it is necessary to specify the largest acceptable error in estimating the true mean delay (rate). This error corresponds to the absolute precision (defined in Section 1.2.4).

The next statement asks if the maximum value of the population standard deviation is known. The standard deviation (i.e., the square root of the variance) is a measure of the dispersion of the values. If $x_1, x_2, \ldots, x_n$ are values from a sample of size n, and $\overline{x}$ is the sample mean, an estimate of the population standard deviation is

$$s = \left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2\right]^{1/2}.$$

(If all outcomes are equal, the standard deviation is zero.) If the maximum value of the population standard deviation is not known, the instructions are

---

[2]Since NTIA implementation results in a single trial per test of User Information Bit Transfer Rate and User Fraction of Input/Output Time, the sample size needn't be determined. However, since the experimenter may have other uses in mind, rate is maintained here as an option.

listed for test B (test E for rates). If the maximum value is known, that value is to be entered. Table C-1 lists values of a factor which, when multiplied by s, estimates the upper 95% confidence limit for the standard deviation - essentially the maximum value. The factor depends upon the sample size (which is to be determined here). However, the table shows the degree of importance of knowing the maximum value; for larger sample sizes, it is less important (Crow et al., 1960, Table 8).

Table C-1. Factor Which, When Multiplied by the Standard Deviation, Estimates the Upper 95% Confidence Limit of the Standard Deviation

n - 1	FACTOR	n - 1	FACTOR	n - 1	FACTOR	n - 1	FACTOR
1	31.910	11	1.698	21	1.429	40	1.280
2	6.285	12	1.651	22	1.415	50	1.243
3	3.729	13	1.611	23	1.403	60	1.217
4	2.874	14	1.577	24	1.391	70	1.198
5	2.453	15	1.548	25	1.380	80	1.183
6	2.202	16	1.522	26	1.370	90	1.171
7	2.035	17	1.499	27	1.361	100*	1.161
8	1.916	18	1.479	28	1.352		
9	1.826	19	1.461	29	1.344		
10	1.755	20	1.444	30	1.337		

*For n-1 > 100, the factor is

$$\frac{1}{1 - \frac{1.960}{[2(n-1)]^{1/2}}}$$

Then the program asks if the trials are statistically independent. Two trials are statistically independent if the previous outcome of one trial has nothing to do with the probable outcome of the other. Strictly, consider the events A and B. Suppose the probability that A will occur is P(A), and the conditional probability that A will occur, given that B has occurred, is P(A|B). Events A and B are statistically independent if

$$P(A|B) = P(A)\cdot$$

If the trials are statistically independent, the instructions for test C (test F for rates) are listed.

If they are not statistically independent, some serial dependence exists among the trials. Serial dependence for the lag k = 1, 2,..., n - 1 is measured by the sample autocorrelation function,

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \overline{x})(x_{i+k} - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.$$

This function has a value approximately between -1 and 1 for large n. (It is a convenient estimate of the population autocorrelation function, ρ_k which occurs between -1 and 1, but the estimate may not occur there for small n; for example, for n = 2, r_1 = -2 whatever the values of x_1 and x_2.) In the case of dependence, the function could have a value near these extremes. But, in the case of independence, the value will tend to be near zero. Serial dependence for lag k can be observed by plotting, in the x-y plane, the n – k points whose coordinates are $x_i = x_i$ and $y_i = x_i + k$. Positive autocorrelation is indicated by points that tend to approximate a curve with a continuously positive slope. Negative autocorrelation is indicated by points that tend to approximate a curve with a continuously negative slope; and zero autocorrelation is indicated by points that have no systematic tendency.

Trials that occur close in time or space are more likely to be either more similar or more dissimilar than are trials that are not close. Hence, if serial dependence exists it probably occurs between adjacent trials (i.e., trials of lag 1). Table C-2 lists some average autocorrelation coefficients observed for a number of tests of three public data networks (labelled A, B, and C).

If the autocorrelation of lag 1 is known, at least approximately, this value should be entered; then the instructions for test C (test F for rates) are listed. Otherwise, instructions for test B (test E for rates) are listed.

