NTIA Report 95-319 (5)

Performance Evaluation of Data Communication Services: NTIA Implementation of American National Standard X3.141 Volume 5. Data Analysis

Martin J. Miles

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

Larry Irving, Assistant Secretary
for Communications and Information

August 1995

CONTENTS

FIGURES

FIGURES (Cont'd)

Figure B-8. Flowchart of subroutine **studnt**. 85

Figure C-1. Structured design diagram of analysis of single tests for time parameters. 88

FIGURES (Cont'd)

FIGURES (Cont'd)

TABLES

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

VOLUME 5. DATA ANALYSIS

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 analyze a performance parameter from a single test, and from multiple tests conducted at selected combinations of levels of variable conditions.

Single and multiple tests can analybe zed in any of four ways: estimation with known precision, acceptance tests, comparison tests, and tests to determine if a variable condition is a factor. Formulas for these analyses are provided. The formulas are incorporated in an interactive FORTRAN program that can be implemented by either a shell script or an operator. In all cases, dependence between trials is estimated by a first-order Markov chain. Performance parameters are analyzed from multiple tests by, first, pooling trials of tests, then means of tests, and, finally, means of levels of a selected variable condition.

Key Words: acceptance test, American National Standards, analysis of variance, comparison test, data communication systems, dependent trials, estimation, factors, performance measurements, precision

1. INTRODUCTION

This volume shows how to analyze the 24 American National Standard X3.102 (ANSI, 1992) performance parameters according to the methods specified by ANS X3.141 (ANSI, 1987).[2]

---

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

[2]The statistical theory for the NTIA implementation of ANS X3.141 comes from more than twenty special publications; it is organized in a book authored by M.J. Miles and E.L. Crow (to be published).

Table 1 is a list of the ANS X3.102 performance parameters organized according to communication functions. Primary performance parameters are shown in bold type, and ancillary time parameters are shown in italic type.

Table 1a lists the performance parameters according to the common performance criteria of speed, accuracy, and dependability.

On the other hand, for the purpose of analysis, Table 1b lists the performance parameters according to the type of random variable. Time parameters (i.e., delays and rates) might be asymptotically normally distributed, gamma distributed, or log normally distributed.[3] Failure probability parameters have a generalized binomial distribution (which considers dependence).

All analyses are applied to the ANS X3.102 performance parameters (only). Additional analysis could be applied to mathematical functions of one or more of the performance parameters. Such analysis could evaluate various data communication services, but it would be beyond ANS X3.141.

Although a great variety of analyses are available, the ANS X3.141 standard recommends one or more of four analyses:

- estimation with known precision,
- acceptance tests,
- comparison tests, and
- tests to determine if a variable condition is a factor.

The analyses selected from this list depend upon the objectives of the experiment. Some common objectives and their plausible analyses are listed in Table 2 and discussed in Section 2.3 of Volume 2. Some analyses, such as comparison tests, and tests to determine if a variable condition is a factor, utilize multiple tests only. Others, such as estimation with known precision, can utilize either a single test or multiple tests.

In all cases, dependence between trials is estimated by a first-order Markov chain. That is, dependence between trials is estimated by the autocorrelation of lag 1. In this model, the outcome of a trial is assumed to

---

[3]Volume 6 shows how to obtain histograms and box plots (i.e., abbreviated histograms) from single tests of primary time parameters.

Table 1. ANS X3.102 Performance Parameters

a. Organization by primary communication function and performance criterion

			PERFORMANCE CRITERIA		
			SPEED	ACCURACY	DEPENDABILITY
PRIMARY COMMUNICATION FUNCTIONS		ACCESS	• ACCESS TIME • *USER FRACTION OF ACCESS TIME*	• INCORRECT ACCESS PROBABILITY	• ACCESS DENIAL PROBABILITY • ACCESS OUTAGE PROBABILITY
	USER INFORMATION TRANSFER	BIT TRANSFER	——	• BIT ERROR PROBABILITY • BIT MISDELIVERY PROBABILITY • EXTRA BIT PROBABILITY	• BIT LOSS PROBABILITY
	USER INFORMATION TRANSFER	BLOCK TRANSFER	• BLOCK TRANSFER TIME • *USER FRACTION OF BLOCK TRANSFER TIME*	• BLOCK ERROR PROBABILITY • BLOCK MISDELIVERY PROBABILITY • EXTRA BLOCK PROBABILITY	• BLOCK LOSS PROBABILITY
	USER INFORMATION TRANSFER — TRANSFER SAMPLE TRANSFER	TRANSFER AVAIL-ABLILTY	——	• TRANSFER DENIAL PROBABILITY	
	USER INFORMATION TRANSFER — TRANSFER SAMPLE TRANSFER	THROUGHPUT	• USER INFORMATION BIT TRANSFER RATE • *USER FRACTION OF INPUT/OUTPUT TIME*	——	——
	DISENGAGEMENT	SOURCE DISENGAGEMENT	• SOURCE DISENGAGEMENT TIME • *USER FRACTION OF SOURCE DISENGAGEMENT TIME*	• SOURCE DISENGAGEMENT DENIAL PROBABILITY	
	DISENGAGEMENT	DESTINATION DISENGAGEMENT	• DESTINATION DISENGAGEMENT TIME • *USER FRACTION OF DESTIN-ATION DISENGAGEMENT TIME*	• DESTINATION DISENGAGEMENT DENIAL PROBABILITY	

b. Organization by primary communication function and random variable

			RANDOM VARIABLES		
			DELAY	RATE	FAILURE
PRIMARY COMMUNICATION FUNCTIONS		ACCESS	• ACCESS TIME	• *USER FRACTION OF ACCESS TIME*	• INCORRECT ACCESS • ACCESS OUTAGE • ACCESS DENIAL
	USER INFORMATION TRANSFER	BIT TRANSFER	——	——	• BIT ERROR • BIT MISDELIVERY • EXTRA BIT • BIT LOSS
	USER INFORMATION TRANSFER	BLOCK TRANSFER	• BLOCK TRANSFER TIME	• *USER FRACTION OF BLOCK TRANSFER TIME*	• BLOCK ERROR • BLOCK MISDELIVERY • EXTRA BLOCK • BLOCK LOSS
	USER INFORMATION TRANSFER — TRANSFER SAMPLE TRANSFER	TRANSFER AVAIL-ABLILTY	——	——	• TRANSFER DENIAL
	USER INFORMATION TRANSFER — TRANSFER SAMPLE TRANSFER	THROUGHPUT	——	• USER INFORMATION BIT TRANSFER RATE • *USER FRACTION OF INPUT/OUTPUT TIME*	——
	DISENGAGEMENT	SOURCE DISENGAGEMENT	• SOURCE DISENGAGEMENT TIME	• *USER FRACTION OF SOURCE DISENGAGEMENT TIME*	• SOURCE DISENGAGEMENT DENIAL
	DISENGAGEMENT	DESTINATION DISENGAGEMENT	• DESTINATION DISENGAGEMENT TIME	• *USER FRACTION OF DESTINATION DISENGAGEMENT TIME*	• DESTINATION DISENGAGEMENT DENIAL

Table 2. Common Experiment Objectives and Plausible Analyses

	PLAUSIBLE ANALYSIS			
EXPERIMENT OBJECTIVES	Estimation	Acceptance	Comparison	Factor
Acceptance/Maintenance		√		
Characterization	√			
Design/Management	√	√	√	√
Optimization				√
Selection			√	

be influenced only by the immediately preceding outcome.[4] Dependence does not affect the estimate of the mean, but it affects its precision; it usually increases the length of a confidence interval beyond that of independence (Appendix A, Equation A-1). If the trials are independent, the interval derived from the Markov model reduces to the classical interval that assumes independence. Chronological plots of trials of time parameters can reveal dependence; these plots can be generated by the methods described in Volume 6.

All four recommended analyses can be accomplished by **star**, an interactive computer program that is implemented by either a shell script or an operator (Miles, 1984).[5] In a given execution, **star** accomplishes one of the following tasks:

- Sample Size. It determines the minimum sample size (i.e., the fewest number of trials or failures) required to obtain a specified precision for estimation of a performance parameter. This procedure is described in Section 8 of Volume 2.

---

[4]In accordance with the standard, delays that are excessive are considered to be failures. In the NTIA implementation, delays having interleaved failures are considered to be consecutive delays - even though they are not. That is, delays surrounding one or more interleaved failures will be treated as if they are consecutive delays, and the autocorrelation will tend to be greater than it is. This error, however, is on the "safe side" because the precision will tend to be less than otherwise.

[5]A shell script is a file of UNIX[tm] commands, sometimes called a command file. The names of all files, directories, shell scripts, programs, and commands in this report are listed in bold type.

- Single Test Analysis. It estimates a performance parameter and its 90% or 95% confidence limits from a single test.

- Multiple Test Analysis. It analyzes a performance parameter from a set of multiple tests (each of which is conducted at a selected combination of levels of variable conditions). Analysis of multiple tests can be interpreted in various ways to accomplish each of the four recommended analyses.

star is written in ANSI FORTRAN 77 to enhance its portability. Results are written to the standard output device (typically the screen). **star** can be implemented by an operator. However, it should be executed by an operating system that supports I/O redirection because the NTIA software is designed so that either a shell script or an operator provides responses to prompts and access to performance data files.

This introduction discusses the analysis of single and multiple tests in a general manner. The four sections following this introduction show how single and multiple tests can be used to accomplish each of the four recommended analyses. The first four appendices apply to analysis of single tests. The next four appendices are analogous to the first four, but they apply to analysis of multiple tests.

1.1 Analysis of a Single Test

star analyzes performance data from a single test by estimating performance parameters and their 90% or 95% confidence limits. This is true of all performance parameters except two: since each test provides only one value for User Information Bit Transfer Rate and User Fraction of Input/Output Time, confidence limits cannot be estimated from single tests; analysis of these parameters requires multiple tests.

If a test results in zero failures or one failure, the estimates of the mean and the lower confidence limit are both zero. The upper confidence limit can be estimated unless there is an insufficient number of trials (causing an arithmetic computer error). The minimum number of trials required to compute the upper limit depends upon the selected confidence level and the conditional probability of a failure given that a failure occurred in the previous trial – a probability that must be estimated. The minimum number of trials is listed in Table 3 for a wide range of conditional probabilities. The numbers of trials

Table 3. Minimum Sample Sizes When the Number of Failures is Zero or One

MAXIMUM VALUE OF CONDITIONAL PROBABILITY	NUMBER OF FAILURES			
	0		1*	
	Confidence Level		Confidence Level	
	90%	95%	90%	95%
0.99	25	13	26	26
0.95	7	9	13 (14)	17 (19)
0.90	6	8	9 (11,12)	11 (15)
0.80	5	6	5 (8,9)	6 (11)
0,70	4	5	4 (7)	4 (9)
0.60	4	4	4 (6)	4 (7,8)
0.50	3	4	4 (5)	4 (5,6)
0.40	3	3	6	4 (5,6)
0.30	3	3	5	6
0.20	2	3	5	5
0.10	2	3	4	5
0	2	2	4	4

*Exclude the sample sizes in parenthesis; they are not acceptable.

shown in parentheses, although greater than the minimum number, also cause an arithmetic computer error.

Appendix A contains formulas required to estimate performance parameters and their confidence limits from single tests. Appendix B contains flowcharts of the subroutine structure of **star** (for analysis of a single test) and flowcharts of each subroutine required to analyze single tests.

1.1.1 Shell Script Implementation of Analysis of a Single Test

When a test is conducted, the performance data are extracted. They are then reduced and stored in text files (called performance outcome files). They are described in Section 4 of Volume 4 and listed here in Table 4.

Table 4. Performance Outcome Files

PERFORMANCE OUTCOMES	FILE NAMES
Access Outcome	ACO
Source Disengagement Outcome	D10
Destination Disengagement Outcome	D20
Bit Transfer Outcome	B10
Block Transfer Outcome	B20
Transfer Sample Outcome	B30
Throughput Sample Outcome	B40

The shell scripts listed in Table 5 implement analysis of single tests. They provide the performance outcome data and predetermined responses to **star**.

Implementation is initiated when the operator types **do** and its arguments during data reduction.

Table 5. Shell Scripts That Implement Single Test Analysis

	PERFORMANCE PARAMETER TYPE	
	Time	Failure Probability
Access-Disengagement	time-a	fail-a
User Information Transfer	time-x	fail-x

Figure 1 is a simple structured design diagram showing shell script implementation of analyses of a single test.[6] Appendix C describes these processes.

Figures 2 and 3 show sample analyses of time parameters for access-disengagement and user information transfer tests, respectively. Figures 4 and 5 show sample analyses of failure probability parameters for access-disengagement and user information transfer tests, respectively.

---

[6]The achieved precision may be more or less than the specified precision. If it is less, the number of additional trials or failures necessary to achieve the specified precision is not computed (a feature available from operator implementation). That is, for time parameters, the shell script assumes that the population variance and autocorrelation of lag 1 are known, and for failure probability parameters the shell script assumes the conditional probability (of a failure given a failure in the previous trial) is known.

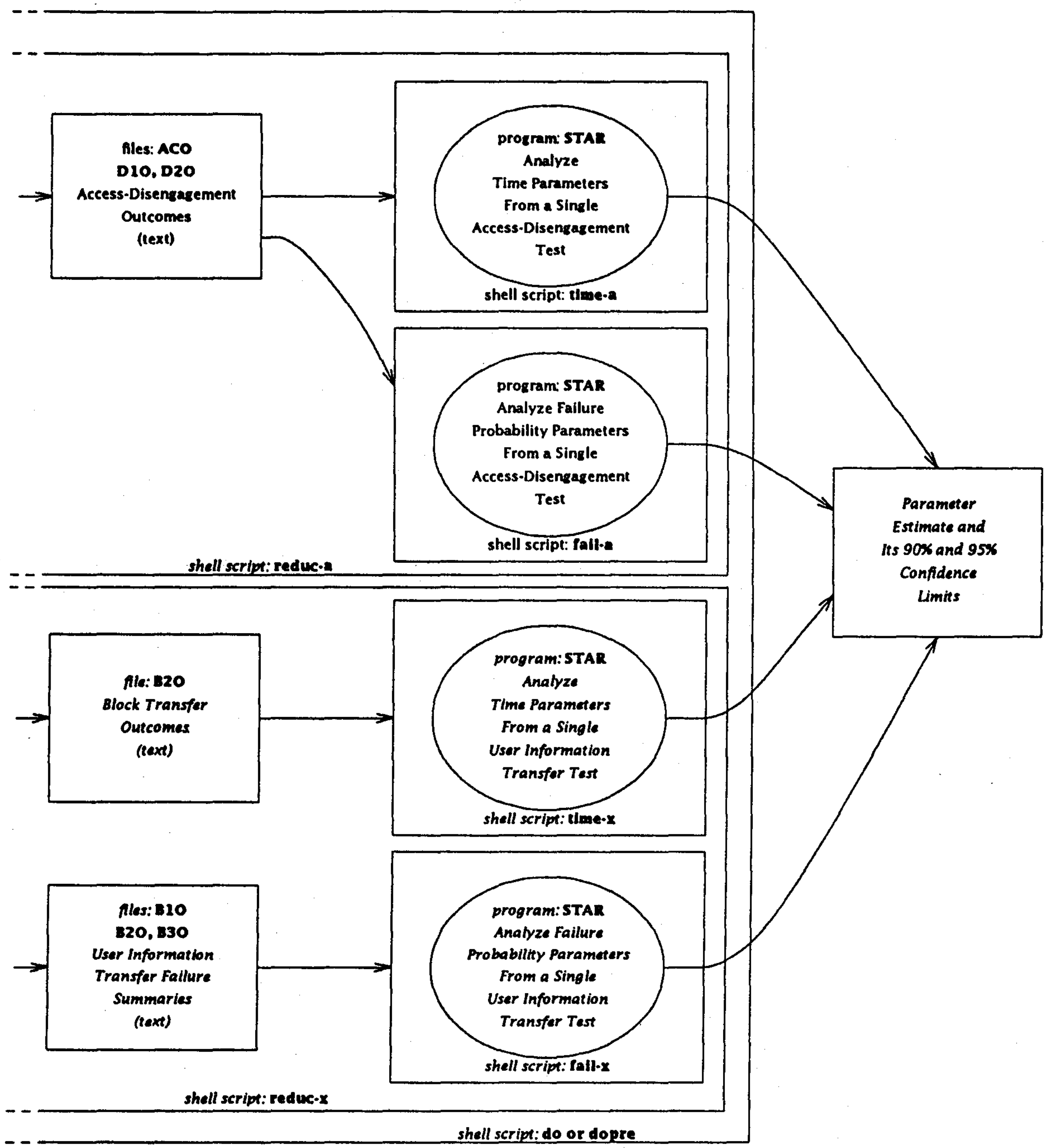

Figure 1. Structured design diagram of shell script implementation of analysis of a single test.

MEASUREMENT RESULTS SUMMARY

NTIA-ITS (Boulder) 2218

PERFORMANCE PARAMETER	SAMPLE SIZE	ESTIMATED VALUE	CONFIDENCE LEVEL (PERCENT)	LOWER CONFIDENCE LIMIT	UPPER CONFIDENCE LIMIT
ACCESS TIME	33	.47858E+02	90	.47333E+02	.48383E+02
			95	.47227E+02	.48490E+02
USER FRACTION OF ACCESS TIME	33	.26687E-02	90	.18527E-02	.34847E-02
			95	.16964E-02	.36409E-02
DISENGAGEMENT TIME (SOURCE)	28	.89793E+01	90	.87887E+01	.91698E+01
			95	.87498E+01	.92087E+01
USER FRACTION OF DISENGAGEMENT TIME (SOURCE)	28	.87136E-02	90	.85151E-02	.89122E-02
			95	.84770E-02	.89503E-02
DISENGAGEMENT TIME (DESTINATION)	33	.86597E+00	90	.81650E+00	.91544E+00
			95	.80648E+00	.92546E+00
USER FRACTION OF DISENGAGEMENT TIME (DESTINATION)	33	.70513E-01	90	.66392E-01	.74634E-01
			95	.65602E-01	.75423E-01

ESTIMATED PERFORMANCE TIMES ARE EXPRESSED IN SECONDS

Figure 2. Example from shell script implementation of analysis of time parameters from an access-disengagement test.

MEASUREMENT RESULTS SUMMARY

NTIA-ITS (Boulder) 2215

PERFORMANCE PARAMETER	SAMPLE SIZE	ESTIMATED VALUE	CONFIDENCE LEVEL (PERCENT)	LOWER CONFIDENCE LIMIT	UPPER CONFIDENCE LIMIT
BLOCK TRANSFER TIME	40	.55582E+01	90	.54810E+01	.56355E+01
			95	.54655E+01	.56510E+01
USER FRACTION OF BLOCK TRANSFER TIME	40	.29819E-02	90	.28642E-02	.30996E-02
			95	.28416E-02	.31222E-02

ESTIMATED PERFORMANCE TIMES ARE EXPRESSED IN SECONDS

Figure 3. Example from shell script implementation of analysis of time parameters from a user information transfer test.

MEASUREMENT RESULTS SUMMARY

NTIA-ITS (Boulder) 2218

PERFORMANCE PARAMETER	SAMPLE SIZE	ESTIMATED VALUE	CONFIDENCE LEVEL (PERCENT)	LOWER CONFIDENCE LIMIT	UPPER CONFIDENCE LIMIT
INCORRECT ACCESS PROBABILITY	36	.00000E+00	90	.00000E+00	.22162E+00
			95	.00000E+00	.26944E+00
ACCESS DENIAL PROBABILITY	36	.83333E-01	90	.28145E-01	.19488E+00
			95	.23488E-01	.21752E+00
ACCESS OUTAGE PROBABILITY	36	.00000E+00	90	.00000E+00	.22162E+00
			95	.00000E+00	.26944E+00
DISENGAGEMENT DENIAL PROBABILITY (SOURCE)	33	.15152E+00	90	.74809E-01	.27216E+00
			95	.66285E-01	.29573E+00
DISENGAGEMENT DENIAL PROBABILITY	33	.00000E+00	90	.00000E+00	.23544E+00
			95	.00000E+00	.28512E+00

WHEN THE OBSERVED NUMBER OF FAILURES IS 0 OR 1, THE CONDITIONAL PROBABILITY OF FAILURE USED TO ESTIMATE CONFIDENCE LIMITS IS 0.8

Figure 4. Example from shell script implementation of analysis of failure probability parameters from an access-disengagement test.

MEASUREMENT RESULTS SUMMARY

NTIA-ITS (Boulder) 2215

PERFORMANCE PARAMETER	SAMPLE SIZE	ESTIMATED VALUE	CONFIDENCE LEVEL (PERCENT)	LOWER CONFIDENCE LIMIT	UPPER CONFIDENCE LIMIT
BIT ERROR PROBABILITY	163840	.00000E+00	90	.00000E+00	.70329E-04
			95	.00000E+00	.91485E-04
BIT LOSS PROBABILITY	163840	.00000E+00	90	.00000E+00	.70329E-04
			95	.00000E+00	.91485E-04
EXTRA BIT PROBABILITY	163840	.00000E+00	90	.00000E+00	.70329E-04
			95	.00000E+00	.91485E-04
BLOCK ERROR PROBABILITY	40	.00000E+00	90	.00000E+00	.20557E+00
			95	.00000E+00	.25107E+00
BLOCK LOSS PROBABILITY	40	.00000E+00	90	.00000E+00	.20557E+00
			95	.00000E+00	.25107E+00
EXTRA BLOCK PROBABILITY	40	.00000E+00	90	.00000E+00	.20557E+00
			95	.00000E+00	.25107E+00
TRANSFER DENIAL PROBABILITY	4	.00000E+00	90	.00000E+00	.63621E+00
			95	.00000E+00	.69390E+00

WHEN THE OBSERVED NUMBER OF FAILURES IS 0 OR 1, THE CONDITIONAL PROBABILITY OF FAILURE USED TO ESTIMATE CONFIDENCE LIMITS IS 0.8

Figure 5. **Example from shell script implementation of analysis of failure probability parameters from a user information transfer test.**

1.1.2 Operator Implementation of Analysis of a Single Test

Performance parameters can be analyzed by operator implementation as well as by shell script implementation. Moreover, if the experimenter had insufficient knowledge of the population prior to the test, he/she was instructed to observe a certain number of trials preliminarily to obtain sufficient knowledge. **star** now determines the number of additional observations, if any, that is required to obtain the specified precision.[7]

Figure 6 is a structured design diagram of operator implementation. Although **star** can analyze rates, this procedure is not described in the figure because NTIA procedures provide only one rate trial per test. The acceptable modes of data entry for operator implementation of analysis of a single test are summarized in Table 6 for each type of random variable.

star can be executed from **/usr/data/5a** by typing

star

and the requested responses. Appendix D shows in detail how an operator implements **star** to analyze a performance parameter from a single test.

Table 6. Acceptable Modes of Performance Data Entry for Operator Implementation of Analysis of a Single Test

	DELAY	RATE	FAILURE
File	√	√	-
Keyboard	√	√	√

[7]This feature is not enabled for shell script implementation of **star**.

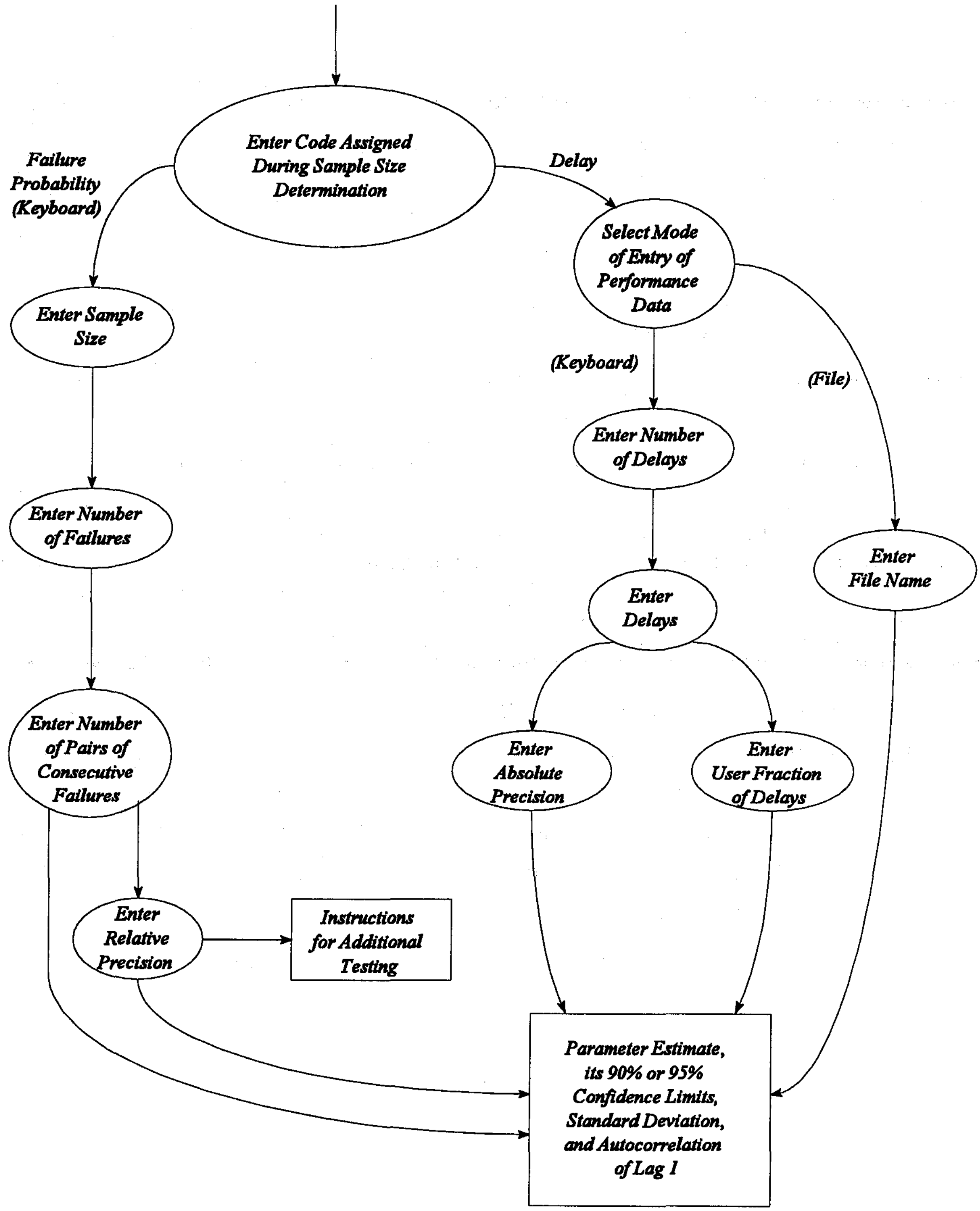

Figure 6. Structured design diagram of operator implementation of analysis of a single test.

A. **Performance Data for Time Parameters**

Performance data for delays can be entered by a file or the keyboard.

For file entry, create a file such that each line contains performance data for a trial: performance time, user portion of performance time, and number of bits (if rates are to be analyzed). The format is

2F16.3, F8.0

After performance data for the last trial are entered, create the last line with -30. in the first field.

For keyboard entry, enter the number of delays, enter the performance times for all trials, then enter either the user portions of performance times for all trials or the absolute precision. The absolute precision is to be entered only if the variance and/or the autocorrelation of lag 1 was not known when the sample size was determined.

If the number of delays is insufficient to obtain the specified precision, **star** computes the number of additional delays that must be observed.

Figure 7 is an example of output from operator implementation of analysis of a test of Access Time. It shows estimates of Access Time, User Fraction of Access Time, and their 95% confidence limits. The estimates of standard deviation and autocorrelation of lag 1 for the primary parameters are also shown to indicate the dispersion and dependence, respectively.

```
YOUR TEST OF     4 TRIALS RESULTED IN AN ESTIMATED MEAN DELAY OF
.41750E+02 .  YOU CAN BE 95 PERCENT CONFIDENT THAT THE
TRUE MEAN DELAY IS BETWEEN .39529E+02 AND .43971E+02 .
 YOUR TEST RESULTED IN AN ESTIMATED MEAN
USER-RESPONSIBLE FRACTIONAL DELAY OF .12465E-01. YOU CAN BE
95 PERCENT CONFIDENT THAT THE TRUE MEAN IS BETWEEN .12465E-01
AND .14595E-01.
(THE ESTIMATE OF THE STANDARD DEVIATION IS .40311E+01, AND
THE ESTIMATE OF THE AUTOCORRELATION OF LAG 1 IS  -.78590E+00.)
```

Figure 7. Example from operator implementation of analysis of a single test of time parameters.

B. Performance Data for Failure Probability Parameters

Performance data for failure probability parameters must be entered from the keyboard. If the number of failures is insufficient, **star** indicates the number of additional failures that must be observed to obtain the specified precision.

Sequentially enter the number of trials, number of failures, number of pairs of consecutive failures, and, possibly, the relative precision (if the conditional probability of a failure, given a failure on the previous trial was unknown when the sample size was determined).

Figure 8 is an example of output from operator implementation of analysis of Access Denial Probability. It shows the estimate of the probability of a failure, its confidence limits, the conditional probability of a failure (given that a failure occurred in the previous trial) and the autocorrelation of lag 1.

1.2 Analysis of Multiple Tests

The combinations of levels of variable conditions should have been selected during experiment design so that all analysis objectives can be achieved.

star analyzes performance parameters from a set of tests conducted at selected combinations of levels of the variable conditions by pooling the data. The tests selected for analysis of multiple tests should have standard deviations that are somewhat similar. Analysis is accomplished by tests of hypotheses at the α = 5% significance level.

1.2.1 Pooling Data from Multiple Tests

Analysis of multiple tests is based on a linear model for the analysis of variance. This model assumes that

- there are three additive components of variation (variation among trials within a test, variation among tests within a level of a selected variable condition, and variation among levels of the variable condition),
- the levels of the variable conditions have been chosen randomly from a set of all possible levels,

```
YOUR TEST OF      860 TRIALS RESULTED IN AN ESTIMATED FAILURE PROBABILITY
OF .34884E-02.  YOU CAN BE 90% CONFIDENT THAT THE TRUE
FAILURE PROBABILITY IS BETWEEN .37315E-03 AND .12230E-01.
(THE CONDITIONAL PROBABILITY OF A FAILURE GIVEN THAT A
FAILURE OCCURRED IN THE PREVIOUS TRIAL IS .33372E+00 AND
THE AUTOCORRELATION OF LAG 1 IS   .33139E+00.)
```

Figure 8. Example from operator implementation of analysis of a single test of a failure probability parameter.

- the tests performed at a given level result in a random sample of all possible tests for that level, and

- dependence among trials in each test is estimated by a first-order Markov chain.

Figure 9 is a flowchart of the scheme for pooling data from multiple tests. **star** analyzes performance parameters from poolings of

- trials from the tests,

- means of the tests, and

- means of the levels of a selected variable condition.

The acceptability of pooling for a delay parameter is determined from (total) performance times, not user performance times. However, **star** estimates both the primary delay parameter and its user fraction of the delay. The acceptability of pooling for the User Information Bit Transfer Rate and the User Fraction of Input/Output Time is determined by their common denominator, Input/Output Time. The acceptability of pooling for a failure probability parameter is determined jointly from observed failures and observed pairs of consecutive failures.

A. Pooling Trials of Tests

To determine if there is a significant difference among test means, **star** tests the null hypothesis that all test means are equal. Specifically, **star** evaluates a statistic that depends on the dispersion of means of the tests about the mean of all trials and has a known distribution under the assumptions of the model. **star** then determines the 5% point of the statistic's distribution.[8] The null hypothesis is accepted if the value of the statistic is less than the 5% point, and it is rejected otherwise. <u>If the null hypothesis of equal means of tests is accepted, trials from all tests are considered to come from the same population and are pooled.</u> **star** estimates the mean of all trials and its confidence limits. This pooling provides the most precision of the three poolings.

---

[8]The 5% point is the value of the abscissa from which 5% of the density is to the right.

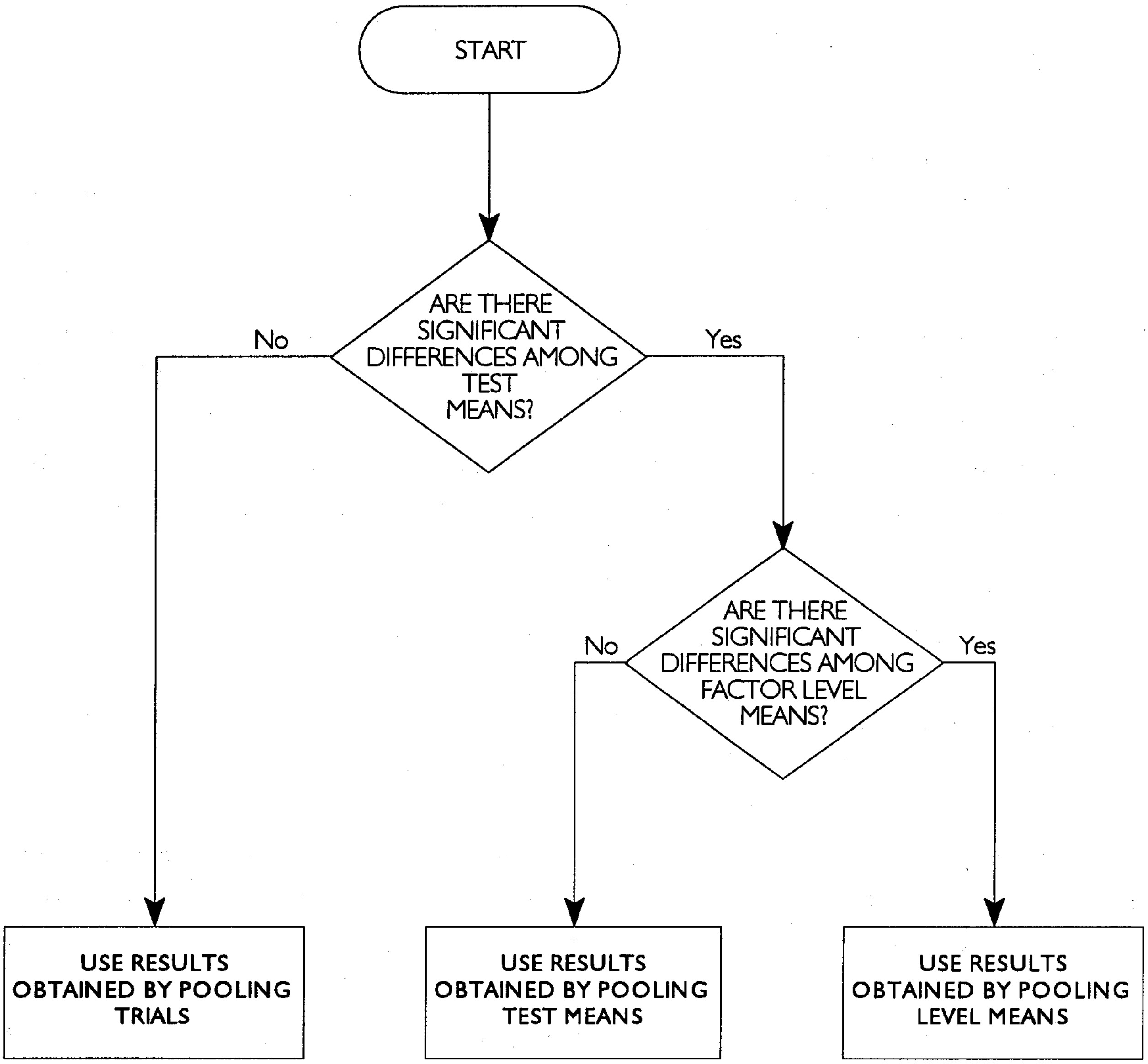

Figure 9. Flowchart of pooling procedure for analysis of multiple tests.

The variable condition selected does not affect the acceptance of the null hypothesis (that test means are equal) nor the estimates of the mean of the trials and its confidence limits.

B. **Pooling Means of Tests**

To determine if there is a significant difference among the means of tests, **star** tests the null hypothesis that the means of all levels of a variable condition are equal. In this case, the program evaluates a statistic that depends on the dispersion of the means of the levels of the variable condition about the mean of all test means. The hypothesis is accepted if the value of the statistic is less than the 5% point of the distribution, and rejected otherwise. If the null hypothesis of equal means of levels is accepted, means from all tests are considered to come from the same population and are pooled. **star** estimates the mean of test means and its confidence limits. This pooling provides less precision than pooling trials of tests.

The variable condition that is selected affects the acceptance of the null hypothesis (that level means are equal) but not the estimates of the mean of test means and its confidence limits.

C. **Pooling Means of Levels**

The means of each level of the selected variable condition are pooled. There is no null hypothesis for this pooling; it is, simply, done. **star** estimates the mean of level means and its confidence limits. This pooling provides the least precision of the three poolings.

The variable condition that is selected affects the estimates of the mean of the level means and its confidence limits.

Appendix E contains formulas for analysis of multiple tests.

Analysis of multiple tests by **star** can be implemented by either a shell script or an operator.

1.2.2 Shell Script Implementation of Analysis of Multiple Tests

When a test is conducted, the performance outcome files (listed in Table 4) are used by shell scripts (listed in Table 5) to produce performance data files for multiple tests. The file **log** is also appended.[9] It contains the test number, date, time, source site, line speed, type of test, number of access attempts, number of block transfers, block size, and intertrial delay (for both access attempts and block transfer attempts).[10] At the conclusion of the experiment, this file should be edited to remove information from flawed tests.

Program **qklog** uses the file **log** to produce the files **log.acc** and **log.xfr**, which contain access-disengagement and user information test identification, respectively.

To create these files type

qklog

Table 7 lists the variable conditions, the number i = 1, 2, ..., N that corresponds to the order of the N variable conditions, and the name of the file or shell script that contains or computes their levels. Levels of two additional variable conditions can be entered in the command line of **runxt**; they are indicated by A_7 and A_8 for access-disengagement tests and by U_8 and U_9 for user information transfer tests.

[9]This is done by program **mklog** and implemented by shell script **runxt** at the source end user site.

[10]The block sizes in **log.xfr** are followed by the character **b** (for bytes) to distinguish this character string from others containing numbers only (such as the test number) when using the UNIX[tm] **grep** utility.

Table 7. Conditions Assumed to be Variable

VARIABLE CONDITION	CODE	FILE/SHELL SCRIPT WHERE LEVELS ARE CONTAINED/COMPUTED
Source Site	1	default netcodes
Network	2	runxt netcodes
Day of Week	3	(computed by runxt)
Time of Day	4	(computed by runxt)
Intersession Delay	5	default
Destination Site	6	default netcodes
A_7	7	runxt
A_8	8	runxt

a. Access-Disengagement Tests

VARIABLE CONDITION	CODE	FILE/SHELL SCRIPT WHERE LEVELS ARE CONTAINED/COMPUTED
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

b. User Information Transfer Tests

Tests selected for pooling are copied from either **log.acc** or **log.xfr** to **log.wrk**. For example, if certain access-disengagement tests are selected because they have a common level, say **xxx**, identification lines of these tests can be copied from **log.acc** to **log.wrk** by typing

grep xxx log.acc > log.wrk

Three shell scripts can implement **star** by providing a performance parameter identification, the number of the variable condition (i.e., i = 1, 2, ..., N) selected for testing the null hypothesis that the means of its levels are equal, and the file **log.wrk**. These shell scripts correspond to the three types of performance parameters and are called **delay**, **rate**, and **fail**. Figure 10 is a structured design diagram of shell script implementation of analysis of multiple tests. Each underscore character in the argument of each shell script command represents a character of the performance parameter identification.

The following three subsections show what must be done to implement **star** by a shell script for delay, rate, and failure probability parameters; Appendix G describes how it is done.

A. Delay Parameters

Table 8 lists the time parameters and the commands to implement shell script analysis of multiple tests. The **i** in the command is the ordinal number of the variable condition for which the means of the levels should be pooled; this is the code number in Table 7. After one of these commands is entered, **star** analyzes the selected delay parameter, whose tests were conducted at the selected combination of levels of variable conditions.

Example: Figure 11 is a sample output of analysis of Access Time and User Fraction of Access Time for 11 tests. Source Site (i.e., i = 1) is the variable condition selected for testing the hypothesis that the means of levels (Fort Worth, Seattle, and Washington D.C.) are equal. The identifying lines of 11 tests were copied from **log.acc** to the file called **log.wrk**, and the command

delay ac 1

was typed.

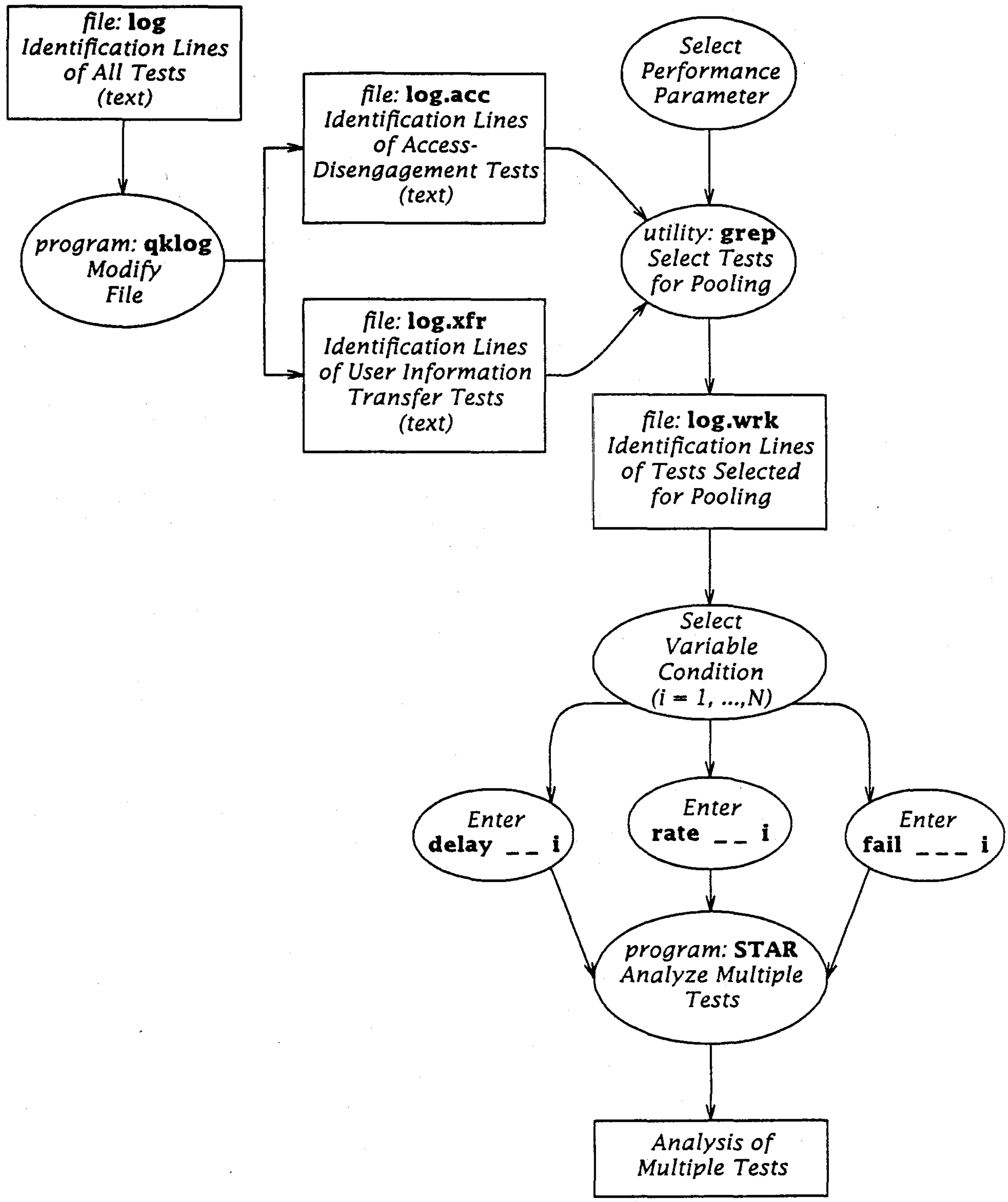

Figure 10. Structured design diagram of operator procedure to analyze multiple tests by shell scripts.

Table 8. Commands for Shell Script Implementation of Analysis Multiple Tests of Time Parameters

PERFORMANCE PARAMETER	COMMAND
Access Time & User Fraction of Access Time	delay ac i
Block Transfer Time & User Fraction of Block Transfer Time	delay b2 i
User Information Bit Transfer Rate & User Fraction of Input/Output Time	rate b4 i
Source Disengagement Time & User Fraction of Source Disengagement	delay d1 i
Destination Disengagement Time & User Fraction of Destination Disengagement Time	delay d2 i

star lists the following data:

- Single Test Data. For each of the 11 tests, **star** lists the test number, the level of each of the six variable conditions, and the number of trials. It also lists the estimate of the mean and the standard deviation for both the primary parameter and its user fraction.

- Quantities that Determine the Degrees of Freedom. It lists the number of trials, the number of tests, and the number of levels of the selected variable condition.

- Autocorrelations. It lists two autocorrelations of lag 1. They are:

  - Weighted average of the autocorrelations of the tests (i.e., the autocorrelation for each test, weighted by the number of its trials). This average modifies the degrees of freedom of the F distribution (i.e., the 5% point) and the value of the F statistic, both of which are used for the hypothesis test: Positive autocorrelation decreases both the 5% point and the F statistic.

  - Average autocorrelation of the trials. This average modifies the degrees of freedom of the Student t distribution for computation of the confidence limits of the pooled data: Positive autocorrelation increases the length of the confidence interval.

Analysis of Multiple Tests

Access Time
Variable Condition 1

Thu Jan 26 14:06:14 MST 1989

Test	--- Variable Conditions ---	Trials	---- Times ---- Mean	Std Dev	User Fractions Mean	Std Dev
775	ftw netA fri 1 A55 bol	20	38.291	1.608	0.0397	0.0199
823	sea netA fri 2 A55 bol	20	42.439	1.527	0.0339	0.0047
815	sea netA fri 6 A55 bol	20	41.576	1.269	0.0352	0.0044
835	sea netA mon 3 A55 bol	15	42.954	1.325	0.0345	0.0053
858	sea netA thu 1 A55 bol	20	42.284	1.338	0.0345	0.0053
876	sea netA thu 4 A55 bol	20	42.313	2.197	0.0350	0.0064
811	sea netA thu 5 A55 bol	19	41.163	1.015	0.0373	0.0065
997	wdc netA thu 3 A55 bol	17	41.751	2.198	0.0356	0.0075
928	wdc netA tue 1 A55 bol	20	44.500	4.380	0.0332	0.0068
952	wdc netA tue 5 A55 bol	20	39.813	1.625	0.0368	0.0043
978	wdc netA wed 4 A55 bol	18	42.304	1.820	0.0351	0.0054

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 209
NUMBER OF TESTS = 11
NUMBER OF LEVELS = 3

WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 11 TESTS = .3927E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 209 TRIALS = .4998E+00 @

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	80	10	.4961E+01	.1963E+01	W	.4112E+02	.4173E+02	.4234E+02	
	-	-	-	-	V	.3423E-01	.3547E-01	.3671E-01	
AMONG TESTS	8	2	.4011E+01	.4460E+01	W	.4067E+02	.4176E+02	.4286E+02	*
	-	-	-	-	V	.3416E-01	.3551E-01	.3686E-01	
AMONG LEVELS	-	-	-	-	W	.3536E+02	.4083E+02	.4631E+02	
	-	-	-	-	V	.2953E-01	.3676E-01	.4417E-01	

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 11. Example of analysis of Access Time and User Fraction of Access Time from multiple access tests.

- Results of Poolings. For each of the three poolings (i.e., trials of tests, means of tests, and means of levels), **star** lists the effective degrees of freedom, the value of the F statistic, and value of the 5% point of the F distribution. Then for the delay (**W**) and the user fraction of delay (**V**), it lists estimates of the lower 95% confidence limit, the mean, and the upper 95% confidence limit. The * to the right of a delay row indicates that pooling is acceptable - for both **W** and V. In this example, the F statistic (4.961) is greater than the 5% point of the F distribution (1.963); hence, pooling the trials is not acceptable. However, the means of tests can be pooled since this F statistic (4.011) is less than this 5% point (4.460).

B. Rate Parameters

Table 8 lists the time parameters and the commands to implement shell script analysis of multiple tests. The **i** in the command is the ordinal number of the variable condition for which the means of the levels should be pooled; this is the code number in Table 7. After one of these commands is entered, **star** analyzes the selected rate parameters, whose tests were conducted at the selected combination of levels of variable conditions.