Instructions for tests A, B, and C (tests D, E, and F for rates) consist of a list of information that will be required when the computer program is re-accessed to analyze the test results.[3] Notice that test B (test E for rates)

[3] If the test results will be entered through a file (instead of through a keyboard) the number of delays (or transfers) needn't be entered.

Table C-2. Autocorrelation of Lag 1 for Time Parameters as Measured on Three Public Data Networks (Labeled A, B, and C). The Time Between Access Attempts was 55 s, and the Time Between Block Transfers was 0 s

Access-Disengagement

	NO. TESTS	NO. TRIALS	AUTOCOR.
ACCESS TIME			
A	11	209	0.182
B	12	212	0.056
C	7	111	-0.008
S. DISENG. TIME			
A	11	194	-0.149
B	12	194	-0.129
C	7	102	0.538
D. DISENG. TIME			
A	11	207	-0.002
B	12	212	-0.017
C	7	111	0.116

User Information Transfer*

	NO. TESTS	NO. TRIALS	AUTOCOR.
BLOCK TRANSFER TIME			
A	7	559	0.349
B	7	560	0.763
C	7	560	0.955

*There is only trial per test for the User Information Bit Transfer Rate and the User Fraction of Input/Output Time. Hence, there is no autocorrelation for these two parameters.

resulted because not much was known about the population; hence, it is possible that the suggested sample size is too small to achieve the desired precision. If so, the analysis portion of the computer program will determine the number of additional trials required to achieve this precision.

The code numbers required for re-access are either 11 or 13 (21 or 23 for rates) for the 90% confidence level and either 12 or 14 (22 or 24 for rates) for the 95% confidence level. The computer program is designed to analyze as many as 200 trials.

Example: Determine the sample size necessary to estimate the Access Time within 0.7 seconds and be 90% confident of this. Assume that the maximum value of the standard deviation of the delays is known to be 1.5 seconds, and the trials are statistically independent.

Solution:

1. Access the computer program.
2. Type **0** (the code number to determine the sample size), and press the return key.
3. Type **1** (for delays), and press the return key.
4. Type **1** (for the 90% confidence level), and press the return key.
5. Type **2** (to obtain precision), and press the return key.
6. Type **0.7** (the largest acceptable error), and press the return key.
7. Type **YES** (the maximum standard deviation is known), and press the return key.
8. Type **1.5** (the maximum value), and press the return key.
9. Type **YES** (the trials are statistically independent), and press the return key.

Now, instructions for the test are listed in Figure C-6.

TO ACHIEVE YOUR TEST OBJECTIVE, YOU MUST GENERATE
AT LEAST 13 DELAYS. WHEN YOU RE-ACCESS THIS
PROGRAM TO ANALYZE YOUR TEST, YOU WILL BE ASKED TO ENTER:
1. **YOUR CODE NUMBER (IT IS 11.),**
2. **THE NUMBER OF DELAYS,**
3. **THE TOTAL DELAY IN EACH TRIAL (IN CHRONOLOGICAL ORDER),**
4. **THE USER-RESPONSIBLE PORTION OF THE DELAY IN EACH TRIAL (IN CHRONOLOGICAL ORDER).**

Figure C-6. Instructions for sample size determination for time parameters.

Appendix B of Volume 5 discusses analyses of the test results for time parameters. The analysis part of this example, which resulted in Test C, is shown at the end of Appendix B of that volume.

C.2 Failure Probability Parameters

Figure C-2 shows the sequence of decisions that results in one of three possible tests of failure probability (Tests G, H, and I).

If the sample size is dictated by budget or time, the test instructions are listed for Test G.

If failure probability is to be tested for a given precision, the desired relative precision must be specified (defined in Section 1.2.4).

Suppose p is the probability of failure on any one trial, and λ is the conditional probability of a failure given that a failure occurred in the previous trial. Then,

- $\lambda > p$ means that failures and successes tend to cluster,
- $\lambda = p$ means that the trials are independent, and
- $\lambda < p$ means that failures and successes tend to alternate.

The operator is now asked to estimate the maximum value of the conditional probability, called λ_{max}. Like all probabilities, the conditional probability is between 0 and 1. The larger the estimate of λ_{max}, the larger the required sample size. If this can be estimated, it is entered, and the instructions for Test I are listed. Table C-3 lists some average autocorrelations of lag 1 obtained from tests of three public data networks (labelled A, B, and C).

If the maximum value cannot be estimated, estimate the amount above the conditional probability that the conditional probability is allowed to exceed only 5% of the time. This amount is the absolute precision for λ. The instructions for Test H are then listed.[4]

---

[4]The computer program cannot accept a value equal to 1.

Table C-3. Autocorrelation of Lag 1 for Failure Probability Parameters for Access-Disengagement Tests as Measured on Three Public Data Networks (Labeled A, B, and C). Time Between Access Attempts was set at 55 s

	NO. PAIRS	NO. FAIL.	NO. TRIALS	AUTOCOR.
INCORRECT ACCESS				
A	0	0	220	indet.*
B	0	0	240	indet.
C	0	0	140	indet.
ACCESS OUTAGE				indet.
A	0	0	220	-0.087
B	0	11	240	indet.
C	0	0	140	
ACCESS DENIAL				
A	4	11	220	0.112
B	2	17	240	-0.010
C	8	29	140	-0.101
S. D. DENIAL				
A	1	15	209	-0.050
B	1	18	212	0.089
C	0	9	111	0.162
D. D. DENIAL				
A	0	2	209	-0.111
B	0	0	212	indet.
C	0	0	111	indet.

*"indet." means the autocorrelation is indeterminate (because the number of failures is zero).

Instructions for Tests G, H, and I list the information that will be required when the computer program is re-accessed to analyze the test results. Notice that Test H results when not much is known about the population; hence, it is possible that the suggested sample size is too small to achieve the desired precision. If so, the analysis portion of the computer program will state the number of additional trials required to achieve this precision.

The re-access code numbers are either 31 or 33 for the 90% confidence level and either 32 or 34 for the 95% confidence level.

Example: Determine the sample size necessary to estimate a failure probability within 30% of its true value and be 90% confident of this. Assume that you cannot estimate the maximum value of the conditional probability (of a failure, given that a failure occurred in the previous trial). However, you want the conditional probability to be exceeded by 0.2 only 5% of the time.

Table C-4. Autocorrelation of Lag 1 for Failure Probability Parameters for User Information Transfer Tests as Measured on Three Public Data Networks (Labeled A, B, and C). Time Between Block Transfers was 0 s

	NO. PAIRS	NO. FAIL.	NO. TRIALS	AUTOCOR.
BIT ERROR				
A	2	4	573,440	0.500
B	0	0	573,440	indet.
C	0	0	573,440	indet.
EXTRA BIT				
A	0	0	573,440	indet.
B	0	0	573,440	indet.
C	0	0	573,440	indet.
BIT LOSS				
A	0	0	573,440	indet.
B	0	0	573,400	indet.
C	0	0	573,000	indet.
BLOCK ERROR				
A	0	0	560	indet.
B	0	0	560	indet.
C	0	0	560	indet.
EXTRA BLOCK				
A	0	0	560	indet.
B	0	0	560	indet.
C	0	0	560	indet.
BLOCK LOSS				
A	0	0	560	indet.
B	0	0	560	indet.
C	0	0	560	indet.
TRANSFER DENIAL				
A	0	0	273	indet.
B	0	0	273	indet.
C	0	0	273	indet.

Solution:

1. Gain access to the computer program.

2. Type 0 (the code number to determine the sample size), and press the return key.

3. Type 1 (for failure probability), and press the return key.

4. Type 1 (for the 90% confidence level), and press the return key.

5. Type **2** (to obtain precision), and press the return key.

6. Type **30** (the percent of relative precision), and press the return key.

7. Type **NO** (since the maximum value of the conditional probability is not known), and press the return key.

8. Type **0.2** (since the conditional probability is to be exceeded by 0.2 with 5% probability), and press the return key.