The **rate** shell script implements analysis of two parameters: The primary parameter, User Information Bit Transfer Rate, and its ancillary parameter, User Fraction of Input/Output Time.

Example: Figure 12 is an example of analysis of the User Information Bit Transfer Rate and User Fraction of Input/Output Time, respectively, using Day of the Week (i.e., i = 3) as the variable condition to be tested. The identifying lines of the five selected tests were copied from **log.xfr** to **log.wrk**, and the command

rate b4 3

was typed.

This output is fundamentally different from that of delays and failure probabilities because **star** is implemented twice. For both implementations, **W** is the Input/Output Time (capitalized because of its importance, but it is not a performance parameter). For the first implementation, **V** is the User Information Bit Transfer Rate, and for the second implementation, **V** is the User Fraction of Input/Output Time.

ANALYSIS OF MULTIPLE TESTS

User Information Bit Transfer Rate
Variable Condition 3

Mon Jan 30 08:58:34 MST 1989

Test	---- Variable Conditions ----	Trials	Time	Rate
998	wdc netA thu 3 B00 bol 128	1	111.796	723.6
1060	den netA thu 4 B00 bol 128	1	88.822	910.8
950	wdc netA tue 5 B00 bol 128	1	104.310	775.5
1025	den netA wed 2 B00 bol 128	1	89.167	907.2
976	wdc netA wed 4 B00 bol 128	1	120.935	668.9

TRANSFER TIMES (W) AND BIT TRANSFER RATES (V)

NUMBER OF TRIALS = 5
NUMBER OF TESTS = 5
NUMBER OF LEVELS = 3

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT

THE DEGREES OF FREEDOM IS NOT POSITIVE, THEREFORE THE SIGNIFICANCE TEST FOR VARIATION AMONG TRIALS CANNOT BE PERFORMED

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT
AMONG TESTS *	2	2	.3203E-01	.1900E+02	W	.8552E+02	.1030E+03	.1205E+03
	-	-	-	-	V	.6621E+03	.7972E+03	.9323E+03
AMONG LEVELS	-	-	-	-	W	.9689E+02	.1032E+03	.1096E+03
	-	-	-	-	V	.7405E+03	.7936E+03	.8467E+03

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICATES THIS POOLING IS ACCEPTABLE AT THE 5% LEVEL.

Figure 12. **(Part 1). Example of analysis of User Information Bit Transfer Rate and User Fraction of Input/Output Time from multiple tests.**

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 5
NUMBER OF TESTS = 5
NUMBER OF LEVELS = 3

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT
THE DEGREES OF FREEDOM IS NOT POSITIVE, THEREFORE THE SIGNIFICANCE TEST FOR VARIATION AMONG TRIALS CANNOT BE PERFORMED								
AMONG TESTS	2	2	.3203E-01	.1900E+02	W	.8552E+02	.1030E+03	.1205E+03
*								
	-	-	-	-	V	.1738E+00	.2093E+00	.2448E+00
AMONG LEVELS	-	-	-	-	W	.9689E+02	.1032E+03	.1096E+03
	-	-	-	-	V	.1946E+00	.2084E+00	.2221E+00

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICATES THIS POOLING IS ACCEPTABLE AT THE 5% LEVEL.

Figure 12. **(Part 2). Example of analysis of User Information Bit Transfer Rate and User Fraction of Input/Output Time from multiple tests.**

Analysis consists of the following:

- Single Test Data. For each test, **star** lists the identification (i.e., test number and the level of each of the seven variable conditions). Since there is but one trial (i.e., one transfer) per test, the standard deviation cannot be estimated. Instead, the User Fraction of Input/Output Time and User Information Bit Transfer Rate for each trial is listed.

- Quantities that Determine the Degrees of Freedom. **star** lists the number of trials, the number of tests, and the number of levels of the variable condition (Day of the Week).

- Autocorrelations. Since there is but one trial per test, the weighted average of the autocorrelations of the tests would equal the average autocorrelation of the trials. However, neither is computed because they are assumed to be zero.

- Results of Poolings. Since there is but one trial per test, pooling among trials is equivalent to pooling among tests - and it is arbitrarily labelled **among tests**. To attempt pooling among tests, the F statistic of the Input/Output Time is compared with the 5% percentage point of the F distribution. Since it is less (0.0320 compared with 19.00), the means of the tests can be pooled.[11] The * to the left of the Input/Output Time row (W) in both parts of Figure 12 indicates that pooling among tests is acceptable for both performance parameters. Finally, the means of the levels are pooled. There is no hypothesis test for this pooling, it is simply done.

C. Failure Parameters

Table 9 lists the failure probability performance parameters and the command to implement shell script analysis of multiple tests. Each parameter is identified by a three letter code. The **i** in the command is the ordinal number of the variable condition for which the null hypothesis of equal means of levels is to be tested. After one of these commands is entered, the program analyzes the selected performance parameter.

---

[11]In the case of rates, there is one trial per test; therefore the mean is simply the value of the single trial.

Table 9. Commands for Shell Script Implementation of Analysis of Multiple Tests of Failure Probability Parameters

PERFORMANCE PARAMETER	COMMAND
Incorrect Access Probability	fail acm 1
Access Outage Probability	fail aco 1
Access Denial Probability	fail acl 1
Bit Error Probability	fail b1e 1
Extra Bit Probability	fail b1x 1
Bit Loss Probability	fail b1l 1
Block Error Probability	fail b2e 1
Extra Block Probability	fail b2x 1
Block Loss Probability	fail b2l 1
Transfer Denial Probability	fail b3l 1
Source Disengagement Denial Probability	fail d1l 1
Destination Disengagement Denial Probability	fail d2l 1

Figure 13 is a sample output of analysis of Source Disengagement Denial Probability for 11 tests.[12] The identifying lines of 11 tests were copied from **log.acc** to **log.wrk**, and the command

fail d1l 1

was typed.

Analysis consists of the following:

- Single Test Data. For each of the 11 tests, **star** lists the test number, the levels of each of the six variable conditions, the number of trials, the number of failures, the number of pairs of consecutive failures, and the estimate of the mean (i.e., proportion).

- Quantities that Determine the Degrees of Freedom. It then lists the number of trials, the number of tests, and the number of levels of the selected variable condition (i.e., Fort Worth, Seattle, and Washington, D.C.).

- Autocorrelation. The weighted average of the autocorrelation of lag 1 is not used for analysis of multiple tests of failure probabilities as it is for time parameters. **star** lists the average autocorrelation of lag 1 for the 209 pooled trials. Its value of -0.002 suggests that the trials are essentially uncorrelated.

[12]They are the same tests selected for pooling delays.

ANALYSIS OF MULTIPLE TESTS

Source Disengagement Denial Probability
Variable Condition 1

Wed Feb 8 16:30:00 MST 1989

Test	------ Variable Conditions ------	Trials	Failures	Pairs	Prob
775	ftw netA fri 1 A55 bol	20	1	0	0.050
823	sea netA fri 2 A55 bol	20	1	0	0.050
815	sea netA fri 6 A55 bol	20	3	1	0.150
835	sea netA mon 3 A55 bol	15	0	0	0.000
858	sea netA thu 1 A55 bol	20	3	0	0.150
876	sea netA thu 4 A55 bol	20	1	0	0.050
811	sea netA thu 5 A55 bol	19	0	0	0.000
997	wdc netA thu 3 A55 bol	17	1	0	0.059
928	wdc netA tue 1 A55 bol	20	1	0	0.050
952	wdc netA tue 5 A55 bol	20	3	0	0.150
978	wdc netA wed 4 A55 bol	18	1	0	0.056

FAILURE PROBABILITY

NUMBER OF TRIALS = 209
NUMBER OF TESTS = 11
NUMBER OF LEVELS = 3

AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 209 TRIALS =-.002 @

	DEGREES OF FREEDOM		X2 STAT.	X2 DIST.(5%)	95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	10	-	.8849E+01	.1831E+02	.41571E-01	.71770E-01	.11642E+00	*
			F STAT.	F DIST (5%)				
AMONG TESTS	8	2	.2018E+00	.4460E+01	.4526E-01	.7906E-01	.1212E+00	*
AMONG LEVELS	-	-	-	-	.4624E-01	.7808E-01	.1174E+00	

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 13. Example of analysis of Source Disengagement Time from multiple tests.

- Results of Poolings. The criteria for pooling differs for trials and test means:

  - Trials of Tests. The pooling of all trials is attempted first. In this case, two chi-squared statistics are computed from the performance data (one for the probability of a failure and one for the conditional probability of a failure, given that a failure occurred during the previous trial). Their values are compared with the values of their respective chi-squared distributions at the 5% points. This output lists the chi-squared values from the probability of a failure only. In this example, the trials can be pooled (as indicated by the * in the **among trials** row). Even though the preferred pooling (i.e., among trials) is acceptable, the program continues to test the pooling of means of the tests.

  - Means of Tests. Acceptability of pooling test means of failure probability parameters is determined by the F test. Whereas autocorrelation must be regarded in the pooling of trials, it is not important when pooling proportions among tests. For failure probabilities, a transformation of the proportions is required (see Appendix E). The F statistic is developed from the transformed proportions, and the value of this statistic is compared with the value of the F distribution at its 5% point. The proportions of the levels are then pooled. There is no hypothesis test for the acceptability of this pooling.

1.2.3 Operator Implementation of Analysis from Multiple Tests

An operator can analyze multiple tests.

From /usr/data/5a, type

star

and supply the responses to the prompts. These responses are the code 40 (indicating analysis of multiple tests), the type of parameter, the confidence level (only 95% is available from the NTIA analysis of multiple tests), the number of levels, and the number of tests at each level.

The mode of entry of performance data is different for time and failure probability parameters. Table 10 shows the acceptable mode of entry of performance data, and Figure 14 is a structured design diagram describing operator implementation.

Table 10. Acceptable Mode of Performance Data Entry for Operator Implementation of Analysis of Multiple Tests

	DELAY	RATE	FAILURE
File	√	√	-
Keyboard	-	-	√

A. Performance Data for Time Parameters

Because keyboard entry of performance data of delays and rates is inefficient, **star** provides only file entry of performance data for time parameters.[13] This file must have the same format as that required of analysis of single tests, including the end of file indicator, -30.

Create the file of performance data for time parameters.

B. Performance Data for Failure Probability Parameters

Performance data of failure probability parameters from multiple tests can be entered by keyboard only.

Enter the performance data for the failure probability parameters.

Appendix H shows details of how analysis of multiple tests is implemented by an operator.

[13]NTIA procedures allow only one trial per test of User Information Bit Transfer Rate and User Fraction of Input/Output Time, so performance data collected by procedures such as this cannot be analyzed by operator implementation.

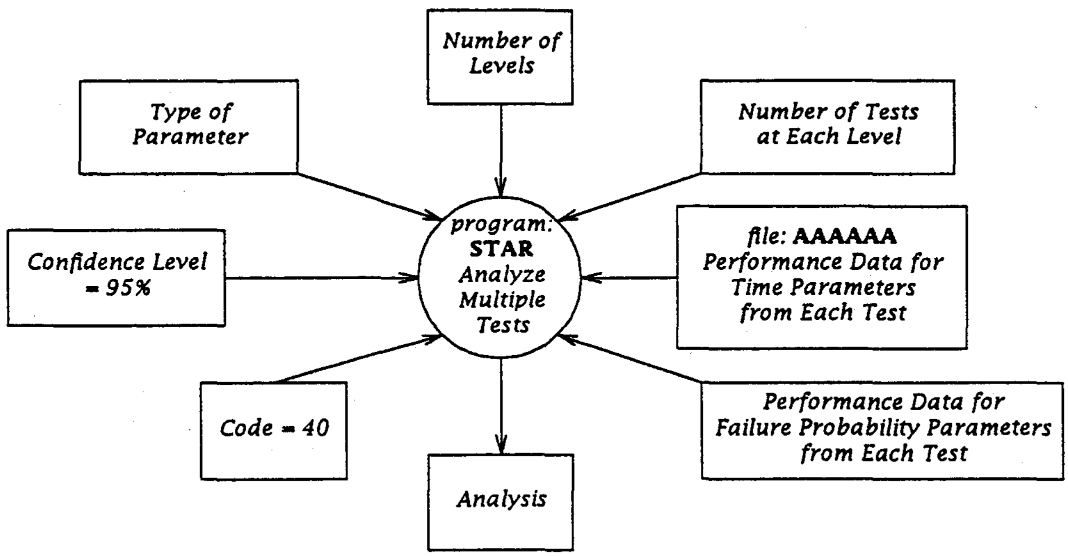

Figure 14. Structured design diagram of operator implementation of multiple test analysis.

2. ESTIMATE A PERFORMANCE PARAMETER

Due to sampling error, a parameter estimate obtained from measurements cannot be expected to equal the population value. Therefore, it is important that any such estimate be accompanied by a measure of its precision (e.g., confidence limits).[14] An estimate can be obtained from a single test (consisting of several trials) or from multiple tests.

2.1 Single Tests

All performance parameters except the User Information Bit Transfer Rate and the User Fraction of Input/Output Time can be estimated with desired precision from a single test. Tests of these parameters result in only one trial and they can be estimated with precision only from multiple tests.

Precision of the estimate is affected by the dispersion (as measured by the standard deviation) and dependence (as measured here by the autocorrelation of lag 1). Dispersion and positive autocorrelation always increase the length of the confidence interval. Negative autocorrelation slightly decreases it.

Estimates of confidence limits can be used to determine the achieved precision and compare it with the specified precision (Section 1.2.4, Volume 2).

The NTIA implementation also allows primary time parameters to be estimated by histograms (sample densities) and box plots (abbreviated histograms). Volume 6 shows how to obtain these two types of plots.

Estimation from a single test can be implemented by shell scripts or by an operator.

2.1.1 Estimation from Implementation by Shell Scripts

The performance parameters are estimated at the conclusion of the data reduction phase of each test. They are estimated at both the 90% and 95% confidence levels. Figures 2-5 are examples of estimates from shell script implementation.

---

[14]Precision for time parameters is defined by absolute precision (i.e., one- half the length of the confidence interval), and precision for failure probability parameters is defined by relative precision (i.e., one half the length of the confidence interval, divided by the estimate) and usually expressed as a percent.

The method by which estimation of a single test is implemented by a script is described briefly in Section 1.1.1 and in detail in Appendix C.

2.1.2 Estimation from Implementation by an Operator

Estimation of a performance parameter can also be implemented by an operator by typing **star** and providing the appropriate responses and performance data. Performance data for time parameters can be entered by either file or keyboard. Performance data for failure probability parameters must be entered by keyboard. Figures 7 and 8 are examples of the output from operator-implemented estimation of a delay parameter and a failure probability, respectively. The method by which this is done is described briefly in Section 1.1.2 and in detail in Appendix D.

2.2 Multiple Tests

Multiple tests of a performance parameter are conducted over selected combinations of levels of variable conditions. Multiple tests can be used for two types of estimates:

- More Precise Estimates. Multiple tests can provide a more precise estimate of a performance parameter, often from tests conducted at a single combination of levels (called replications). However, there are practical limitations to replication: Time and, often location, cannot be identical.

- Representative Estimates. Multiple tests can provide a single, representative estimate of a performance parameter from tests conducted at multiple combinations of levels.[15] This single estimate characterizes the performance parameter.

---

[15]The variable condition selected for analysis has the following effects:

- Pooling Among Trials: Selection does not affect the acceptance of the null hypothesis (that test means are equal) nor the estimate of the mean of the trials and its confidence limits.

- Pooling Among Test Means: Selection affects the acceptance of the null hypothesis (that level means are equal) but not the estimate of the mean of test means and its confidence limits.

- Pooling Among Level Means: There is no null hypothesis, but selection affects the estimate of the mean of level means and its confidence limits.

2.2.1 More Precise Estimate

To obtain a more precise estimate, the analyst can select tests at his/her discretion; however, the standard deviations should be somewhat the same. Usually, the set of tests consists of either replicated tests or tests conducted at the same combination of levels - except, perhaps, one.

If the trials of each test can be pooled, none of the variable conditions is a factor. The estimate from pooled trials from multiple tests should be more precise than that from a single test. If the trials cannot be pooled, perhaps either a subset of these tests or a different set of tests should be analyzed.

Example: Analyze multiple tests to obtain a more precise estimate of Access Time and User Fraction of Access Time.

Solution: A set of six tests having the same Source Site (i.e., Seattle) is selected. Use the **grep** utility to copy the identification of tests conducted from Seattle from **log.acc** to **log.wrk**. Type

grep sea log.acc > log.wrk

These tests have the same levels of Source Site (i.e., Seattle), Network (i.e., A), Interaccess Delay (i.e., 55 s), and Destination Site (i.e., Boulder). They have different levels of Day of the Week and Time of Day. For this set of tests, these are the only variable conditions. The purpose of pooling is to obtain a more precise estimate, so the selection of the variable condition to test the null hypothesis of equal level means is relatively unimportant. Arbitrarily select Day of the Week (whose code is 3). The code for Access Time is **ac** (Table 8). Type

delay ac 3

Figure 15 is the output of the analysis of the six tests. The * in the **among trials** row indicates that the trials of the tests can be pooled. This estimate can be used for Access Time (the **W** row) and User Fraction of Access Time (the **V** row). The absolute precision for Access Time at the 95% confidence level is 0.385 s. Conversely, the average absolute precision of the six single tests is 0.633 s. Therefore, pooling the trials has provided a more precise estimate.

Analysis of Multiple Tests

Access Time
Variable Condition 3

Thu Jan 26 16:41:08 MST 1989

			Times		User Fractions	
Test	--- Variable Conditions ---	Trials	Mean	Std Dev	Mean	Std Dev
823	sea netA fri 2 A55 bol	20	42.439	1.527	0.0339	0.0047
815	sea netA fri 6 A55 bol	20	41.576	1.269	0.0352	0.0044
835	sea netA mon 3 A55 bol	15	42.954	1.325	0.0345	0.0053
858	sea netA thu 1 A55 bol	20	42.284	1.338	0.0345	0.0053
876	sea netA thu 4 A55 bol	20	42.313	2.197	0.0350	0.0064
811	sea netA thu 5 A55 bol	19	41.163	1.015	0.0373	0.0065

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 114
NUMBER OF TESTS = 6
NUMBER OF LEVELS = 3

WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 6 TESTS = .3280E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 114 TRIALS = .2650E+00 @

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	52	5	.1579E+01	.2402E+01	W	.4171E+02	.4209E+02	.4248E+02	*
	-	-	-	-	V	.3398E-01	.3505E-01	.3613E-01	
AMONG TESTS	3	2	.1023E+01	.9550E+01	W	.4145E+02	.4212E+02	.4280E+02	*
	-	-	-	-	V	.3379E-01	.3506E-01	.3632E-01	
AMONG LEVELS	-	-	-	-	W	.4087E+02	.4229E+02	.4372E+02	
	-	-	-	-	V	.3337E-01	.3448E-01	.3639E-01	

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 15. Example of a precise estimate from multiple tests.

2.2.2 Representative Estimate

To obtain a representative estimate, the analyst can select tests at his/her discretion; however, the standard deviations should be somewhat the same. Usually the set of tests is selected so that, instead of many estimates, there will be a single, representative estimate.

If the trials of each test can be pooled, none of the variable conditions is a factor, and the estimate from pooling is representative.

If the trials cannot be pooled, at least one variable condition is a factor. Analysis continues by testing whether the means of each test can be pooled. If so, they are pooled, and the next best estimate of the performance parameter has been obtained (i.e., precision is less than when trials of tests are pooled).

If the means of each test cannot be pooled, the means of each level of the selected variable condition are pooled (there is no hypothesis test for this pooling, and precision is less than if means of tests can be pooled).

Example: It is desired to characterize Block Transfer Time and User Fraction of Block Transfer Time for network A. Seven tests have been conducted. Some variable conditions have a single level (and could be considered to be fixed conditions for this set of tests): All user information blocks contain 128 bytes, and they have been transferred without an interblock time gap (i.e., high utilization, denoted by B00). However, the tests were conducted from two cities, during four days and four time periods. So a single estimate must represent $2 \times 4 \times 4 = 32$ estimates.

Solution: Transfer the identification of the seven tests from **log.xfr** to **log.wrk**. To obtain a representative estimate, arbitrarily select Source Site (whose code is 1) as the variable condition for which to test the null hypothesis (that the means of levels are equal).[16] Type

delay b2 1

Figure 16 shows the result of this analysis. Since the **among trials** value of the F statistic (4.636) exceeds the value of the F distribution at the 5% point (2.100), at least one of the three variable conditions is a factor, and the trials cannot be pooled. The absolute precision from the test would have been 0.059 s.

However, the **among tests** pooling passes the hypothesis test, and the means of the tests can be pooled. This estimate is used as the single, representative estimate for these levels. The absolute precision is 0.118 s. Conversely, the average absolute precision of the seven single tests is 0.109 s.

We needn't use the least satisfactory, **among levels**, pooling whose absolute precision is 0.536 s.

---

[16]The variable condition selected for analysis has the following effects:

- Pooling Among Trials: Selection does not affect the acceptance of the null hypothesis (that test means are equal) nor the estimate of the mean of the trials and its confidence limits.

- Pooling Among Test Means: Selection affects the acceptance of the null hypothesis (that level means are equal) but not the estimate of the mean of test means and its confidence limits.

- Pooling Among Level Means: There is no null hypothesis, but selection affects the estimate of the mean of level means and its confidence limits.

ANALYSIS OF MULTIPLE TESTS

Block Transfer Time
Variable Condition 1

Thu Jan 26 16:44:06 MDT 1989

Test	--- Variable Conditions ---							Trials	---- Times ---- Mean	Std Dev	User Fractions Mean	Std Dev
1077	den	netA	fri	2	B00	bol	128	80	3.671	0.451	0.0981	0.0170
1060	den	netA	thu	4	B00	bol	128	80	3.660	0.413	0.0999	0.0167
1064	den	netA	thu	4	B00	bol	128	79	3.685	0.425	0.1016	0.0164
1025	den	netA	wed	2	B00	bol	128	80	3.707	0.439	0.0981	0.0159
998	wdc	netA	thu	3	B00	bol	128	80	3.914	0.577	0.0734	0.0149
950	wdc	netA	tue	5	B00	bol	128	80	3.701	0.428	0.0804	0.0151
976	wdc	netA	wed	4	B00	bol	128	80	4.175	0.756	0.0713	0.0166

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 559
NUMBER OF TESTS = 7
NUMBER OF LEVELS = 2

WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 7 TESTS = 0.4055E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 559 TRIALS = 0.2667E+00 @

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT
AMONG TRIALS	229	6	.4636E+01	.2100E+01	W	.3729E+01	.3788E+01	.3847E+01
	-	-	-	-	V	.8533E-01	.8906E-01	.9279E-01
AMONG TESTS	5	1	.4671E+01	.6610E+01	W	.3610E+01	.3788E+01	.3965E+01 *
	-	-	-	-	V	.7699E-01	.8960E-01	.1022E+00
AMONG LEVELS	-	-	-	-	W	.3270E+01	.3806E+01	.4341E+01
	-	-	-	-	V	.3438E-01	.8783E-01	.1413E+00

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICATES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 16. Example of a representative estimate from multiple tests.

3. DETERMINE ACCEPTABILITY OF A PERFORMANCE PARAMETER

An acceptance test is a hypothesis test that can determine whether the mean of a performance parameter equals or exceeds an acceptable (threshold) value.[17] Hence, an acceptance test is appropriate to determine the acceptability of a performance parameter for at least two experiment objectives:

- Acceptance. An acceptance test can determine if the mean of a performance parameter is acceptable (e.g., for purchase of a system or service).

- Maintenance. An acceptance test can determine if the mean of a performance parameter is acceptable (e.g., for users of an installed system or service). If not, the system may require maintenance to return the mean to an acceptable value.

The concepts described in Section 1.2.5 of Volume 2 are used in acceptance testing:

- Threshold. The (threshold) acceptable value is specified. This is a value that can be accepted with indifference.

- Interval of Uncertainty. Because a sample has a finite number of trials, an interval of uncertainty exists about the threshold value. 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.

- Null Hypothesis. The null hypothesis states that the population value of the performance parameter is equal to the totally satisfactory value.[18] Because we are interested in whether the parameter value is better than the totally satisfactory value, this hypothesis is tested by a one-sided test.