Now, the instructions for the test are listed in Figure C-7.

```
  TO ACHIEVE YOUR TEST OBJECTIVE, YOU MUST GEN-
ERATE AT LEAST     17 FAILURES.  AFTER THE TEST YOU WILL
RE-ACCESS THIS PROGRAM TO ANALYZE THE PERFORMANCE OF
THE COMMUNICATION SYSTEM.  YOU WILL BE ASKED TO ENTER:
   1. YOUR CODE NUMBER.  (IT IS 33.),
   2. THE SAMPLE SIZE,
   3. THE NUMBER OF FAILURES IN THE SAMPLE,
   4. THE NUMBER OF PAIRS OF CONSECUTIVE FAILURES
      IN THE SAMPLE,
   5. THE SPECIFIED RELATIVE PRECISION.
```

Figure C-7. Instructions for sample size determination for failure probabilities.

Appendix D of Volume 5 discusses analysis for failure probability parameters. The analysis part of this example, which resulted in Test H, is shown at the end of Appendix D of that volume.

C.3 References

Crow, E.L., F.A. Davis, and M.W. Maxfield (1960), *Statistics Manual* (Dover Publications, Inc., New York, NY).

APPENDIX D: SHELL SCRIPT IMPLEMENTATION OF THE PRELIMINARY CHARACTERIZATION TEST AND THE RESULTANT TEST DURATION

This appendix describes the shell script **dopre**. It implements the program star

- to analyze a preliminary test (intended to characterize the performance parameters), and
- to use these results to accurately determine the sample size required to obtain the five specified precisions and the resultant test durations.

D.1 Preliminary Characterization Tests

There is one preliminary characterization test for access-disengagement tests and one for user information transfer tests. The results are stored in **/usr/net/.../data/2p/nnnn.pre**.

D.1.1 Access-Disengagement Tests

The shell script **dopre** calls the shell script **reduc-a** to analyze the performance parameters for the access and disengagement functions.[1] Figure D-1 is a structured design diagram showing this procedure.

A. Time Parameters

reduc-a calls the shell script **time-a** to analyze time parameters. **time-a** uses the three files **ACO** (access outcomes), **D1O** (source disengagement outcomes), and **D2O** (destination disengagement outcomes) and removes the header and trailer records to produce the three modified files called **statin**. Then **time-a** creates the prompt files, **prompt.90** and **prompt.95**, calls **star**, and uses **prompt.90** and **prompt.95** and the three **statin** files to estimate the mean, standard deviation, and autocorrelation of lag 1. The shell script **pretabt** prepares a line of this information and places it in the temporary file **prelim.tab**.

B. Failure Probability Parameters

reduc-a calls the shell script **fail-a** to analyze failure probability parameters. **fail-a** allows the three files **ACO**, **D1O**, and **D2O** to be used by the

[1]Both **reduc-a** and **reduc-x** (Section D.1.2) are discussed in Volume 4.

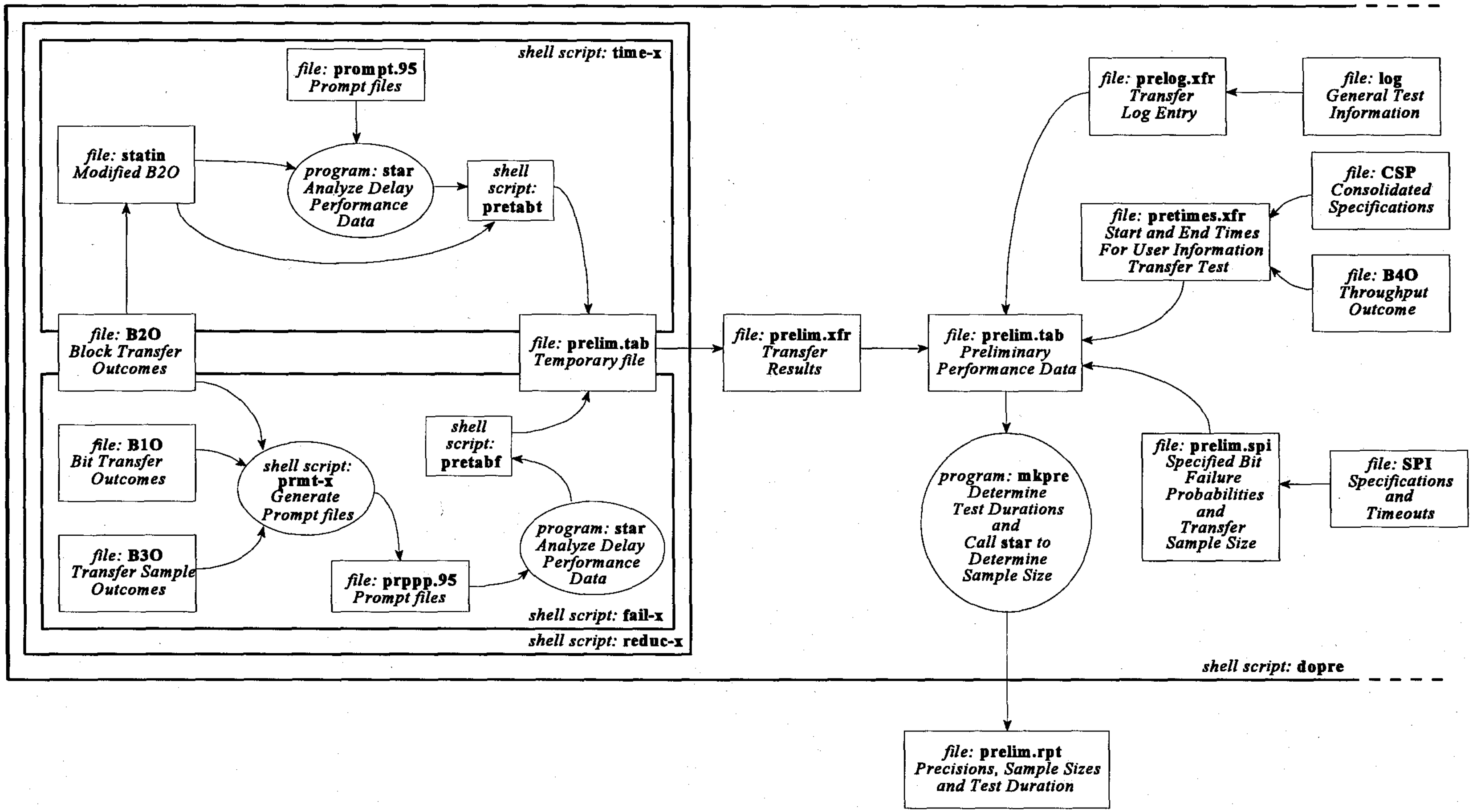


Figure D-1. Structured design diagram of shell script implementation on an access-disengagement preliminary characterization test.

shell script **prmt.x** to generate the prompt files **prppp.90** and **prppp.95**. The characters **ppp** in the file name act as a place holder representing three characters that refer to different parameters (e.g., **ppp** refers to the **acl** of Access Denial). Then **fail-a** calls **star** which estimates the mean, standard deviation, and autocorrelation of lag 1. The shell script **pretabf** prepares a line of this information and places it in the temporary file **prelim.tab**.

After **fail-a** and **time-a** have executed, **prelim.tab** is moved to **nnnn.pre** in **data/2p**.

D.1.2 **User Information Tests**

dopre calls the shell script **reduc-x** to analyze the user information transfer parameters. This procedure is shown in Figure D-2.

A. **Time Parameters**

reduc-x calls **time-x**. **time-x** creates the files **prompt.90** and **prompt.95** and uses the file **B20** and removes the header and trailer records to produce the modified file **statin**.[2] Then **time-x** calls **star** which uses **prompt.90** and **prompt.95** and **statin** to estimate the mean, standard deviation, and autocorrelation of lag 1. The shell script **pretabt** creates a line with this information and places it in the temporary file **prelim.tab**.