Acceptance tests involve two precision objectives:

- Incorrect Rejection. The probability of incorrectly rejecting a performance value that is totally satisfactory is to be $\alpha = 0.05$ or less (a probability called the significance

---

[17]This threshold value applies to acceptance tests; it has nothing to do with threshold values of the support parameters that determine the Transfer Denial Probability.

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

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. 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 interval, and rejected otherwise. Since NTIA analysis uses 90% or 95% confidence limits, α should be 5% or 2.5% respectively.[19]

- Incorrect Acceptance. The probability of (incorrectly) accepting a 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 (Section 8.1 of Volume 2).

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 17. 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 value is 0.5, and
- the probability of (incorrectly) accepting the hypothesis when performance is totally unsatisfactory is β.

Confidence limits obtained from pooled data from multiple tests are not appropriate for acceptance tests.

3.1 Time Parameters

Suppose that a mean delay of μ_0 would be barely acceptable (i.e., μ_0 is the threshold value).[20] The true delay cannot be known with certainty from a (finite) sample. In other words, we cannot achieve the ideal OC, a curve with the probability of acceptance of unity for $\mu < \mu_0$ and of zero for $\mu > \mu_0$; there

---

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

[20]The following discussion is in terms of delays. Rates would cause the discussion to be reversed in the sense that small delays and large rates are desirable.

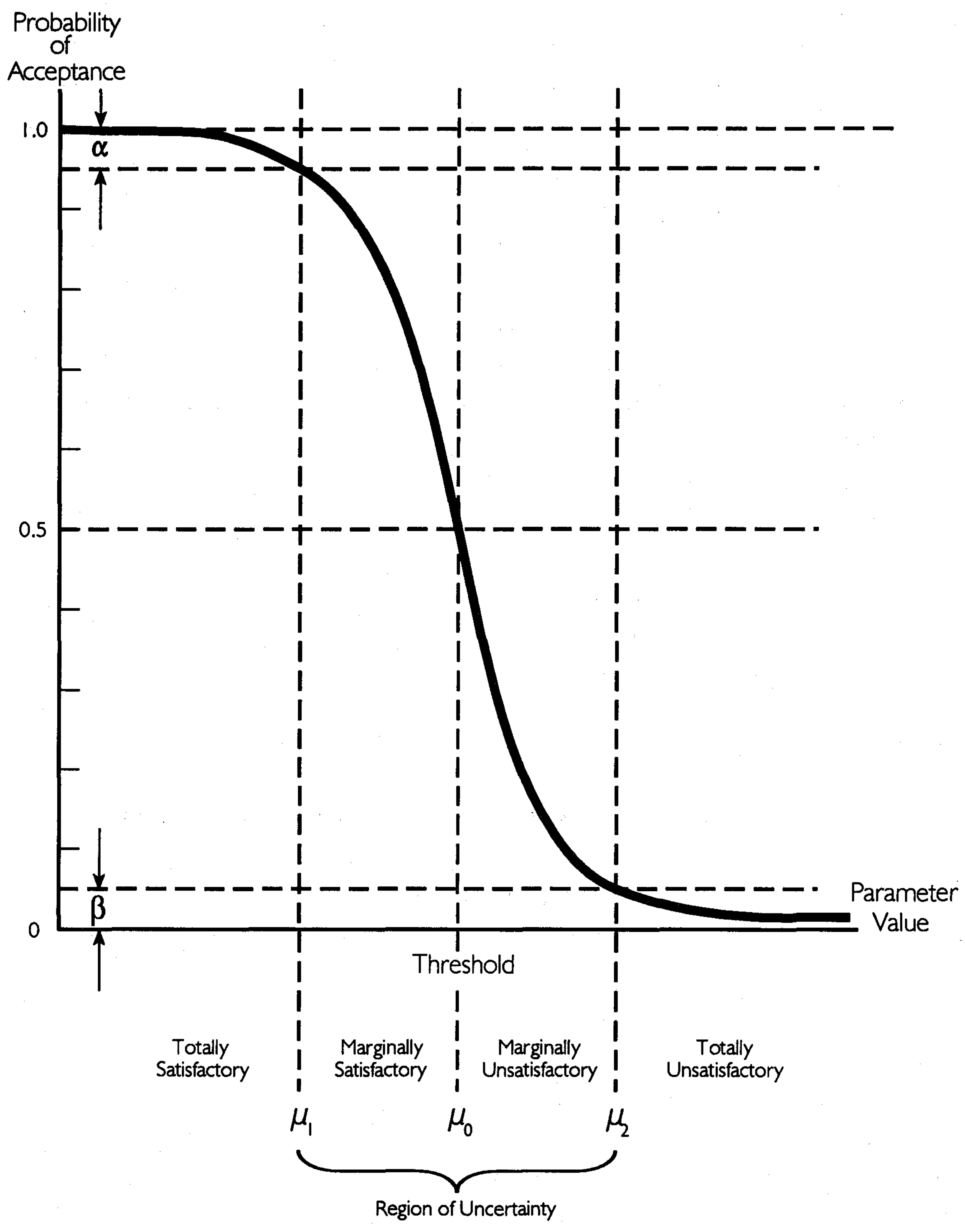

Figure 17. Schematic operating characteristic curve of the sampling plan for an acceptance test.

interval of uncertainty from the totally satisfactory performance and totally unsatisfactory performance are μ_1 and μ_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:

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

It is assumed that time parameters are normally distributed (at least approximately).

Example (delay): Suppose a proposed data communication service is specified 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 has approximately a normal distribution with standard deviation $\sigma = 8$ s, and the autocorrelation of lag 1 is estimated to be 0.5. Determine, at the $\alpha = 5\%$ significance level, if this performance parameter is acceptable .

Solution: Twenty-four trials are required to achieve the precision as expressed by the two probabilities specified in this example (Section 8.1.1 of Volume 2). That is,

- the probability of accepting a totally satisfactory time is $1 - \alpha = 0.95$, and
- the probability of accepting a totally unsatisfactory time is $\beta = 0.10$.

Since $\alpha = 5\%$, acceptability will be determined by $(1 - 2\alpha)100\% = 90\%$ confidence limits. The twenty-four trials resulted in lower and upper 90% confidence limits of 28.2 s and 51.4 s, respectively. Since the lower confidence limit is less than the totally satisfactory Access Time of 40 s, this performance parameter is acceptable.

Example (rate): A network is considered to be acceptable if the long-term User Information Bit Transfer Rate (i.e., 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. Determine if this performance parameter is acceptable at the $\alpha = 5\%$ significance level.

Solution: Since there is one throughput trial per test, the trials are considered to be independent. Twenty tests are required to achieve the two probabilities specified in this example (Section 8.1.1 of Volume 2). They are

- the probability of accepting a totally satisfactory time is $1 - \alpha = 0.95$, and
- the probability of accepting a totally unsatisfactory time is $\beta = 0.05$.

Since the trials are thought to be independent, $r_1 = 0$, and the required number of tests remains 20. Since $\alpha = 5\%$, acceptance will be determined by $(1 - 2\alpha)100\% = 90\%$ confidence limits. The 20 tests resulted in lower and upper 90% confidence limits of 1.3 Mbps and 3.4 Mbps, respectively. The performance parameter is not acceptable since the upper confidence limit is less than the totally acceptable value of 3.6 Mbps.

3.2 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 cannot be known with certainty from a finite sample. In other words, we cannot achieve the ideal OC, a curve with the probability of acceptance of unity for $p < p_0$ and of zero for $p > p_0$; there is an interval of uncertainty about this value. 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:

$$H_0: p = p_1$$

A performance parameter is considered to be acceptable if the lower confidence limit is less than the totally satisfactory value.

Example: A proposed data communication service is specified to have a Bit Error Probability not greater than $p_0 = 10^{-4}$, a value that is accepted with indifference. 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 thought to be dependent, and the autocorrelation of lag 1 is estimated to be 0.4. Determine if the Bit Error Probability is acceptable.

Solution: In Section 8.1.2 of Volume 2, it was determined that 59,501 bits must be transferred, and the service would be accepted if 4 or fewer bit errors were observed. A user information transfer test of 60,000 bits was conducted in which 3 bit errors and 1 pair of consecutive bit errors were observed. Since $\alpha = 0.05$, 90% confidence limits must be determined. The lower and upper 90% confidence limits are 5.33×10^{-6} and 1.77×10^{-4}, respectively. The lower confidence limit exceeds the totally acceptable level of

$$p_1 = 10^{-0.5} \times p_0 = 3.16 \times 10^{-4}.$$

Since these two values are really quite close, the Bit Error Probability could probably be considered either acceptable or unacceptable.

4. COMPARE A PERFORMANCE PARAMETER FROM TWO SERVICES

Performance parameters from two services or systems can be compared. The null hypothesis states that the means of the two performance parameters are equal:[21]

$$H_0: \mu_1 = \mu_2.$$

If estimates of the two means, each obtained at the same conditions, are significantly different at the $\alpha = 5\%$ significance level, the performance parameter from one service or system is preferred.

Hypotheses are tested by hypothesis tests, and the appropriate hypothesis test depends upon whether the performance parameter is a time parameter or a failure probability parameter. In either case, program **star** can be used to compare a performance parameter from two tests, each conducted at the same combination of levels - except Network, of course.[22]

4.1 Time Parameters

Use the UNIX[tm] **grep** utility to copy the identification of the two tests from the file called **log.acc** or **log.xfr** (depending upon whether the parameter is an access-disengagement or a user information transfer parameter, respectively) into the file called **log.wrk**. Then proceed as in Section 1.2.2 (subsections A or B). If the hypothesis test shows that the trials from the two tests can be combined (as indicated by the * in the **among trials** row), neither service/system can be preferred for that performance parameter. If the trials from the two tests cannot be combined, the performance parameter values from the two services/systems are significantly different and one can be preferred.

Example: Access Times for systems B and C were estimated in tests having eighteen trials (test number 867) and ten trials (test number 796), respectively. Determine if either system has a significantly shorter estimated Access Time.

---

[21]This is a special case of the analysis of multiple tests in Section 5 (because only two tests are used). The purpose there is to determine if a variable condition is a factor, not to compare two performance parameter values.

[22]Since Network is the only variable condition, no discussion of the implications of selecting a variable condition is needed (as it is for estimation and for determining if a variable condition is a factor).

Solution: Program **star** was executed and produced the results shown in Figure 18. The * to the right of the **among trials** row indicates there is no significant difference between the two systems at the 5% level (i.e., the pooling is acceptable at that level). Hence, neither system has a significantly shorter Access Time. Even though the estimate of one is 2.4% less than the other, the standard deviations are large enough that it is not significantly shorter.

4.2 Failure Probability Parameters

Copy the identification of the two tests from the file called **log.acc** or **log.xfr** (depending upon whether the parameter is an access-disengagement or a user information transfer parameter, respectively) into the file called **log.wrk**. Then proceed exactly as in Section 1.2.2 (subsection C). If the results of the hypothesis test show that the trials from the two tests can be combined (as indicated by the * in the **among trials** row), neither of the two services/systems can be preferred for that performance parameter. If the trials from the two tests cannot be combined, there is a significant difference between the systems or services (for that performance parameter). Choose the service/system having the smaller estimate of failure probability.

Example: Transfer Denial Probability was measured for systems A and C. Determine if either system has a significantly smaller Transfer Denial Probability.

Solution: Program **star** was executed and produced the results shown in Figure 19. The * to the right of the **among trials** row indicates there is no significant difference between the two systems at the 5% level (i.e., the pooling is acceptable at that level). Hence, neither system has a significantly smaller Transfer Denial Probability.

ANALYSIS OF MULTIPLE TESTS

Access Time
Variable Condition 1

Thu Feb 9 14:05:54 MST 1989

Test	--- Variable Conditions ---	Trials	---- Times ---- Mean	Std Dev	User Fractions Mean	Std Dev
867	sea B thu 3 L f-on tone	18	43.891	2.183	0.0334	0.0039
796	sea C thu 3 L foff tone	10	44.963	2.038	0.0351	0.0081

TIMES (W) AND FRACTION OF TIMES (V)

```
NUMBER OF TRIALS = 28
NUMBER OF TESTS  =  2
NUMBER OF LEVELS =  1
```

```
WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
                  OF LAG 1 OVER THE    2 TESTS =  .2253E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
                OF LAG 1 OVER THE   28 TRIALS =  .2223E+00 @
```

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT
AMONG TRIALS	16	1	.9983E+00	.4490E+01	W	.4320E+02	.4427E+02	.4534E+02
*								
	-	-	-	-	V	.3137E-01	.3405E-01	.3673E-01
AMONG TESTS	-	-	-	-	W	.4212E+02	.4443E+02	.4673E+02
	-	-	-	-	V	.3082E-01	.3427E-01	.3773E-01

THE F STATISTIC IS INDETERMINATE FOR THE AMONG TESTS POOLING,
SO THIS POOLING CANNOT BE TESTED.

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICATES THIS POOLING
  IS ACCEPTABLE AT THE 5% LEVEL.

Figure 18. Example of a comparison test for a time parameter.

ANALYSIS OF MULTIPLE TESTS

Transfer Denial Probability
Variable Condition 4

Thu Feb 9 14:31:34 MST 1989

Test	------ Variable Conditions ------	Trials	Failures	Pairs	Prob
901	wdc netA mon 2 B00 bol 128	19	3	1	0.158
911	wdc netC mon 4 B00 bol 128	19	1	0	0.053

FAILURE PROBABILITY

NUMBER OF TRIALS = 38
NUMBER OF TESTS = 2
NUMBER OF LEVELS = 2

AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 38 TRIALS =-.177 @

	DEGREES OF FREEDOM		X2 STAT.	X2 DIST.(5%)	95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	1	-	.7964E+00	.3841E+01	.22372E-01	.10526E+00	.28079E+00	*
			F STAT.	F DIST (5%)				
AMONG TESTS	-	-	-	-	.2038E-03	.1190E+00	.4052E+00	

THE F STATISTIC IS INDETERMINATE FOR THE AMONG TESTS POOLING,
SO THIS POOLING CANNOT BE TESTED.

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 19. Example of a comparison test for a failure probability parameter.

5. DETERMINE IF A VARIABLE CONDITION AFFECTS A PERFORMANCE PARAMETER

Two methods from Analysis of Variance are available to determine if a variable condition is a factor for a performance parameter: linear regression analysis (for primary time performance parameters) and hypothesis tests of the null hypothesis of equal means of tests (for any performance parameter).

5.1 Linear Regression Analysis

The NTIA implementation allows as many as eight variable conditions for access-disengagement tests or nine variable conditions for user information transfer tests. Linear regression analysis can be used to determine if a certain variable condition (having quantifiable levels) is a factor for a time parameter. The values of this primary time parameter (i.e., trial values or estimates from more than one trial) can be plotted at the measured levels of the variable condition. The levels of the variable condition are values of the independent variable, and the parameter values are values of the dependent variable. In the absence of measurement error, these points would lie on a curve. For simplicity, assume the curve is a straight line. The line is determined by the method of least squares (i.e., the line that minimizes the sum of the squares of the vertical distances between the points and the line). This line, called a regression line, is defined by

$$y = a + bx$$

where

$$b = \frac{\Sigma (y_i - \bar{y})(x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2},$$

and

$$a = \bar{y} - b\bar{x}.$$

In this case, x_i is the value of the ith level, and y_i is the value of the trial or estimate at that level.

The slope, b, of the regression line (called the regression coefficient) indicates the degree to which the performance parameter depends upon the variable condition; a slope of zero suggests independence, in which case we would conclude that the variable condition is not a factor for that parameter. For example, measurements over some public data networks of Block Transfer Time as a function of block length have shown that Block Transfer Time increases about 1 s for each additional 128 characters of block length (Spies et al., 1988).

We can test the null hypothesis that b = b′ where b′ is some selected slope. For this application, we want to determine if a variable condition is a factor for a performance parameter, so we would test the hypothesis that the slope is zero (i. e., b = 0).[23]

Define

$$t = \frac{(n - 1)^{1/2} b s_x}{[(n - 1)/(n - 2)](s_y^2 - b s_x^2)}$$

where

$$s_x = [1/(n - 1)\Sigma(x_i - \overline{x})^2],$$

$$s_y = [1/(n - 1)\Sigma(y_i - \overline{y})^2],$$

and n is the number of trials or estimates. Reject the null hypothesis

$$H_0: b = 0$$

at the α significance level if

$$|t| > t_{n-2,\ \alpha/2}.$$

Volume 6 of this report shows how regression lines can be plotted for primary time parameters. However, no analysis is supplied there.

5.2 Tests of Hypotheses

This analysis of multiple tests assumes that the experiment is designed to investigate the effect of a single variable condition upon a performance parameter. Hence, the design can be considered to be either a completely randomized design or a randomized block design (each with one variable

---

[23]The hypothesis test for the null hypothesis, b = b′, and confidence intervals for the slope are available (Crow et al., 1960, p. 160).

condition). Performance data from tests having different levels of this variable condition are pooled.

The experiment may have been designed to determine if one or more of N identified variable conditions are factors.[24] To determine if they are, one or more samples have probably been obtained at different combinations of levels of the variable conditions. Each combination of levels of the variable conditions defines a population, and the null hypothesis states that the means of the populations, say k of them, are equal:

$$H_0: \mu_1 = \ldots = \mu_k.$$

The null hypothesis is tested by a hypothesis test. The statistic is compared with an appropriate distribution at, say, the α = 5% point. If the value of the statistic is less than this value, the null hypothesis is accepted, and none of the tested variable conditions are factors. Otherwise, at least one is a factor.

Program **star** can determine if any of the variable conditions are factors. Suppose a set of tests having j variable conditions is to be analyzed (i.e., $1 \leqq j \leqq N$). Tests from the experiment can be selected for two purposes:

- Determine if One Variable Condition is a Factor. Select tests having the same combination of levels except for those of one variable condition (i.e., j = 1). The hypothesis test determines if that variable condition is a factor.

- Determine if at Least One of the j Variable Conditions is a Factor. Select tests having the same combination of levels except for j variable conditions (i.e., j > 1). The hypothesis test determines if any of those j variable conditions is a factor. However, it will not indicate which are. Of the j variable conditions, one must be selected to test the null hypothesis that the means of its levels are equal. Selection has the following effects:

  - Pooling Among Trials. Selection does not affect the acceptance of the null hypothesis that test means are equal nor the estimates of the mean of the trials and its confidence limits.

---

[24]Methods to define statistics that test whether multiple variable conditions are factors can be found in many texts concerning experiment design.

- Pooling Among Test Means. Selection affects the acceptance of the null hypothesis that level means are equal but not the estimates of the mean of test means and its confidence limits.

- Pooling Among Level Means. There is no null hypothesis, but selection affects the estimates of the mean of level means and its confidence limits.

5.2.1 Time Parameters

Use the UNIX[tm] **grep** utility to copy the identification of the selected tests from the file called **log.acc** or **log.xfr** (depending upon whether the parameter is an access-disengagement parameter or a user information transfer parameter, respectively) into the file called **log.wrk**. Then type one of the commands from Table 8 to implement **star**. If the hypothesis test determines that the trials can be pooled (as indicated by the * in the **among trials** row), none of the conditions is a factor.[25] Otherwise, any of the j variable conditions with different levels is a factor. Other statistics are required to determine which of the j variable conditions are factors.

Example 1: Eleven tests of Access Time have been conducted on System A. The five identified variable conditions are

- Source Site (three levels: Fort Worth, Seattle, and Washington D.C.),

- Day of Week (five levels: Monday, Tuesday, Wednesday, Thursday, and Friday),

- Time of Day (six levels, each containing 4 hours: identified by 1, 2, 3, 4, 5, and 6),

- Interaccess Delay (One level: 55 s), and

- Destination Site (One level: Boulder).

Since two of the variable conditions have only one level, there are really only three variable conditions. Determine if any of these three conditions are factors for Access Time on System A.

---

[25]Only the **among trials** pooling can be used to determine this.

Solution: Suppose the identification of these 11 tests is in **log.wrk**. Then type

delay ac 1

where variable condition 1 has been arbitrarily selected.[26] Program **star** is executed by the shell script, and the results are shown in Figure 20. Since the * does not appear in the **among trials** row, at least one of the three variable conditions is a factor.

Example 2: Determine if the variable condition, Source Site, is a factor for Access Time in the 11 tests of Example 1.

Solution: Form two subsets of tests: those from Seattle and those from Washington D.C.

A. A Subset from the Source Site, Seattle

Select the tests conducted in Seattle by typing

grep sea log.wrk > log.aaa

Then place their identifications into **log.wrk** by typing

cp log.aaa log.wrk

To implement **star**, type

delay ac 1

where Source Site has been arbitrarily selected as the variable condition to test the null hypothesis that the test means are equal.[27] Program **star** is executed using these six tests, and the results are shown in Figure 21. There are only two variable conditions: Day of the Week and Time of Day. Since the * appears in the **among trials** row, neither are factors, and Source Site must have been a factor.

---

[26]Its selection will not affect the test of the null hypothesis that the test means are equal.

[27]Since Source Site has only one level in this set of tests, the **among level** means is not analyzed.

Analysis of Multiple Tests

Access Time
Variable Condition 1

Thu Jan 26 14:06:14 MST 1989

Test	--- Variable Conditions ---	Trials	---- Times ---- Mean	Std Dev	User Fractions Mean	Std Dev
775	ftw netA fri 1 A55 bol	20	38.291	1.608	0.0397	0.0199
823	sea netA fri 2 A55 bol	20	42.439	1.527	0.0339	0.0047
815	sea netA fri 6 A55 bol	20	41.576	1.269	0.0352	0.0044
835	sea netA mon 3 A55 bol	15	42.954	1.325	0.0345	0.0053
858	sea netA thu 1 A55 bol	20	42.284	1.338	0.0345	0.0053
876	sea netA thu 4 A55 bol	20	42.313	2.197	0.0350	0.0064
811	sea netA thu 5 A55 bol	19	41.163	1.015	0.0373	0.0065
997	wdc netA thu 3 A55 bol	17	41.751	2.198	0.0356	0.0075
928	wdc netA tue 1 A55 bol	20	44.500	4.380	0.0332	0.0068
952	wdc netA tue 5 A55 bol	20	39.813	1.625	0.0368	0.0043
978	wdc netA wed 4 A55 bol	18	42.304	1.820	0.0351	0.0054

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 209
NUMBER OF TESTS = 11
NUMBER OF LEVELS = 3

WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 11 TESTS = .3927E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 209 TRIALS = .4998E+00 @

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	80	10	.4961E+01	.1963E+01	W	.4112E+02	.4173E+02	.4234E+02	
	-	-	-	-	V	.3423E-01	.3547E-01	.3671E-01	
AMONG TESTS	8	2	.4011E+01	.4460E+01	W	.4067E+02	.4176E+02	.4286E+02	*
	-	-	-	-	V	.3416E-01	.3551E-01	.3686E-01	
AMONG LEVELS	-	-	-	-	W	.3536E+02	.4083E+02	.4631E+02	
	-	-	-	-	V	.2953E-01	.3676E-01	.4417E-01	

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 20. Example of a test of variable conditions for a time parameter.

Analysis of Multiple Tests

Access Time
Variable Condition 1

Thu Jan 26 16:47:48 MST 1989

Test	--- Variable Conditions ---	Trials	---- Times ---- Mean	Std Dev	User Fractions Mean	Std Dev
823	sea netA fri 2 A55 bol	20	42.439	1.527	0.0339	0.0047
815	sea netA fri 6 A55 bol	20	41.576	1.269	0.0352	0.0044
835	sea netA mon 3 A55 bol	15	42.954	1.325	0.0345	0.0053
858	sea netA thu 1 A55 bol	20	42.284	1.338	0.0345	0.0053
876	sea netA thu 4 A55 bol	20	42.313	2.197	0.0350	0.0064
811	sea netA thu 5 A55 bol	19	41.163	1.015	0.0373	0.0065

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 114
NUMBER OF TESTS = 6
NUMBER OF LEVELS = 1

WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 6 TESTS = .3280E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 114 TRIALS = .2650E+00 @

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	52	5	.1579E+01	.2402E+01	W	.4171E+02	.4209E+02	.4248E+02	*
	-	-	-	-	V	.3398E-01	.3505E-01	.3613E-01	
AMONG TESTS	-	-	-	-	W	.4145E+02	.4212E+02	.4280E+02	
	-	-	-	-	V	.3379E-01	.3506E-01	.3632E-01	

THE F STATISTIC IS INDETERMINATE FOR THE AMONG TESTS POOLING,
SO THIS POOLING CANNOT BE TESTED.

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 21. Example of a test of variable conditions for a time parameter using the source site, Seattle.