B. **Failure Probability Parameters**

fail-x uses the three files **B10**, **B20**, and **B30** in another shell script, **prmt.x** to create the prompt files **prppp.90** and **prppp.95** (where **ppp** is a place holder as described above). The values in this file are used by **star** to estimate the mean, standard deviation, and autocorrelation of lag 1. The shell script **pretabf** places a line of information in the temporary file **prelim.tab**.

After **fail-x** and **time-x** have executed, the file **prelim.tab** is moved to **prelim.xfr**.

[2]Since each test of User Information Bit Transfer Rate and User Fraction of Input/Output Time has but one trial, precision cannot be estimated for these parameters.

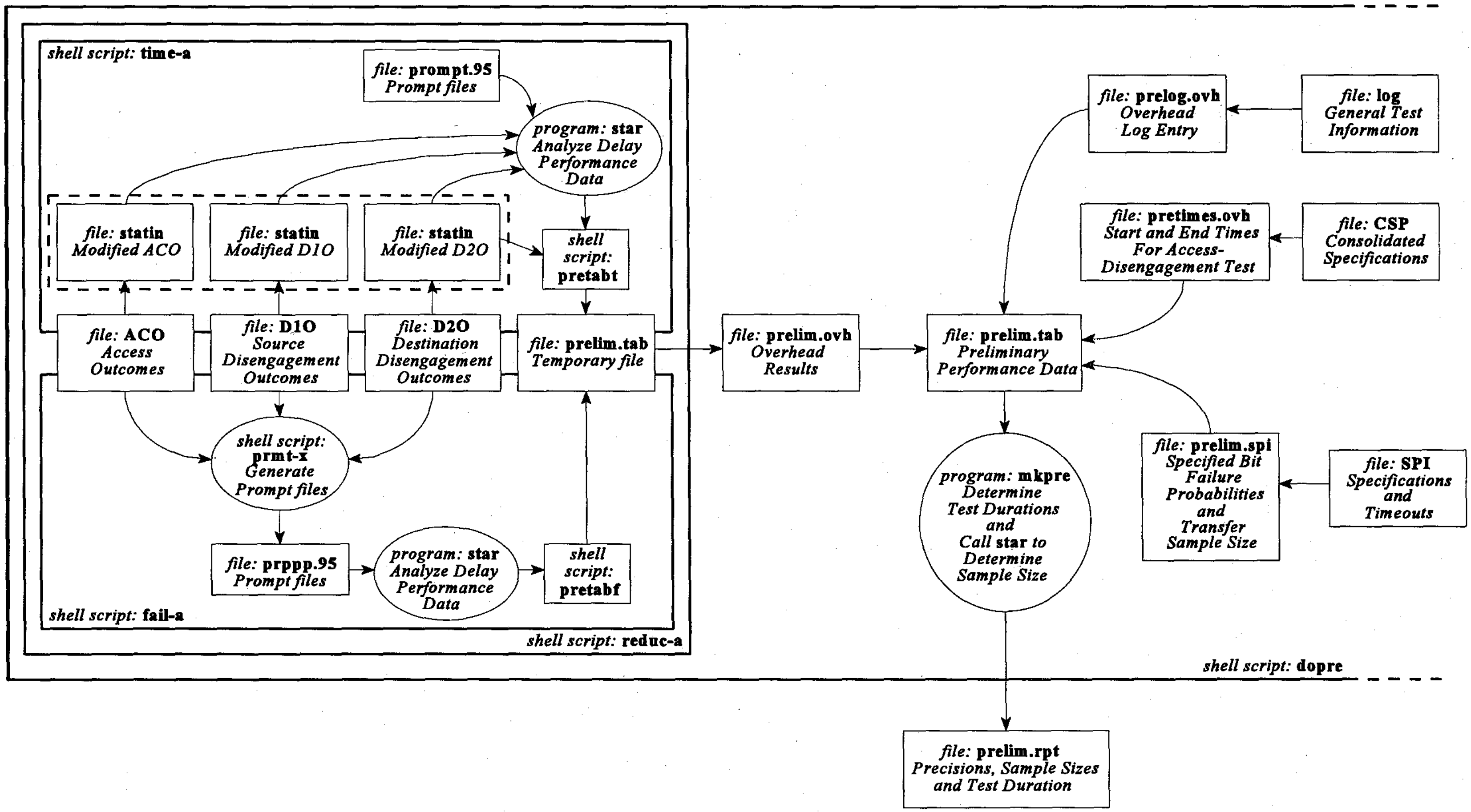


Figure D-2. Structured design diagram of shell script implementation of a user information transfer preliminary characterization test.

D.2 Sample Size Determination

In the following discussion, it is convenient to describe the processing of both types of tests together; with the exception of the **B40** file, present in user information transfer tests, they are the same.

The shell script **dopre** creates the files **prelog.ovh** and **prelog.xfr** which contain the log entry from file **log** (a single line of information for the corresponding test).

The files **pretimes.ovh** and **pretimes.xfr** contain the starting and ending times of the preliminary tests (recorded as **tbat** and **tbatz** in **CSP** and given in seconds since January 1, 1970). **pretimes.xfr** also contains a line of throughput outcome from the file **B40**. This is the Input/Output Time, the user portion of Input/Output Time, and the number of bits transferred.

The file **prelim.spi** contains a line from the file **spi.xfr** which indicates the specified performance values of Bit Error Probability, Extra Bit Probability, and Lost Bit Probability, and the size of the transfer sample.

The seven files **prelim.xfr**, **prelim.ovh**, **prelog.xfr**, **prelog.ovh**, **pretimes.ovh**, **pretimes.xfr**, and **prelim.spi** are concatenated (in this order) to form the file **prelim.tab**. This file is submitted as input to the program **mkpre** which summarizes the data and calls **star** to determine the sample size required to achieve a specified precision. It does this for five selected precisions for each performance parameter.

It provides the necessary responses to the prompts from **star**:

- 0 for sample size determination,
- 1 or 3 for delays or failure probabilities, respectively,[3]
- 2 for the 95% confidence level, and
- 2 to compute the sample size that will provide the required precision.

[3] 2 (for rate) is not entered because there is only one trial of User Information Bit Transfer Rate per test, and, hence, there is no precision.

D.2.1 **Time Parameters**

The absolute precision achieved during the preliminary characterization test is partitioned linearly into five fractions of the achieved absolute precision, P_a:

$$P_{a1} = (1/5)P_a,\ P_{a2} = (2/5)P_a, \ldots, P_{a5} = (5/5)P_a.$$

- Each of the above absolute precisions is supplied to **star**.
- The estimate of the standard deviation (obtained from the preliminary test) is supplied to **star** as the requested "maximum standard deviation".[4]
- The estimate of the autocorrelation of lag 1 (obtained from the preliminary test) is supplied to **star**.

star determines the sample size required to achieve each of the five absolute precisions; the sample sizes for ancillary parameters are those determined from their primary parameters.

D.2.2 **Failure Probability Parameters**

Five reasonable relative precisions are used:

P_{r1} = 20%, P_{r2} = 50%, P_{r3} = 100%, P_{r4} = 200%, and P_{r5} = 500%.

- Each of the above relative precisions is supplied to **star**.
- The conditional probability of a failure given that a failure occurred on the previous trial must be entered. There are two scenarios:
  - <u>More than one failure.</u> If more than one failure occurred during the preliminary characterization test, the conditional probability was estimated, and it is supplied to **star** to help determine the sample size.
  - <u>Zero or one failure.</u> If zero or one failure occurred, the estimate of the conditional probability is either indeterminate or zero, respectively, and the shell script must supply a value for this probability; the rather

[4]The estimated value should be supplied as the maximum value because it is reasonably well known from the preliminary characterization test.

conservative value of 0.8 is used, and its maximum value is either the value computed or 0.9, whichever is smaller.[5]

Star provides the number of failures required to achieve the five specified relative precisions.