B. A Subset from the Source Site, Washington D.C.

Select only those tests conducted in Washington D.C. Proceed as above except use **wdc** instead of **sea** with the **grep** utility. The results are shown in Figure 22.

- Figure 21 shows that neither Day of Week nor Time of Day is a factor in Seattle.
- Figure 22 shows that Day of Week and/or Time of Day is a factor in Washington D.C.

Therefore, the fact that Source Site is a factor for Access Time in the experiment of Example 1 is probably due to temporal phenomena in Washington D.C.

5.2.2 Failure Probability Parameters

Use the UNIX^{tm} **grep** utility to copy the identification of the selected tests from **log.acc** or **log.xfr** (depending upon whether the performance parameter is an access-disengagement parameter or a user information transfer parameter) into **log.wrk**. Then proceed as indicated in subsection C of Section 1.2.2 (subsection C). If the hypothesis test determines that the trials can be pooled (as indicated by the * in the **among** trials row), none of the variable conditions is a factor.[28] Otherwise, any of the j variable conditions is a factor, and other statistics are required to determine which are.

Example: Eleven tests of Source Disengagement Denial Probability have been conducted on system A. (These are the same tests as discussed in Example 1 of the previous section.) Since two of the variable conditions are represented by only one level, there are really only three variable conditions. Determine if any of these three variable conditions are factors for Source Disengagement Denial Probability on system A.

Solution: To implement **star**, type

fail dll 1

Program **star** is executed using these eleven tests, and the results are shown in Figure 23. Since the * appears in the **among trials** row, none of the three variable conditions is a factor for Source Disengagement Denial.

---

[28]Only the **among trials** row is relevant to analyzing factor effects.

Analysis of Multiple Tests

Access Time
Variable Condition 1

Thu Jan 26 16:49:28 MST 1989

Test	--- Variable Conditions ---	Trials	---- Times ---- Mean	Std Dev	User Fractions Mean	Std Dev
997	wdc netA thu 3 A55 bol	17	41.751	2.198	0.0356	0.0075
928	wdc netA tue 1 A55 bol	20	44.500	4.380	0.0332	0.0068
952	wdc netA tue 5 A55 bol	20	39.813	1.625	0.0368	0.0043
978	wdc netA wed 4 A55 bol	18	42.304	1.820	0.0351	0.0054

TIMES (W) AND FRACTION OF TIMES (V)

NUMBER OF TRIALS = 75
NUMBER OF TESTS = 4
NUMBER OF LEVELS = 1

WEIGHTED AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 4 TESTS = .3628E+00 #
AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 75 TRIALS = .4447E+00 @

	EFFECTIVE DEGREES OF FREEDOM		F STAT.	F DIST. (5%)		95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	31	3	.4197E+01	.2912E+01	W	.4088E+02	.4210E+02	.4332E+02	*
	-	-	-	-	V	.3335E-01	.3495E-01	.3655E-01	
AMONG TESTS	-	-	-	-	W	.3902E+02	.4209E+02	.4516E+02	
	-	-	-	-	V	.3229E-01	.3502E-01	.3774E-01	

THE F STATISTIC IS INDETERMINATE FOR THE AMONG TESTS POOLING,
SO THIS POOLING CANNOT BE TESTED.

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 22. **Example of a test of variable conditions for a time parameter using the source site, Washington, D.C.**

ANALYSIS OF MULTIPLE TESTS

Source Disengagement Denial Probability
Variable Condition 1

Fri Jan 27 10:39:17 MST 1989

Test	------ Variable Conditions ------	Trials	Failures	Pairs	Prob
775	ftw netA fri 1 A55 bol	20	1	0	0.050
823	sea netA fri 2 A55 bol	20	1	0	0.050
815	sea netA fri 6 A55 bol	20	3	1	0.150
835	sea netA mon 3 A55 bol	15	0	0	0.000
858	sea netA thu 1 A55 bol	20	3	0	0.150
876	sea netA thu 4 A55 bol	20	1	0	0.050
811	sea netA thu 5 A55 bol	19	0	0	0.000
997	wdc netA thu 3 A55 bol	17	1	0	0.059
928	wdc netA tue 1 A55 bol	20	1	0	0.050
952	wdc netA tue 5 A55 bol	20	3	0	0.150
978	wdc netA wed 4 A55 bol	18	1	0	0.056

FAILURE PROBABILITY

NUMBER OF TRIALS = 209
NUMBER OF TESTS = 11
NUMBER OF LEVELS = 3

AVERAGE AUTOCORRELATION COEFFICIENT
OF LAG 1 OVER THE 209 TRIALS =-.002 @

	DEGREES OF FREEDOM		X2 STAT.	X2 DIST.(5%)	95% LOWER CONFIDENCE LIMIT	ESTIMATE OF THE MEAN	95% UPPER CONFIDENCE LIMIT	
AMONG TRIALS	10	-	.8849E+01	.1831E+02	.41571E-01	.71770E-01	.11642E+00	*
			F STAT.	F DIST (5%)				
AMONG TESTS	8	2	.2018E+00	.4460E+01	.4526E-01	.7906E-01	.1212E+00	*
AMONG LEVELS	-	-	-	-	.4624E-01	.7808E-01	.1174E+00	

USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE F TEST.
@ USED TO DETERMINE THE EFFECTIVE DEGREES OF FREEDOM FOR THE CONFIDENCE LIMITS.
* AT RIGHT OF UPPER CONFIDENCE LIMIT INDICTES THIS POOLING
IS ACCEPTABLE AT THE 5% LEVEL.

Figure 23. **Example of a test of variable conditions for a failure probability parameter.**

6. ACKNOWLEDGMENTS

The author wishes to thank Edwin L. Crow for extensive technical advice, particularly with acceptance tests and to Earl D. Eyman and Catherine L. Edgar for testing many examples of the computer program; he wishes to also thank Lewis E. Vogler for suggesting an improvement to subroutine **limit**, Margaret H. Pinson for many suggestions to improve this volume, and Rob Reichert for preparing the figures and tables for publication.

7. 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, X3.102,- 1992, (American National Standards Institute, Inc., New York, NY).

Crow, E.L. (1974), Confidence limits for digital error rates, Office of Telecommunications Report 74-51, November. [Derivation of confidence limits applicable to small probabilities. Independence of successive trials assumed.] (NTIS Order No.: COM 7510793)

Crow, E.L. (1978), Relations between bit and block error probabilities under Markov dependence, Office of Telecommunications Report 78-143, March. [Derivation of confidence limits for block error probability. Approximation for the probability of m errors in a block of n bits.] (NTIS Order No.: PB 281545)

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

Crow, E.L., and M.J. Miles (1976), A minimum cost, accurate statistical method to measure bit error rates, Proceedings of the International Conference on Computer Communications, Toronto, Canada, August, pp. 631-635 (International Council for Computer Communication, Washington, D. C.). [Practical procedure for obtaining bit error rate measurements with a specified precision.]

Crow, E.L., and M.J. Miles (1977), Confidence limits for digital error rates from dependent transmissions, Office of Telecommunications Report 77-118, March. [Extension of E. L. Crow's 1974 report on the case of dependent errors. Two-state (Markov) error mode assumed.] (NTIS Order No.: PB267777).

Fisz, M. (1963), *Probability Theory and Mathematical Statistics*, Third edition, (John Wiley & Sons, Inc., New York, NY). [Comprehensive text book.]

Miles, M.J. (1984), Sample size and precision in communication performance measurements, NTIA Report 84-153, August, 130 pp. (NTIS Order No.: PB 85-114270). [Report describing FORTRAN program to implement sample size determination to achieve specified precision and subsequent analysis of single and multiple tests.]

Mood, A.M., F.A. Graybill, and D.C. Boes (1974), *Introduction to the Theory of Statistics*, 3rd Edition (McGraw-Hill Book Company, Inc., 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 82-238, August, 294 pp. (NTIS Order No.: PB 89-117519/AS).

APPENDIX A: FORMULAS FOR ANALYSIS OF A SINGLE TEST

Analysis of a single test consists of estimating the population mean and its confidence interval for a given confidence level.

In all cases, dependence between trials is estimated by a first order Markov chain. This type of dependence can be estimated by the following expression (which might be called the "dependence factor"):

$$c_n^2(r_1) = 1 + \frac{2r_1}{n(1 - r_1)} \cdot \left(n - \frac{1 - r_1^n}{1 - r_1}\right) \tag{A-1}$$

where r_1 is the estimate of autocorrelation of lag 1, and n is the number of trials. This expression is multiplied by certain terms to estimate serial dep ndence; it applies to both time and failure probability parameters.

A.1 Time Parameters

There are two types of time parameters: delay and rate parameters. The delay can be the total delay or the part of the total delay for which the user is responsible. The rate is the number of elements transferred during a certain period of time (e.g., the User Information Bit Transfer Rate). For time parameters, the autocorrelation of lag 1 can be estimated by the autocorrelation coefficient,

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

where $\bar{w}$ is the estimate of the mean, and s^2 is the estimate of the variance.[1]

A.1.1 Total Delay

The population mean delay, W, is estimated from n delays by

$$\bar{w} = \frac{1}{n} \sum_{i=1}^{n} w_i\ , \tag{A-3}$$

and the lower and upper 100(1-2α)% confidence limits for W are

[1]Equation A-23 estimates r_1 for failure probabilities.

$$\left.\begin{matrix} w_L \\ w_U \end{matrix}\right\} = \overline{w} \pm t_{n-1,\alpha} \cdot (s/\sqrt{n}) \cdot c_n[r_1(w)] . \qquad \text{(A-4)}$$

where $t_{n-1,\alpha}$ is the upper 100α% point of the Student t distribution for n-1 degrees of freedom.

A.1.2 User Fraction of the Delay.

If W is the mean delay and T is the mean user-responsible delay, the mean of the user-responsible fraction of the delay is

$$V = \frac{T}{W}. \qquad \text{(A-5)}$$

An unbiased estimate of V is

$$\overline{v} = \frac{\overline{t}}{\overline{w}} \cdot \left[1 + \frac{1}{n}\left(\frac{s_{tw}}{\overline{t}\overline{w}} - \frac{s_w^2}{\overline{w}^2}\right)\right] \qquad \text{(A-6)}$$

where

$$\overline{t} = \frac{1}{n} \sum_{i=1}^{n} t_i , \qquad \text{(A-7)}$$

$$s_t^2 = \frac{1}{n-1} \sum_{i=1}^{n} (t_i - \overline{t})^2 \cdot c_n^2 \, [r_1(t)] , \qquad \text{(A-8)}$$

$$s_w^2 = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \overline{w})^2 \cdot c_n^2 \, [r_1(w)] , \qquad \text{(A-9)}$$

$$s_{tw} = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \overline{w})(t_i - \overline{t}) \cdot c_n[r_1(t) \cdot r_1(w)] , \qquad \text{(A-10)}$$

and

$$s_{\overline{v}}^2 = \frac{\overline{v}^2}{n} \cdot \left(\frac{s_t^2}{\overline{t}^2} + \frac{s_w^2}{\overline{w}^2} - 2 \frac{s_{tw}}{\overline{tw}} \right) . \qquad \text{(A-11)}$$

The confidence limits for v are

$$\left. \begin{matrix} v_L \\ v_U \end{matrix} \right\} = \overline{v} \pm u_\alpha s_{\overline{v}} \qquad \text{(A-12)}$$

where u_α is the upper 100α% point of the normal density.

A.1.3 Rate

NTIA test procedures provide one trial per test of User Information Bit Transfer Rate and its ancillary parameter, User Fraction of Input/Output Time; hence, there is no precision (i.e., confidence limits cannot be computed). However, analysis of rates for single tests is included here because other implementations might provide more than one trial per test.

If b is the number of elements (e.g., bits) successfully transferred during a performance measurement period and w is the duration of the period, the transfer rate for a particular period is

$$r = \frac{b}{w} . \qquad \text{(A-13)}$$

The transfer rate of the data communication system is

$$R = \lim_{w \to \infty} \frac{b}{w} . \qquad \text{(A-14)}$$

It can be estimated by

$$\overline{R} = \frac{\overline{b}}{\overline{w}} \qquad \text{(A-15)}$$

where

$$\bar{b} = \frac{1}{n} \sum_{i=1}^{n} b_i \, , \tag{A-16}$$

and $\bar{w}$ is defined in Equation A-3. Each b_i should be nearly equal, and each w_i should be allowed to vary. The confidence limits for the system user information transfer rate, R, are

$$R_L = \frac{\bar{b}}{W_U}$$

and (A-17)

$$R_U = \frac{\bar{b}}{W_L}$$

where W_L and W_U are determined in Equation A-4.

A.2 Failure Probability Parameters

Suppose P_L and P_U are the lower and upper confidence limits for the failure probability, p. We seek a $100(1-2\alpha)\%$ confidence interval for p such that

$$\sum_{i=s}^{n} f(i|p_L,\lambda,n) = \alpha, \text{ and } \sum_{i=0}^{s} f(i|p_U,\lambda,n) = \alpha \tag{A-18}$$

where $f(i|p,\lambda,n)$ is the probability function of s with parameters p, λ, and n.

If λ is known, these sums determine the exact confidence limits for p. However, the procedure requires significant computer time and storage for n exceeding, say, 500 (Crow and Miles, 1977). Furthermore, for a large sample size and small probabilities, exact confidence limits are unnecessary.

When the number of failures exceeds one, the confidence limits can be approximated satisfactorily by using the normal approximation and the Poisson approximation; these two approximations are then averaged. To obtain the normal approximation, the sums (Equation A-18) are replaced by the normal integral with the mean and variance of s (Crow and Miles, 1977). For small p, the binomial distribution (for the number of failures) can be well approximated by the Poisson distribution as modified by Anderson and Burstein, and this was adapted to a generalized binomial distribution involving λ (Crow and Miles, 1977).

Analysis of the failure probability parameters involves estimating the mean failure rate, p, and the upper and the lower confidence limits, p_U and p_L. The unbiased estimate of the mean failure probability is

$$\overline{P} = \frac{s}{n} \tag{A-19}$$

where s is the number of failures, and n is the sample size.

Formulas for the 100(1-2α)% confidence limits for p depend upon whether the number of failures exceeds 1 or not.

A.2.1 Number of Failures Exceeds 1.

In this case, the confidence limits for p are the average of the 100(1-2α)% confidence limits for an approximation via the normal density and an approximation via the Poisson density

$$P_L = \frac{P_{LN} + P_{LP}}{2} ,$$

and (A-20)

$$P_U = \frac{P_{UN} + P_{UP}}{2} .$$

As seen in Equations A-25 and A-29, both approximations utilize the dependence factor[2] where

$$\overline{Q} = 1 - \overline{P} , \tag{A-21}$$

$$\overline{\lambda} = \frac{r}{s - \overline{P}} , \tag{A-22}$$

[2]Often r_1 is very small and n is very large. In such cases, r_1^n would be small enough to cause an exception in the execution of **star**. To avoid this, the program assigns to r_1^n the maximum of r_1^n or 1×10^{-20}. See subroutine **limit**.

and

$$r_1 = \frac{\overline{\lambda} - \overline{P}}{\overline{Q}}. \tag{A-23}$$

A. **Normal Approximation Confidence Limits**

These limits are given by

$$P_{LN} = \frac{nV + 2s - 1 - R_-}{2n\ (1+V)}$$

and (A-24)

$$P_{UN} = \frac{nV + 2s + 1 + R_+}{2n\ (1 + V)}$$

where

$$\overline{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{P}\overline{Q}}{n}} \cdot c_n(r_1), \tag{A-25}$$

$$V = \frac{(nu_\alpha \overline{\sigma}_{\overline{p}})^2}{s(n - s)}, \tag{A-26}$$

$$R_+ = [(nV + 2s + 1)^2 - (2s + 1)^2\ (1 + V)]^{1/2}, \tag{A-27}$$

$$R_- = [(nV + 2s - 1)^2 - (2s - 1)^2(1 + V)]^{1/2}, \tag{A-28}$$

and u_α is the upper $100\alpha\%$ point of the normal density.

B. **Poisson Approximation Confidence Limits**

These confidence limits depend upon whether P_{LP} exceeds zero or not:

- if $P_{LP} \geq 0$,

$$P_{LP} = \overline{P} - (\overline{P} - P_{LI}) \cdot c_n(r_1),$$
and (A-29)
$$P_{UP} = \overline{P} + (P_{UI} - \overline{P}) \cdot c_n(r_1).$$

- If $P_{LP} < 0$,

$$P_{LP} = 0,$$
and (A-30)
$$P_{UP} = \overline{P} + (P_{UI} - \overline{P}) \cdot c_n(r_1).$$

The confidence limits, P_{LI} and P_{UI}, are approximate confidence limits for p assuming the trials are independent (i.e., assuming λ = p). These confidence limits are

$$P_{LI} = \frac{L}{n - \left(\frac{s-1-L}{2}\right)},$$

and (A-31)

$$P_{UI} = \frac{U}{n + d + \left(\frac{U-s}{2}\right)}$$

where L and U are confidence limits for the mean of a Poisson distribution and d is a numerical adjustment. (U, L, and d are determined from tables in subroutine **poiss**.)

A.2.2 Number of Failures is 0 or 1

In this case, the confidence limits are obtained from the cumulative probability function of s (i.e., $\sum_{i=0}^{1} f(i|p,\lambda,n) = \alpha$):

$$P_L = 0 \,,$$

and (A-32)

$$P_U = \frac{1 - x}{2 - \lambda - x}$$

where

$$x = \begin{cases} \left(\dfrac{\alpha}{Q_u}\right)^{\frac{1}{n-1}} & \text{for } s = 0 \\ \left(\dfrac{\alpha Q_u}{1 + z_2 P_U - z_1 P_U^2}\right)^{\frac{1}{n-3}} & \text{for } s = 1, \end{cases} \tag{A-33}$$

$$z_1 = (n-1)\ (1-\lambda)^2 - 1,$$

and (A-34)

$$z_2 = (n-2)\ (1-\lambda)^2 - 2\,.$$

The value of P_U is obtained by iteration. The first value of Q_U is

$$Q_U = 1 - P_{UP} \tag{A-35}$$

where

$$P_{UP} = (P_{UI} - P_{LI}) \cdot c_n(r_1)\ . \tag{A-36}$$

P_{UP} is the upper Poisson approximate confidence limit when $s > 1$, and $\overline{P}_{LP} < 0$, but $\overline{\lambda}$ is replaced by λ in Equation A-23 for computing $c_n(r_1)$. Subsequently, where P_U was obtained in the previous iteration (i.e., if indices are used, $Q_{U(i)} = 1 - P_{U(i-1)}$ for i = 2, 3,...). Iteration continues until

$$1 - 10^{-k} < \frac{P_{U(i)}}{P_{U(i-1)}} < 1 + 10^{-k} \tag{A-38}$$

where $k > 0$. In **star**, $k = 4$.

APPENDIX B: FLOWCHARTS OF ANALYSIS OF A SINGLE TEST

This appendix is a set of flowcharts for each subroutine from **star** that analyze a single test.

Figure B-1 shows the relationship of the subroutines of **star**. Before conducting the experiment, the experimenter may have had insufficient information about a population and required a preliminary sample. In such case, when he/she analyzes the preliminary sample, subroutines **ssdtim** and **ssdflr** are used prior to **anztim** and **anzflr**, respectively, to determine the size of an additional sample that is required to achieve the specified precision. These subroutines are not shown in Figure B-1 (because they may be used only when **star** is implemented by an operator), but their flowcharts are included as Figures B-7 and B-8.

The flowcharts of each subroutine are listed alphabetically.[1] In these figures, diamonds indicate decisions, rectangles indicate arithmetic operations, and parallelograms indicate input (output is generally omitted).

---

[1]Three subroutines (called **entera**, **enteri**, and **enterx**) that allow entry of data and responses from a keyboard are omitted.

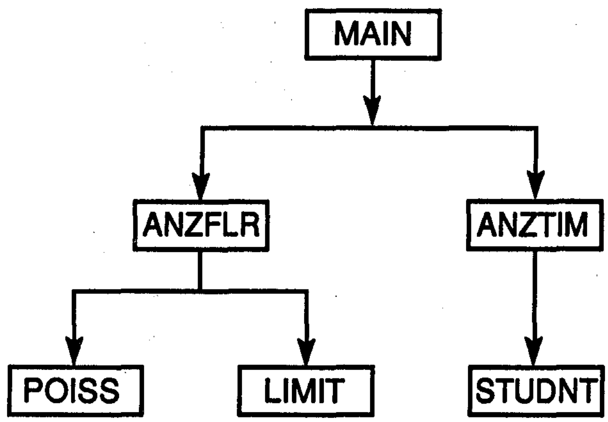

Figure B-1. Flowchart of relationship among subroutines of **star**.

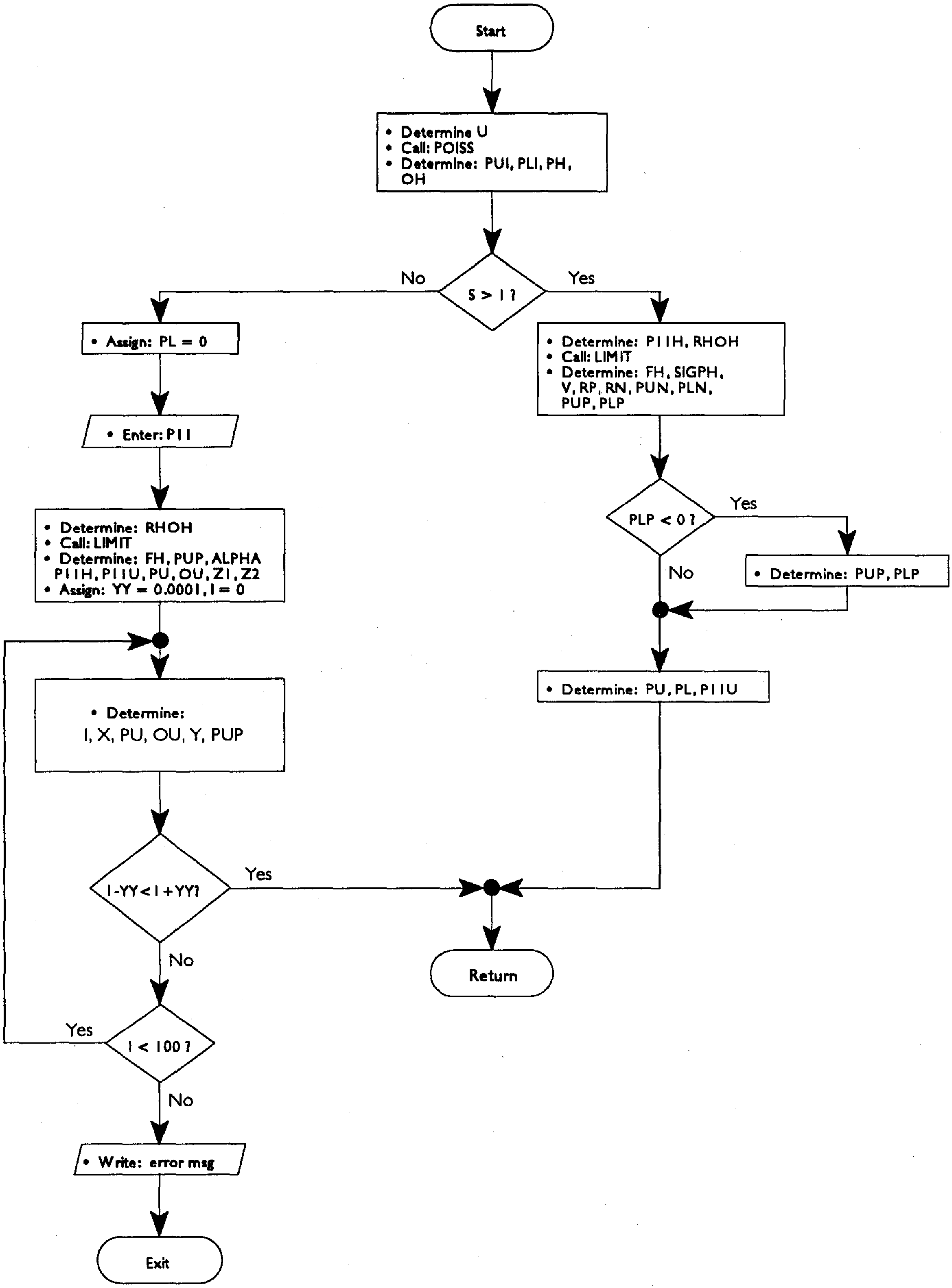

Figure B-2. Flowchart of subroutine anzflr.

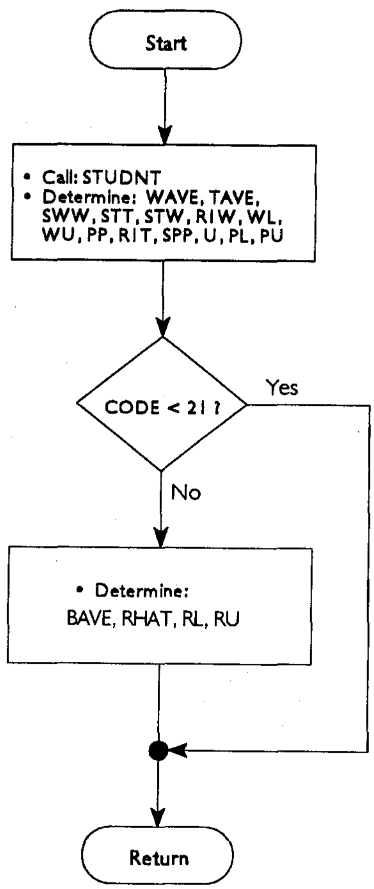

Figure B-3. Flowchart of subroutine **anztim**.

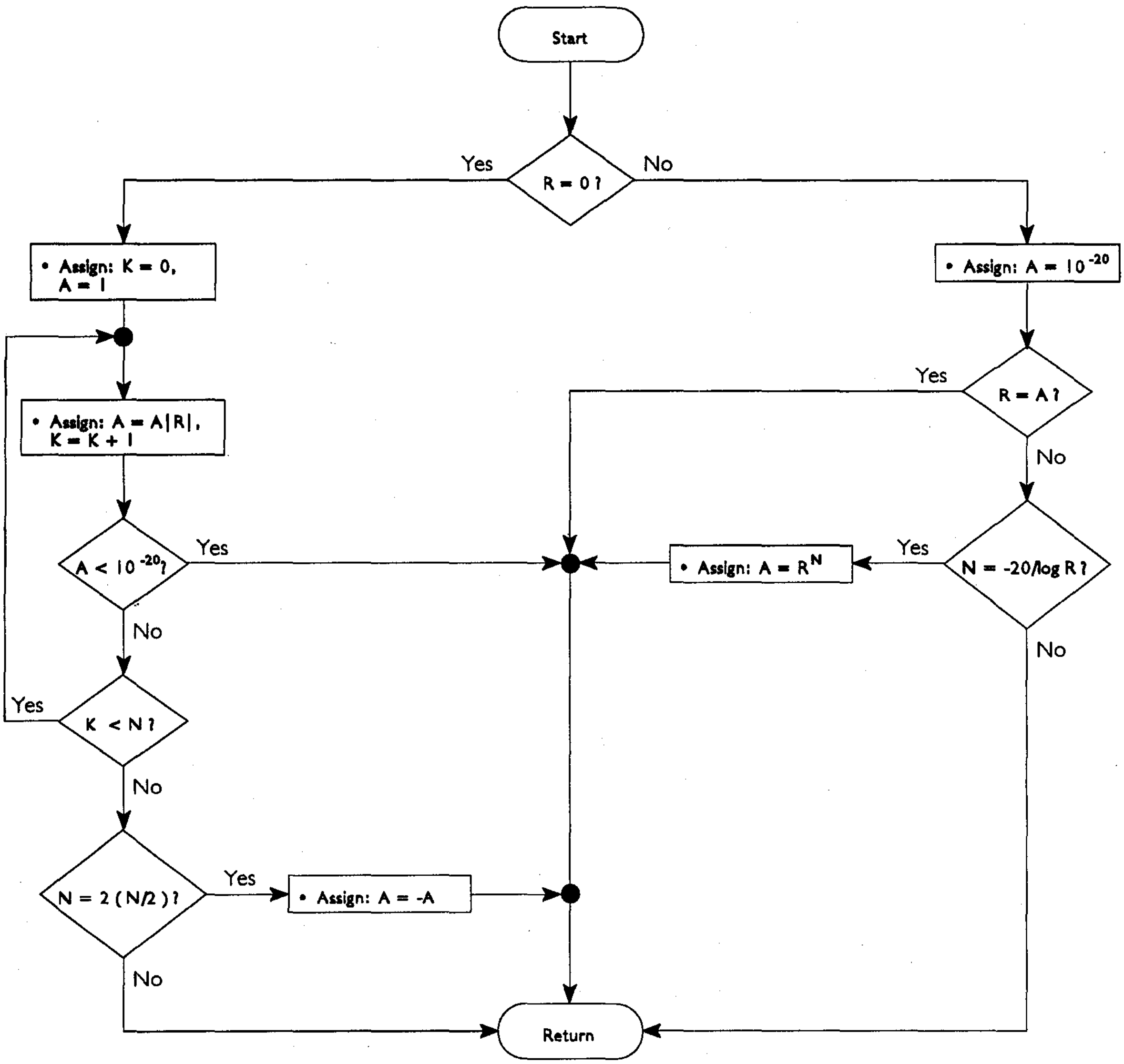

Figure B-4. Flowchart of subroutine limit.

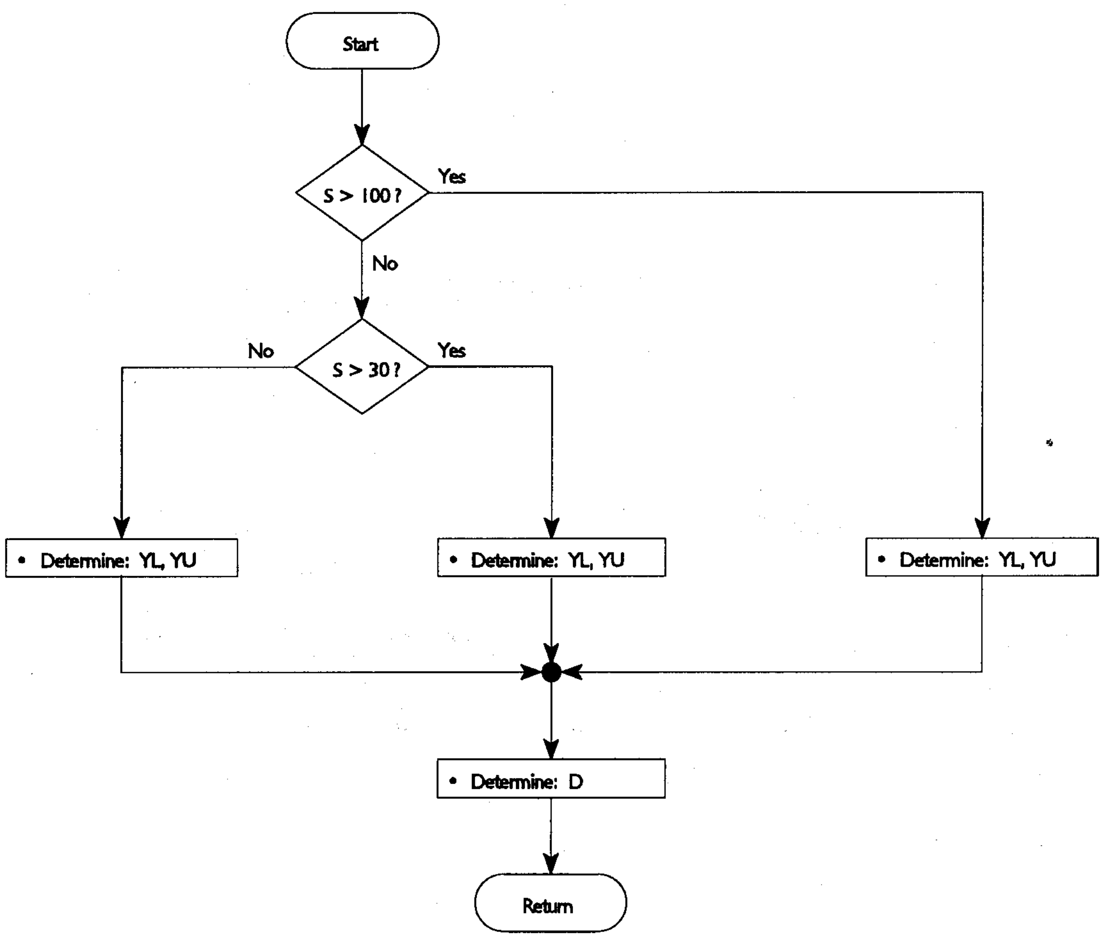

Figure B-5. Flowchart of subroutine poiss.

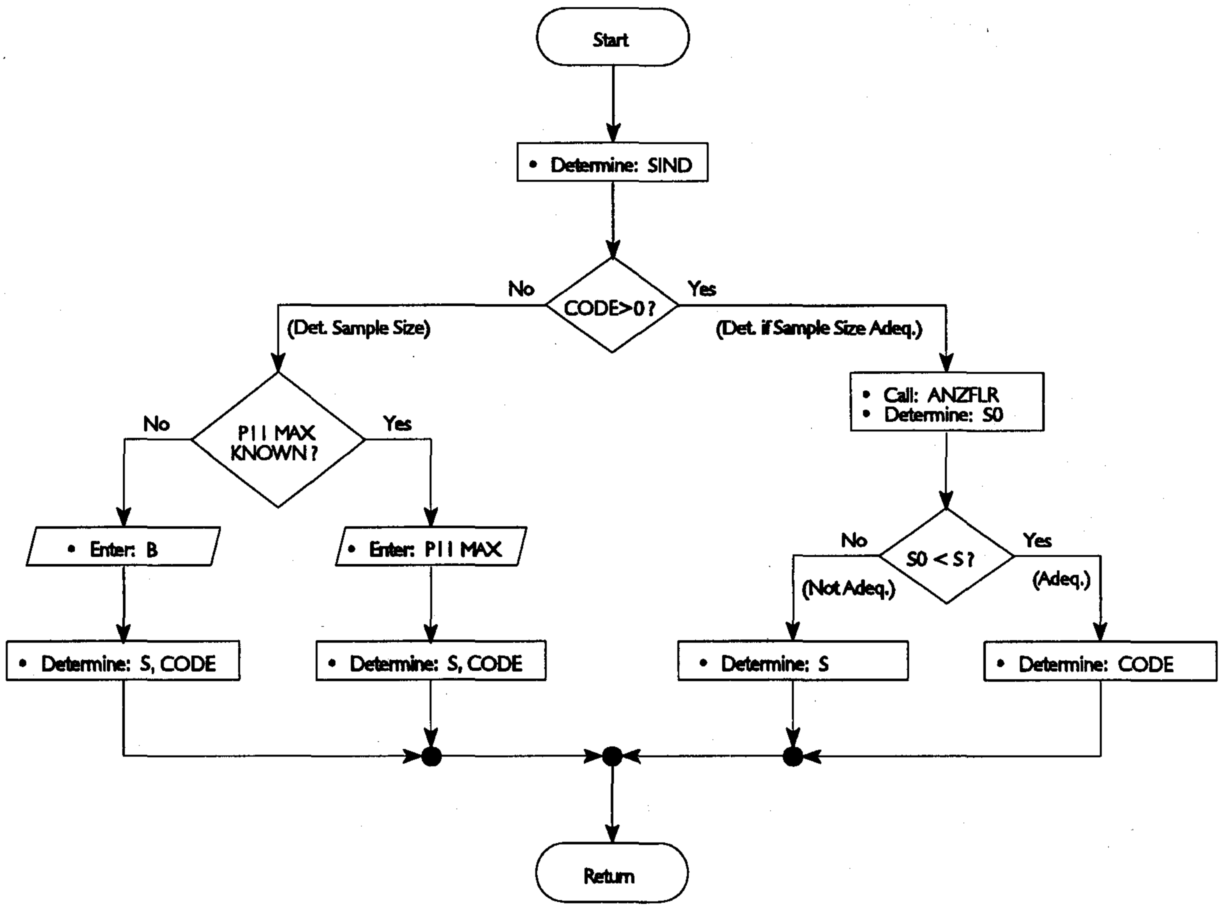

Figure B-6. Flowchart of subroutine **ssdflr**.

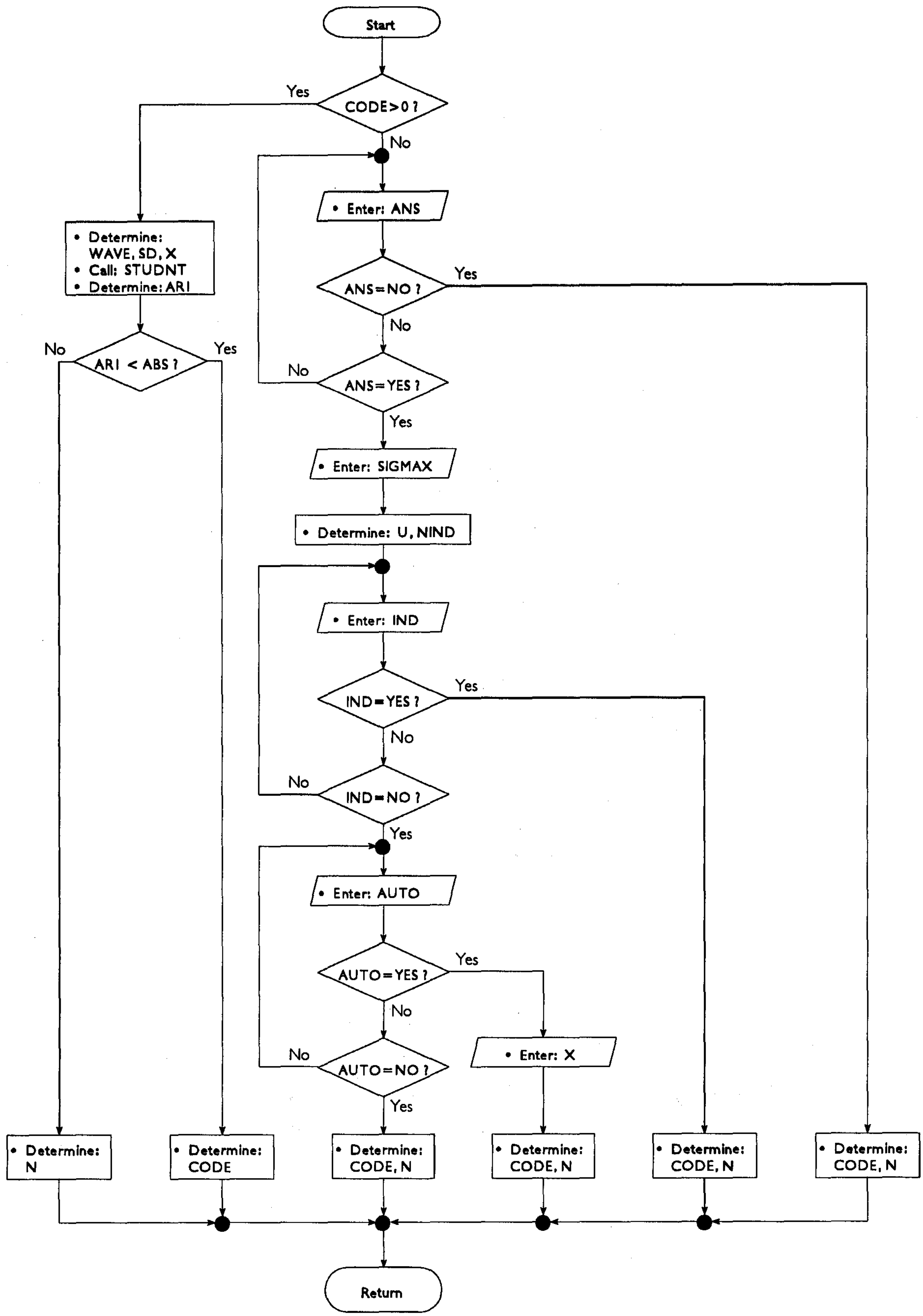

Figure B-7. Flowchart of subroutine ssdtim.

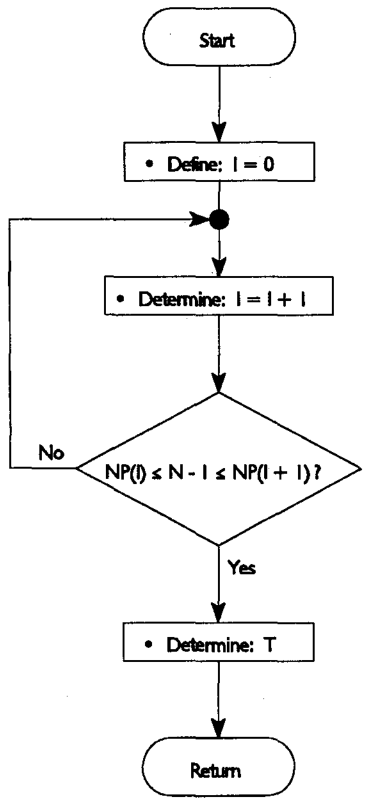

Figure B-8. Flowchart of subroutine studnt.

APPENDIX C: SHELL SCRIPT IMPLEMENTATION OF ANALYSIS OF A SINGLE TEST

This appendix shows how the performance outcomes from data reduction are used by **star** to analyze single tests. These performance outcome files are listed in Table C-1 and are described in Section 4 of Volume 4.[1]

Table C-1. File Names of Performance Outcomes

PERFORMANCE OUTCOMES	FILE
Access Outcomes	ACO
Source Disengagement Outcomes	D10
Dest. Disengagement Outcomes	D20
Bit Outcomes	B10
Block Outcomes	B20
Transfer Sample Outcomes	B30
Throughput Sample Outcomes	B40

The user activates the shell scripts **do** or **dopre** which process data through the analysis phase of single tests, until the measurement report summary file is created for both the 90% and 95% confidence levels.

These shell scripts also contain the shell scripts **reduc-a** and **reduc-x** which reduce the data according to Volume 4.

C.1 Shell Scripts for Time Parameters

Figure C-1 is a structured design diagram of shell script implementation of analysis of a single test for time parameters.

C.1.1 Access-Disengagement Tests

A single execution of the shell script **time-a** produces estimates and 90% and 95% confidence limits for access-disengagement time parameters. UNIX[tm] utilities edit the performance outcome files **ACO**, **D10**, and **D20**. They remove

[1]Throughput sample outcome files (i.e., B40) are not analyzed for single tests because they provide only one trial of the performance parameter, User Information Bit Transfer Rate, and its ancillary parameter, User Fraction of Input/Output Time.

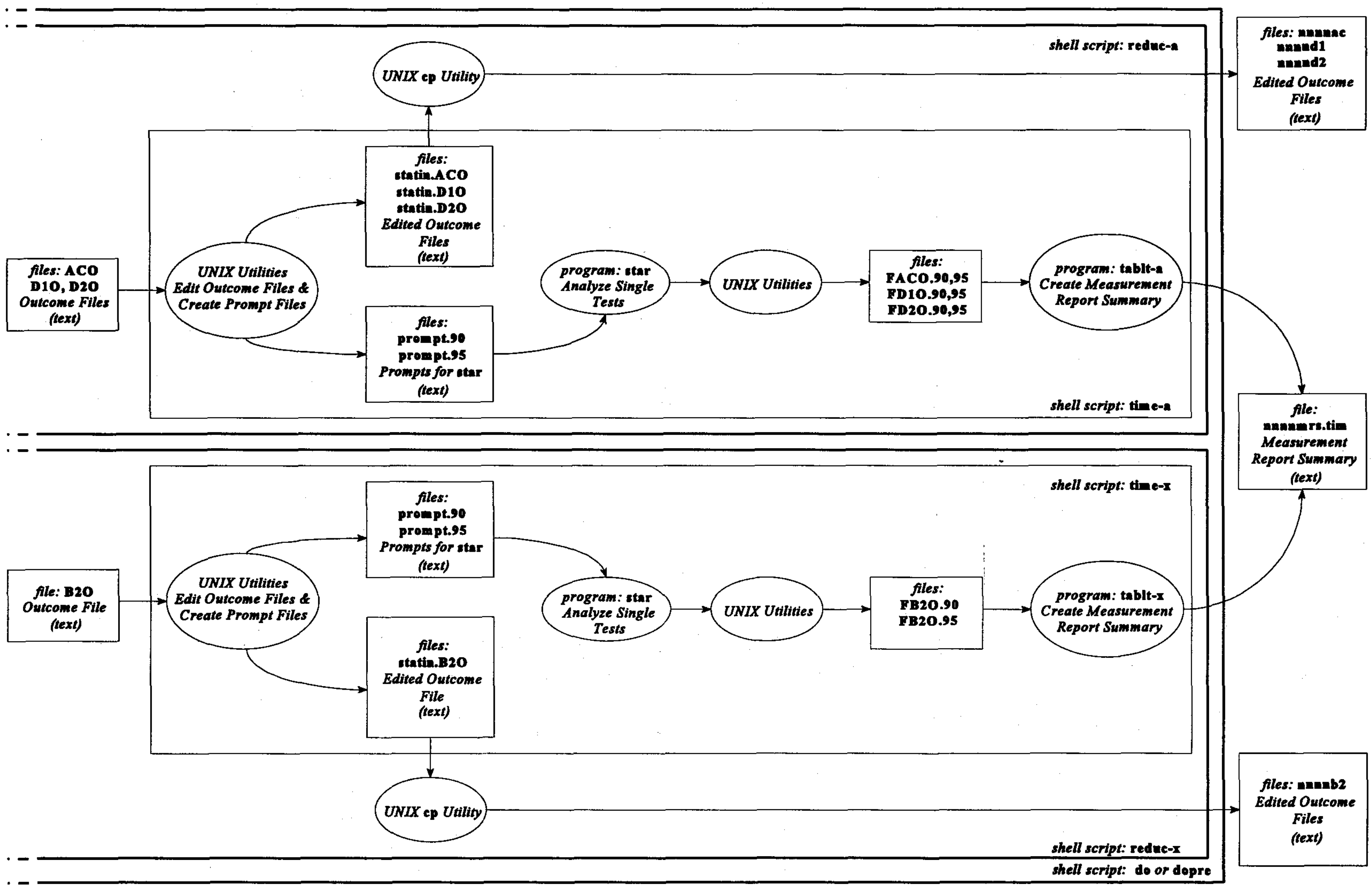

Figure C-1. Structured design diagram of analysis of single tests for time parameters.

header and trailer records, and all records having negative outcome codes (that represent failure outcomes), and store the results in files called **statin.ACO**, **statin.D1O**, and **statin.D2O**, respectively. A copy of these files (called **nnnnac**, **nnnnd1**, and **nnnnd2**, respectively) will be used for analysis of multiple tests. These UNIX^{tm} utilities also produce two files (**prompt.90** and **prompt.95**) that contain responses required by **star** to analyze tests at the 90% and 95% confidence level, respectively. These data and prompt files are used by **star** to analyze the performance parameters.

Other UNIX^{tm} utilities edit the output from **star** to produce the six temporary files **FACO.90**, **FACO.95**, **FD1O.90**, **FD1O.95**, **FD2O.90**, and **FD2O.95**.

The C program **tablt-a** operates on these files to produce the measurement results summary file, **nnnnmrs.tim**, an example of which is shown in Figure 2 (for test 2218).

C.1.2 User Information Transfer Tests

A single execution of the shell script **time-x** produces estimates and 90% and 95% confidence limits for user information transfer time parameters. UNIX^{tm} utilities edit the block outcome file, **B2O**. They remove header and trailer records, and all records with negative outcome codes (that represent failure outcomes), and store the results in a file called **statin.B2O**. A copy of this file (called **nnnnb2**) will be used for analysis of multiple tests. These UNIX^{tm} utilities also produce two files (**prompt.90** and **prompt.95**) that contain responses required by **star** to analyze tests at the 90% and 95% confidence level, respectively. These data and prompt files are used by **star** to analyze the performance parameters.

Other UNIX^{tm} utilities edit the output from **star** to produce the two temporary files **FB2O.90**, and **FB2O.95**.

The C program **tablt-x** operates on these files to produce the measurement results summary file, **nnnnmrs.tim**, an example of which is shown in Figure 3 (for test 2215).

C.2 Shell Scripts for Failure Probability Parameters

Figure C-2 is a structured design diagram of shell script implementation of a single test for failure probability parameters.