D.3 Test Duration

After the sample size is determined, the duration of each test can be estimated.

Define

n′ = number of trials in the preliminary test, and

T′ = duration of the preliminary test as determined from the difference between the beginning and ending time.

D.3.1 Access-Disengagement Tests

It is assumed that the block size, the delay between access attempts, and specified performance values will be the same as during the preliminary test. The number of access attempts, n_i, will be as determined in D.2 (for i = 1, ... , 5 precisions).

A. Time Parameters

The estimated duration of an access-disengagement test of n_i access attempts required to achieve the absolute precision P_{ai} for the selected time parameter is

$$T_i = (n_i/n')T'.$$

B. Failure Probability Parameters

Let s′ denote the number of observed failures and p′ denote the estimate of the probability (for the most important failure probability parameter) during the preliminary test. Then $n' = s'/p'$. The estimated duration of an access-

[5]Otherwise, the computed maximum is often very large - such as 0.99.

disengagement test of n_i access attempts required to achieve the relative precision for the selected failure probability parameter depends upon s':

if s' > 0,

$$T_i = (n_i/n')T'$$

and

if s' = 0,

T_i cannot be estimated.

D.3.2 **User Information Transfer Test**

It is assumed that the block size, the delay between blocks, and the specified performance values will remain as during the preliminary test. The number of blocks, n_i, will be as determined in D.2 (for i = 1, ... ,5 precisions).

A. **Time Parameters**

If the most important parameter is a time parameter, the estimated duration of a user information transfer test is

$$T_i = (n_i/n')T' .$$

B. **Failure Probability Parameters**

Let n' = s'/p'. If the most important parameter is a failure probability parameter the estimated duration of a user information transfer test depends upon s':

- if s' > 0

  $$T_i = (n_i/n')T',$$

  and
- if s' = 0

  T_i can not be estimated.

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

	1. PUBLICATION NO.	2. Gov't Accession No.	3. Recipient's Accession No.
	95-319 (2)		

4. TITLE AND SUBTITLE
Performance Evaluation of Data Communication Services: NTIA Implementation of American National Standard X3.141, Volume 2. Experiment Design

5. Publication Date
August 1995

6. Performing Organization Code
NTIA/ITS.N3

7. AUTHOR(S)
Martin J. Miles

9. Project/Task/Work Unit No.

8. PERFORMING ORGANIZATION NAME AND ADDRESS
National Telecommunications and Information Admin.
Institute for Telecommunication Sciences
325 Broadway
Boulder, CO 80303

10. Contract/Grant No.

11. Sponsoring Organization Name and Address
National Telecommunications and Information Admin.
Herbert C. Hoover Building
14th and Constitution Avenue, NW
Washington, DC 20230

12. Type of Report and Period Covered

13.

14. SUPPLEMENTARY NOTES

15. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

This volume shows how to design an experiment to evaluate the performance of a data communication service as specified by ANS X3.141. It reviews some statistical concepts required for experiment design and analysis, including dependence between consecutive trials. It discusses objectives from both the user's and the vendor's point of view. It describes the criteria for selecting performance parameters and their appropriate analysis. It discusses the conditions that define the population of each function of a data communication service. It shows how to assign performance values (some of which cause excessive delays to be considered failures). It shows how to conduct a preliminary test to characterize the service. This test provides sample sizes required to achieve specified precisions, and it estimates the test duration for each. It discusses possible designs for the experiment and how to determine the sample size if a preliminary characterization test is either not conducted or if the objective of the experiment is to determine the acceptability of a performance parameter value.

Key words: American National Standards; analysis of variance; confidence level; data communication services; dependent trials; factors; performance measurements; performance parameters; precision; sample size

17. AVAILABILITY STATEMENT	18. Security Class. *(This report)*	20. Number of pages
☒ UNLIMITED.	Unclassified	124
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. *(This page)* Unclassified	21. Price:

*U.S. GOVERNMENT PRINTING OFFICE: 1995-676-490/25108