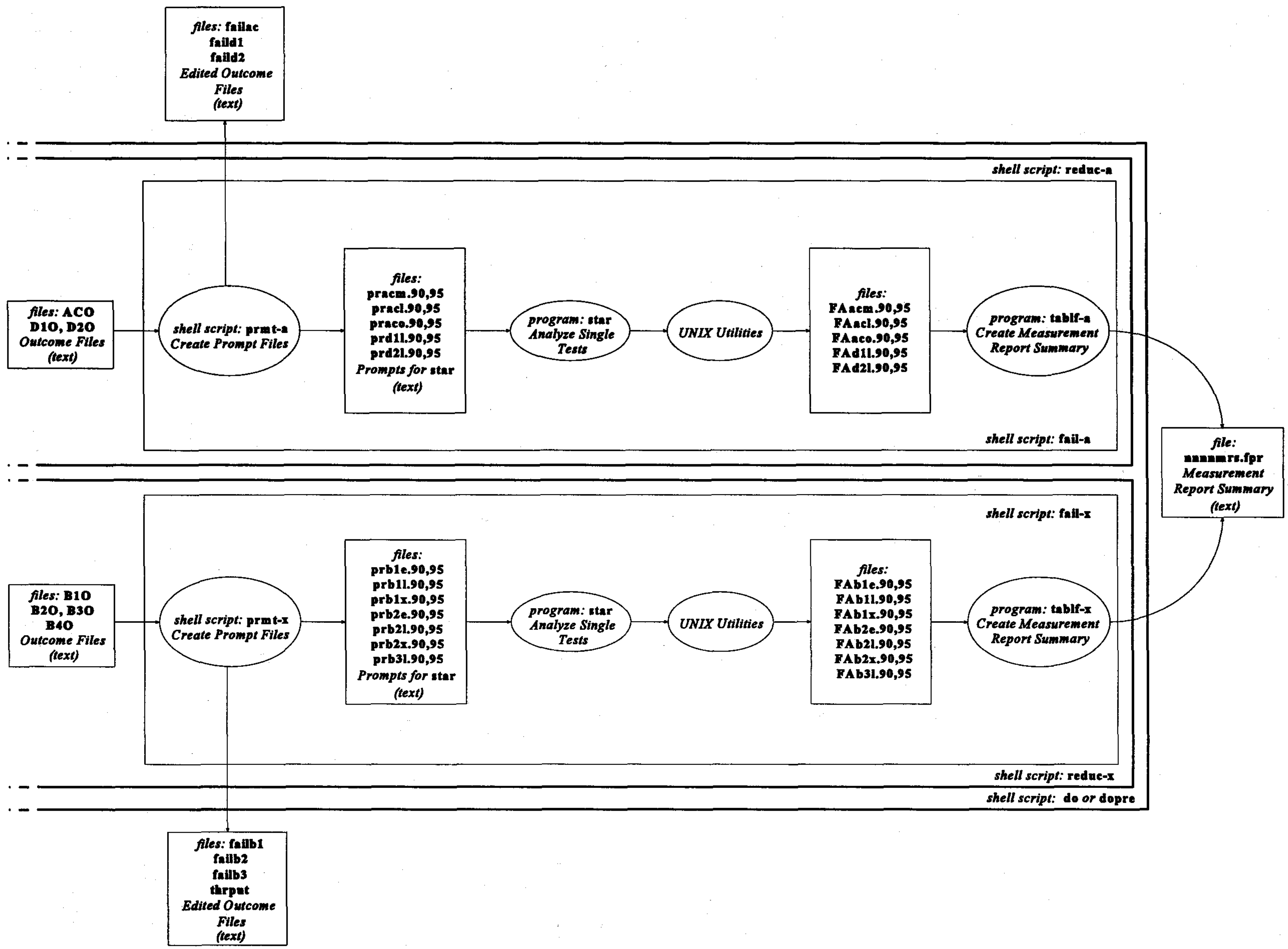

Figure C-2. Structured design diagram of analysis of single tests for failure probability parameters.

C.2.1 Access-Disengagement Tests

A single execution of the shell script **fail-a** produces estimates and 90% and 95% confidence limits for all access-disengagement failure probability parameters.

fail-a first calls the shell script **prmt-a** to generate two sequences of prompt files that contain keyboard responses to **star** program prompts: One sequence of five prompt files is used for estimating 90% confidence limits, and the other sequence of five prompt files is used for estimating 95% confidence limits. **prmt-a** uses UNIX^{tm} utilities to edit the access-disengagement outcome files (**ACO**, **D1O**, and **D2O**); it removes header and trailer records and stores the results in temporary files (**ac**, **d1**, and **d2**, respectively). Key items in the prompt files are the number of trials, the number of failures, and the number of pairs of consecutive failures. **prmt-a** extracts the number of trials and the number of failures and calls a C program (**countum**) to determine the number of pairs of consecutive failures by examining the failure outcomes recorded in the relevant performance outcome file.

prmt-a also produces the files **failac**, **faild1**, and **faild2** for analysis of multiple tests.

fail-a then calls **star** twice for each estimated parameter; the first call obtains the 90% confidence limits, and the second call obtains the 95% confidence limits. Output from **star** is piped through the UNIX^{tm} **tail** utility. It deletes all output except the final statement of results.

fail-a concludes by calling the C program **tablf-a** to generate the measurement results summary file, **nnnnmrs.fpr**, an example of which is shown in Figure 4 (for test 2218).

C.2.2 User Information Transfer Tests

A single execution of the shell script **fail-x** produces estimates and 90% and 95% confidence limits for all user information transfer failure probability parameters.

fail-x first calls the shell script **prmt-x** to generate two sequences of prompt files that contain keyboard responses to **star** program prompts: one sequence of seven prompt files is used for estimating 90% confidence limits and the other sequence of seven prompt files is used for estimating 95% confidence limits. **prmt-x** uses UNIX^{tm} utilities to edit the user information transfer outcomes files (**B10**, **B20**, and **B30**); it removes header and trailer records and stores the results in temporary files (**b1**, **b2**, and **b3**, respectively). Key items in the prompt files are the number of trials, the number of failures, and the number of pairs of consecutive failures. **prmt-x** extracts the number of trials and the number of failures and calls a C program (**countum**) to determine the number of pairs of consecutive failures by examining the failure outcomes recorded in the relevant performance outcome file.

prmt-x also produces the files **failb1**, **failb2**, **failb3**, and **thrput** for analysis of multiple tests.

fail-x then calls **star** twice for each estimated parameter; the first call obtains the 90% confidence limits, and the second call obtains the 95% confidence limits. Output from **star** is piped through the UNIX^{tm} **tail** utility. It deletes all output except the final statement of results.

fail-x concludes by calling the C program **tablf-x** to generate the measurement results summary file, **nnnnmrs.fpr**, an example of which is shown in Figure 5 (for test 2215).

APPENDIX D: OPERATOR IMPLEMENTATION OF ANALYSIS OF A SINGLE TEST

Figure D-1 is an operator-decision diagram for analysis of single tests by program **star**. It shows the operator decisions required for analysis of each of the nine possible scenarios (labelled A through I). If **star** was accessed earlier to determine the sample size required to achieve a specified precision, a code number for subsequent analysis was assigned.[1] Otherwise, obtain the code number from Table D-1. This table contains

- the adequacy of the sample size (as dictated by budget/time or as determined by the desired precision),
- the confidence level,
- the code numbers (i.e., 11 through 34), and
- the test labels (i.e., A through I).

Section D.1 discusses analysis of the time parameters (i.e., delays and rates). Performance data for time parameters can be entered either by a keyboard or by a file.

Section D.2 discusses analysis of the failure probability parameters. Performance data for failure probability parameters can be entered by a keyboard only.

From /usr/data/5a, type

star

and select your test number from Table D-1.

D.1 Time Parameters

There are three possible tests of delay parameters (tests A, B, and C) and three possible tests of the rate parameter (tests D, E, and F):

- Tests A and D result from specifying a sample size.

[1]This number will direct **star** to the proper analysis formulas.

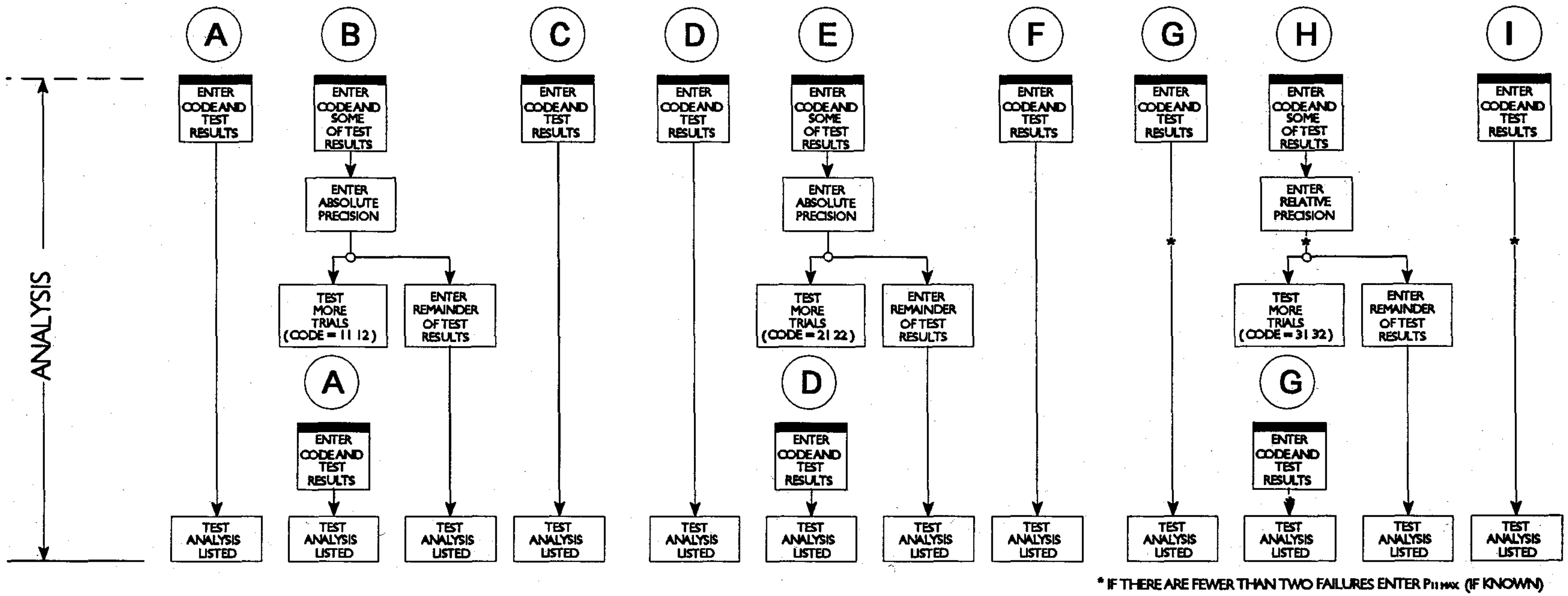

Figure D-1. Operator-decision diagram for analysis of single tests.

Table D-1. Code Numbers and Corresponding Test Labels Resulting from Sample Size Determination

	SAMPLE SIZE KNOWN TO BE ADEQUATE BEFORE TEST		SAMPLE SIZE NOT KNOWN TO BE ADEQUATE BEFORE TEST	
	Confidence Level		Confidence Level	
	90%	95%	90%	95%
Delays	11 - - - - - A or C	12 - - - - - A or C	13 - - - - - B	14 - - - - - B
Rate	21 - - - - - D or F	22 - - - - - D or F	23 - - - - - E	24 - - - - E
Failure Probability	31 - - - - - G or I	32 - - - - - G or I	33 - - - - - H	34 - - - - - H

- Tests B and E result from specifying a desired absolute precision and either not knowing the maximum standard deviation of the delays (Input/Output Time for rates) or realizing some statistical dependence exists but not knowing the autocorrelation of lag 1.

- Tests C and F result from specifying a desired absolute precision, knowing the maximum standard deviation, and knowing the trials are statistically independent.

Figures D-2, D-3, and D-4 are subdiagrams of the operator-decision diagram (Figure D-1); they show the sequence of events that results in analysis of delays.

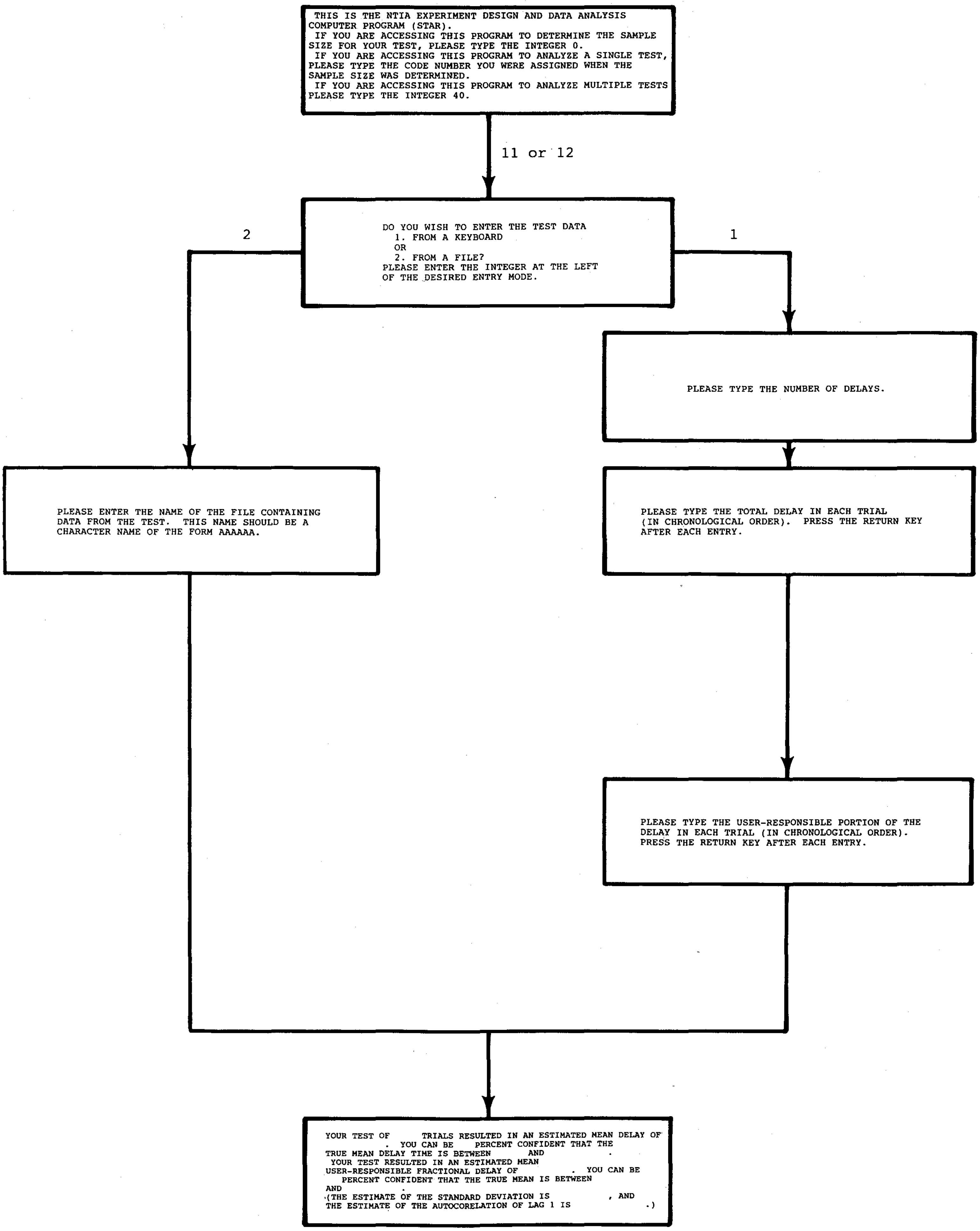

Figure D-2. Program messages for analysis of delays when budget is the criterion (test A).

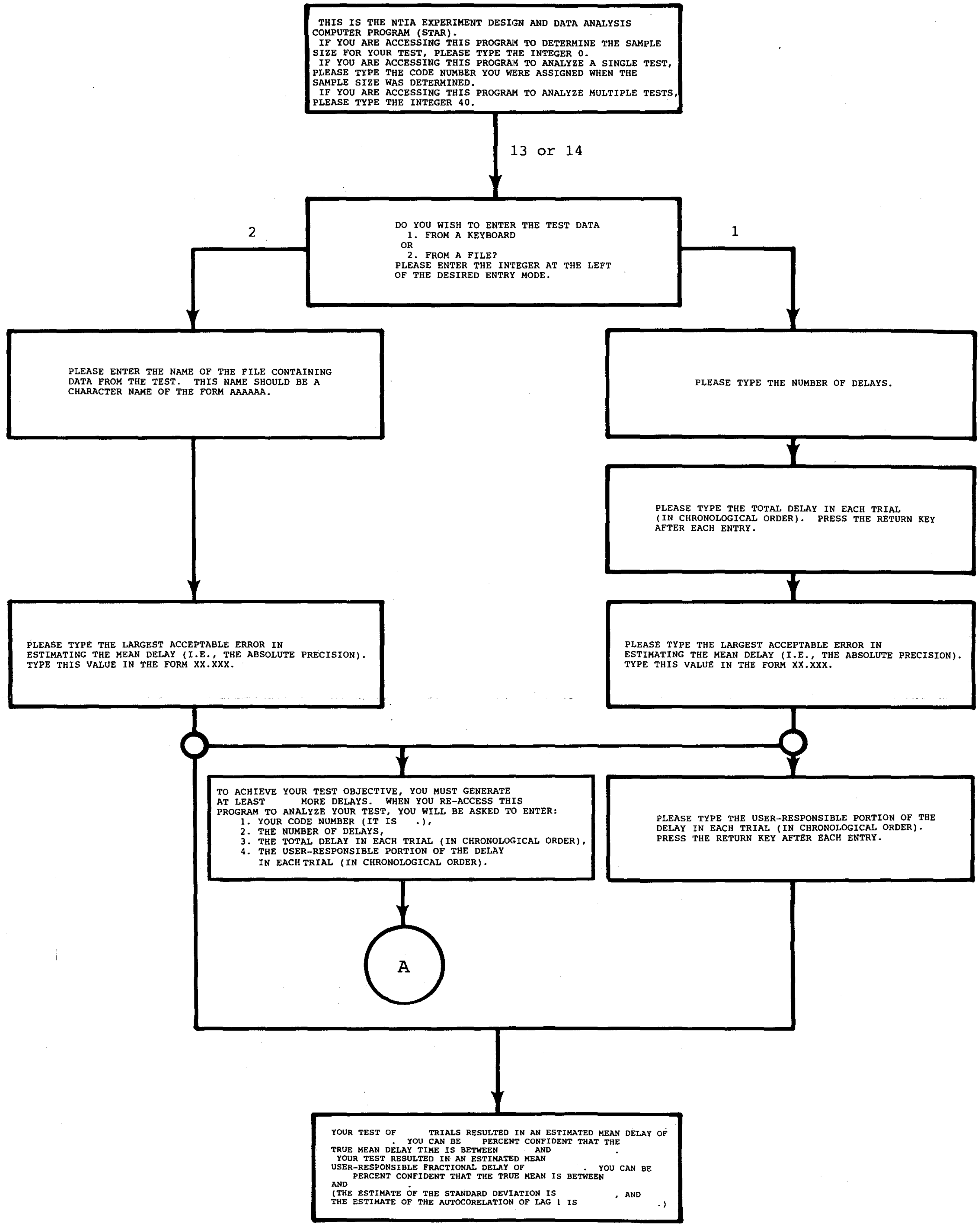

Figure D-3. Program messages for analysis of delays when precision is the criterion and the population is not known (test B).

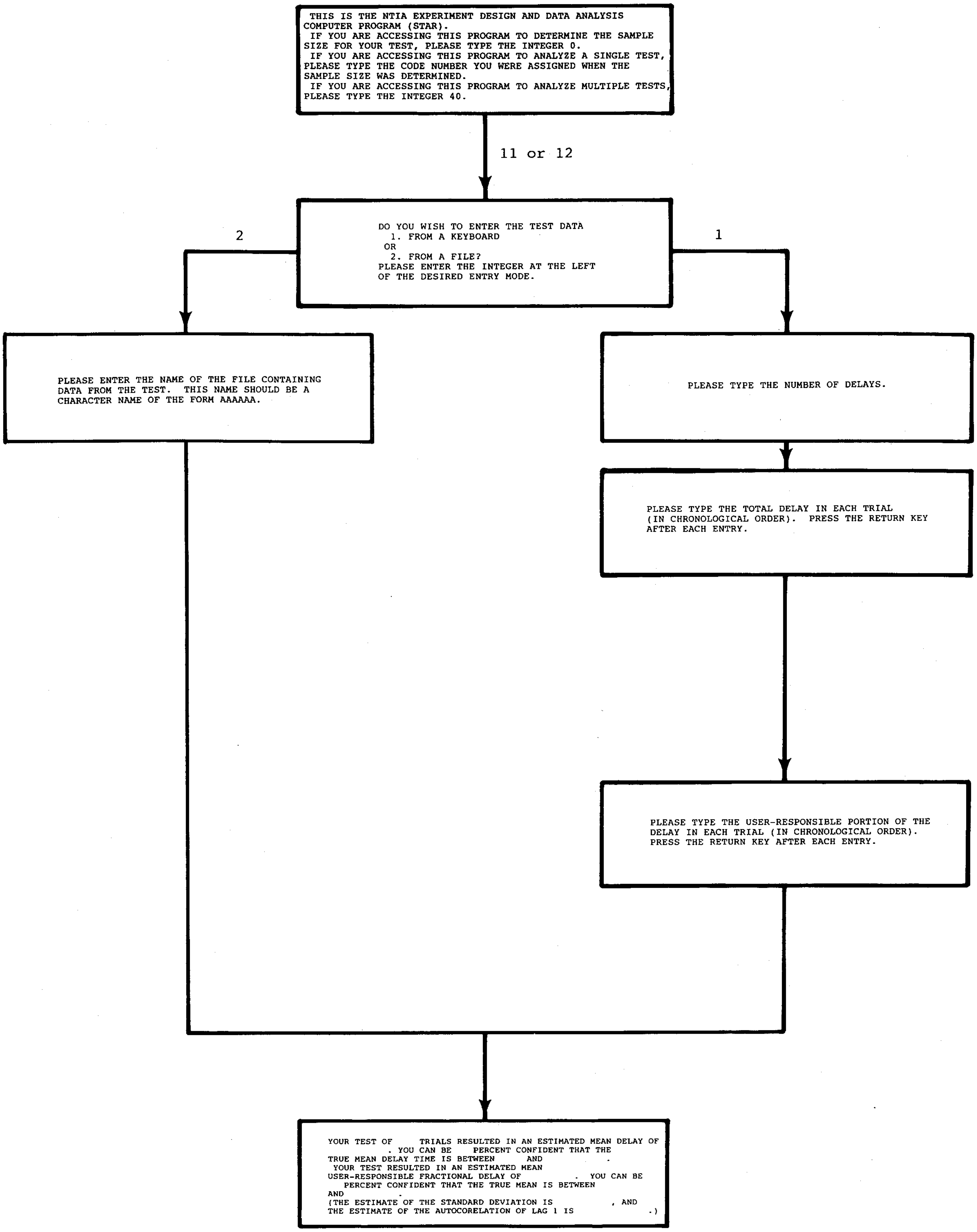

Figure D-4. Program messages for analysis of delays when precision is the criterion and the population is known (test C).

Test A

Enter the Following:

- Code Number. Enter 11 if the 90% confidence level was selected when the sample size was determined or 12 if the 95% confidence level was selected.

- Mode of Data Entry. Enter 1 if data are to be entered from a keyboard or 2 if data are to be entered from a file.

- If data are to be entered from a keyboard, enter:

    - Number of Total delays. This is an integer greater than 1. star is designed for as many as 200 total delays.

    - Total Delay in Each Trial. This is a positive decimal number in the form XXXXXX.XXX; not all delays can be equal. Enter the delays in chronological order, including the decimal point.

    - User-Portion of the Total Delay in Each Trial. The same restrictions apply as for the total delay. The units must be the same as those of the total delays.

- If data are to be entered from a file, enter:

    - File Name. The name should be a character name of the form **aaaaaa**. The file format must be 2F16.3. Columns one and two contain the total delays and the user-portions of the total delays, respectively.

Analysis consists of

- the estimate of the mean delay and its confidence limits, and

- the estimate of the mean user-fraction of the total delay and its confidence limits.

Test B

Enter the following:

- Code Number. Enter 13 if the 90% confidence level was selected when the sample size was determined or 14 if the 95% confidence level was selected.

- Mode of Entry. Enter 1 if data are to be entered from a keyboard or 2 if data are to be entered from a file.

- If data are to be entered from a keyboard, enter:

    - Number of Trials. This is an integer greater than 1. **star** is designed for as many as 200 trials.

    - Total Delay in Each Trial. This is a positive decimal number in the form XXXXXX.XXX, not all equal. Enter the Input/Output Time in chronological order, and include the decimal point.

    - Absolute Precision. This is a positive decimal number as defined in Section 1.2.4 of Volume 2. Depending upon the total delay, the confidence level, and the absolute precision, **star** determines whether more trials are required.

      (If no more trials are required, enter:)

    - User-Portion of the Total Delay in Each Trial. The same restrictions apply as for the total delays. The units must be the same as those of the total delays.

- If data are to be entered from a file, enter:

    - File Name. The name should be a character name of the form **aaaaa**. The file format must be 2F16.3. Columns one and two contain the total delays and the user-portions of the total delays, respectively. The data are to be listed in chronological order.

    - Absolute Precision. This is a positive decimal number as defined in Volume 2. Depending upon the total delays, the confidence level, and the absolute precision, **star** determines whether more trials are required.

If no more trials are required, analysis consists of

- the estimate of the mean delay and its confidence limits, and
- the estimate of the mean user-fraction of the total delay and its confidence limits.

On the other hand, if more trials are required, analysis consists of

- determining the number of additional trials required, and
- assigning a new code number for the next analysis (11 for the 90% confidence level or 12 for the 95% confidence level).

The number of required trials is now known (i.e., the total from the preliminary Test B and this one). Hence, after this test, this re-entry code will cause analysis to proceed as Test A (which results from specifying that the sample size is sufficient).

Test C

Enter responses and data as in Test A.

Figures D-5, D-6, and D-7 are subdiagrams of the operator-decision diagram (Figure D-1); they show the sequence of events that results in analysis of rates.

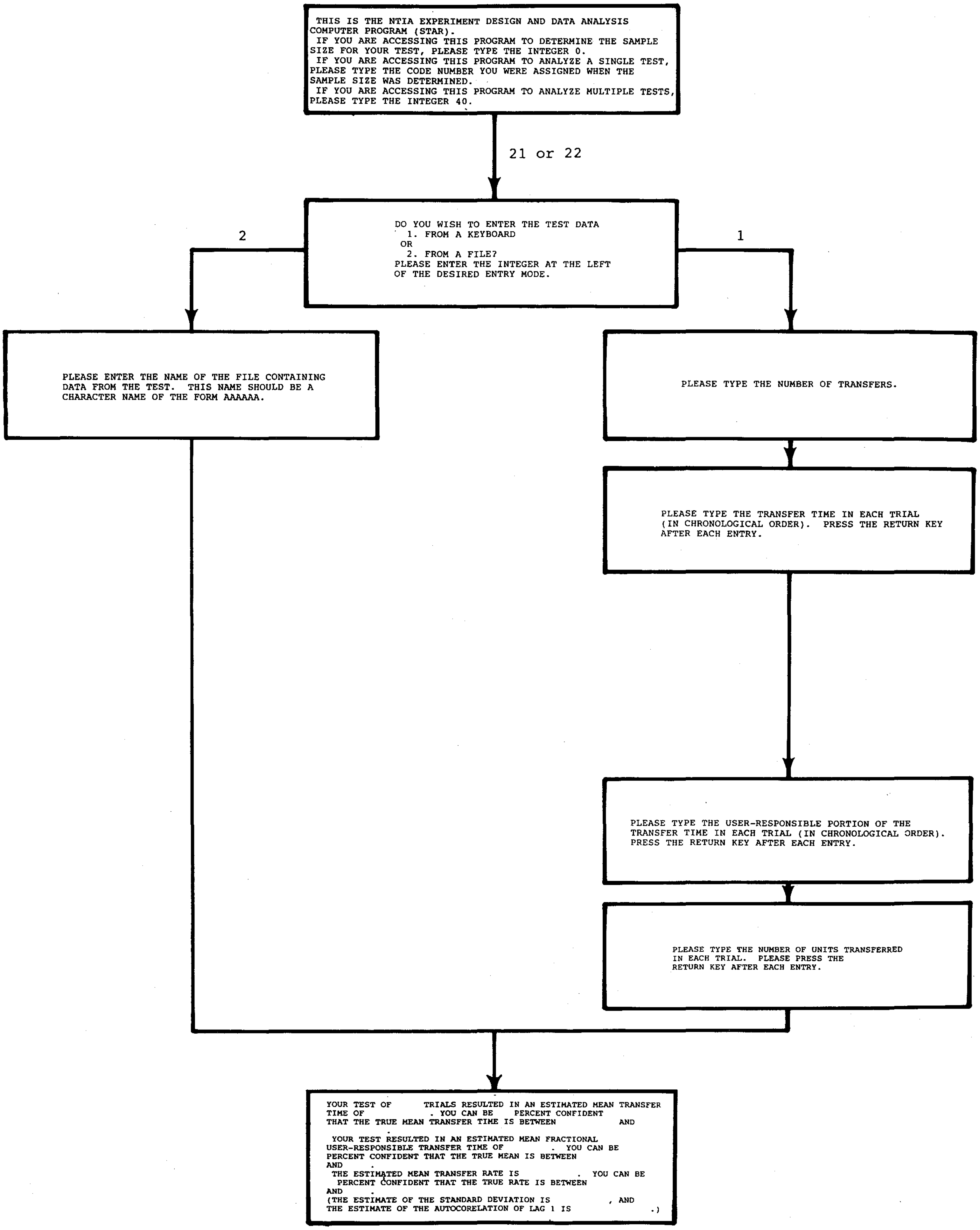

Figure D-5. Program messages for analysis of rates when budget is the criterion (test D).

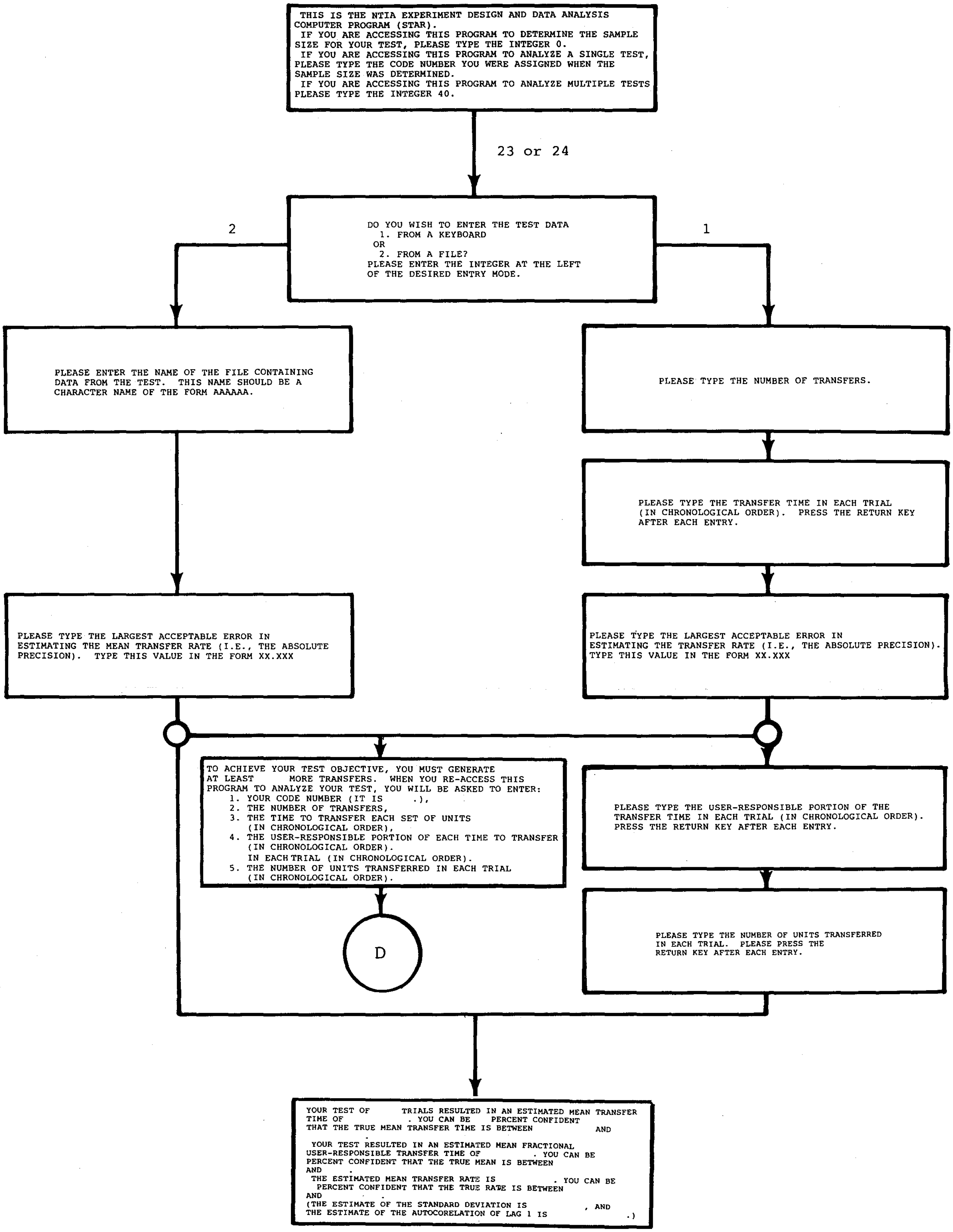

Figure D-6. Program messages for analysis of rates when precision is the criterion and the population is not known (test E).

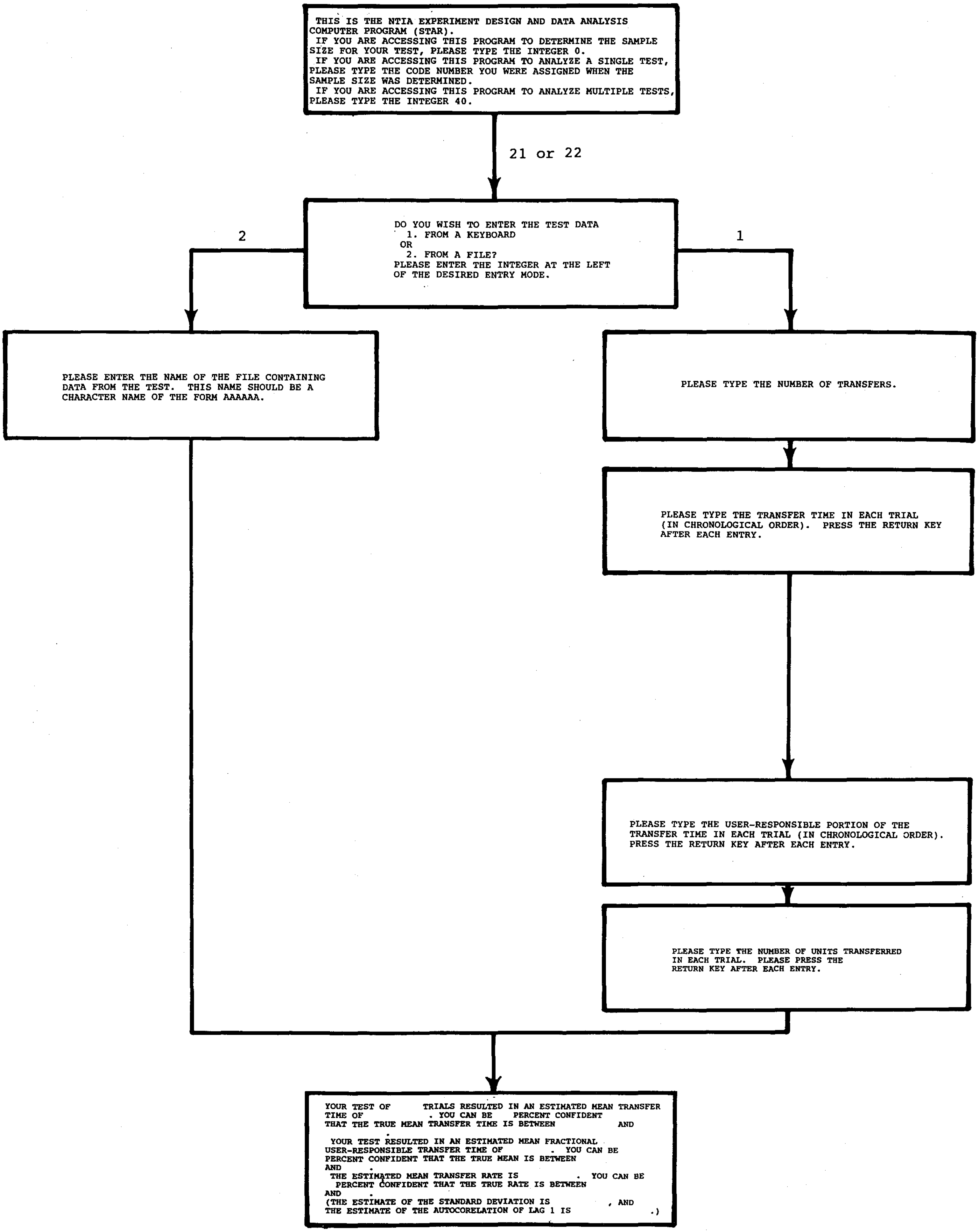

Figure D-7. Program messages for analysis of rates when precision is the criterion and the population is known (test F).

Test D

- Code Number. Enter 21 if the 90% confidence level was selected when the sample size was determined or 22 if the 95% confidence level was selected.

- Mode of Entry. Enter 1 if data are to be entered from a keyboard or 2 if data are to be entered from a file.

- If data are to be entered from a keyboard, enter:

    - Number of Trials. This is an integer greater than 1. **star** is designed for as many as 200 trials.

    - Input/Output Time in Each Trial. This is a positive decimal number in the form XXXXXX.XXX, not all equal. Enter the Input/Output Times in chronological order, and include the decimal point.

    - User-Portion of the Total Input/Output Time in Each Trial. The same restrictions apply as for the total Input/Output Times. The units must be the same as those of the total Input/Output Times.

    - Number of Bits Transferred in Each Trial. This is a positive integer having from one to ten digits.

- If data are to be entered from a file, enter:

    - File Name. The name should be a character name of the form **aaaaaa**. The file format must be 2F16.3, F16.0. Columns one, two, and three contain the total Input/Output Time in each trial, the user-fraction of the total Input/Output Time in each trial, and the number of bits transferred in each trial, respectively. The data are to be entered in chronological order.

Analysis consists of

- the estimate of the mean Input/Output Time and its confidence limits,

- the estimate of the mean User-Fraction of Input/Output Time and its confidence limits, and

- the estimate of the mean User Information Bit Transfer Rate and its confidence limits.

Test E

Enter the following:

- Code Number. Enter 23 if the 90% confidence level was selected when the sample size was determined or 24 if the 95% confidence level was selected.

- Mode of Entry. Enter 1 if data are to be entered from a keyboard or 2 if data are to be entered from a file.

- If data are to be entered from a keyboard, enter:

    - Number of Trials. This is an integer greater than 1. **star** is designed for as many as 200 trials.

    - Input/Output Time in Each Trial. This is a positive decimal number in the form XXXXXX.XXX, not all equal. Enter the Input/Output Times in chronological order, and include the decimal point.

    - Absolute Precision. This is a positive decimal number as defined in Section 1.2.4 of Volume 2. Depending upon the Input/Output Times, the confidence level, and the absolute precision, **star** determines whether more trials are required.

      (If no more trials are required, enter:)

    - User-Portion of the Input/Output Time in Each Trial. The same restrictions apply as for the Input/Output Time.

    - Number of Bits Transferred in Each Trial. This is a positive integer having from 1 to 8 digits.

(continued on next page)

Test E (continued)

Continue entering:

- If data are to be entered from a file, enter:

    - File Name. The name should be a character name of the form **aaaaaa**. The file format must be 2F16.3, F8.0. Columns one, two, and three contain the Input/Output Time in each trial, the user-portion of the Input/Output Time in each trial, and the number of bits transferred in each trial, respectively.

    - Absolute Precision. This is a positive decimal number as defined in Volume 2. Depending upon the total delays, the confidence level, and the absolute precision, **star** determines whether more trials are required.

If no more trials are required, analysis consists of

- the estimate of the mean Input/Output Time and its confidence limits. The estimate of the mean User Fraction of Input/Output Time and its confidence limits, and

- the estimate of the mean User Information Bit Transfer Rate and its confidence limits.

On the other hand, if more trials are required, analysis consists of

- determining the number of additional trials required, and

- assigning a new code number for the next analysis (21 for the 90% confidence level or 22 for the 95% confidence level).

The number of required trials is now known (i.e., the total from the preliminary Test E and this one). Hence, after the test, this re-entry code will cause analysis to proceed as after Test D, which results from specifying that the sample size is sufficient.

Test F

Enter responses and data as in Test D.

Example: The test as in the example in Appendix C of Volume 2 (Test C) has been conducted. It produced the following 13 delays:

5., 7., 6., 5., 4., 5., 8., 5., 6., 7., 6., 6., and 5.

It also produced the following 13 user-responsible portion of delays:

3., 4., 4., 4., 2., 3., 5., 3., 3., 4., 5., 4., and 3.

Enter the data from a keyboard, and analyze the test.

Solution:

- Type, **star**.
- Type 11 (the assigned code number), and press the return key.
- Type 1 (for keyboard entry), and press the return key.
- Type 13 (the number of delays tested), and press the return key.
- Type 5., and press the return key.
  Type 7., and press the return key.
  ...
  Type 5., and press the return key.
- Type 0 (since all delays were entered correctly), and press the return key.
- Type 3., and press the return key.
- Type 4., and press the return key.
  Type 4., and press the return key.
  ...
  Type 3., and press the return key.

The following analysis of the test is listed:

Your test resulted in an estimated mean delay of .57692E+01. You can be 90 percent confident that the true mean delay is between .52757E+01. and .62627E+01.

Your test resulted in an estimated mean user-fraction delay of .62664E+00. You can be 90 percent confident that the true mean is between .58465E+00 and .66864E+00.[2]

D.2 Failure Probability Parameters

There are three possible tests of failure probability (Tests G, H, and I):

- Test G results from specifying a given sample size.
- Test H results from specifying a desired relative precision but not knowing the conditional probability, λ.
- Test I results from specifying the desired relative precision and knowing the maximum value of the conditional probability, λ.

Figure D-8 is an operator-decision subdiagram of Figure D-1. It shows the sequence of events leading to analysis of failure probability for tests G, H, and I. Since analysis of the failure probability requires entry of the code number and only three other numbers there is no provision for entry from a file.

Test G

Enter the following:

- Code Number. Enter 31 if the 90% confidence level was selected when the sample size was determined or 32 if the 95% confidence level was selected.
- Number of Trials. This is a positive integer having from 1 to 8 digits.
- Number of Failures. This is an integer from zero to one less than the sample size.
- Number of Pairs of Consecutive Failures. This is an integer from zero to one less than the number of failures. (However, enter zero if there are also zero failures.)

[2] The confidence limits are closer to the estimate of the mean than the specified 0.7 seconds (i.e., 0.49) because the sample standard deviation of the delays is 1.05 (i.e., smaller than the 1.5 maximum entered when the sample size was to be determined).

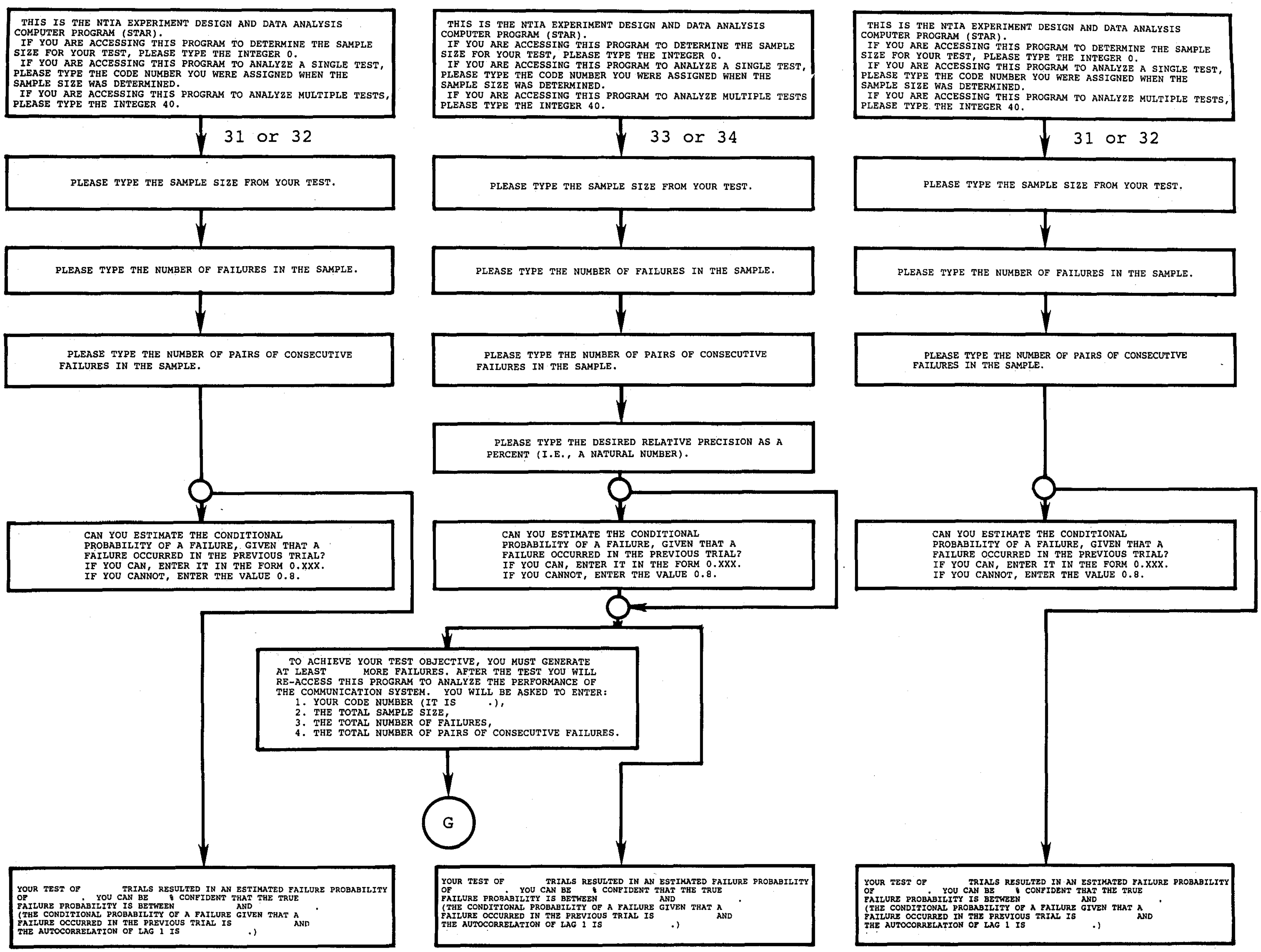

Figure D-8. Program messages for analysis of failure probabilities (tests G, H, and I).

Analysis consists of

- the estimate of the mean failure rate and its confidence limits.

Test H

Enter the following:

- Code Number. Enter 33 if the 90% confidence level was selected when the sample size was determined or 34 if the 95% confidence level was selected.

- Number of Trials. This is a positive integer having from 1 to 8 digits.

- Number of Failures. This is an integer from zero to one less than the sample size.

- Number of Pairs of Consecutive Failures. This is an integer from zero to one less than the number of failures. (However, enter zero if there are also zero failures.)

- Relative Precision. This is a one- or two-digit positive integer as defined in Section 1.2.4 of Volume 2.

If the number of failures is fewer than 2, enter (if known):

- Conditional Probability of a Failure, Given that a Failure Occurred in the Previous Trial. This is a number between 0 and 1.

If no more trials are required, analysis consists of

- the estimate of the mean failure rate and its confidence limits.

On the other hand, if more trials are required, analysis consists of

- determining the number of additional trials required, and

- assigning the new code number (31 for the 90% confidence level or 32 for the 95% confidence level).

The number of required trials is now known (i.e., the total from the preliminary Test H and this one). Hence, after this test, this re-entry code will cause analysis to proceed as in Test G (which results from knowing, initially, that the sample size is sufficient).

Test I

Enter responses and data as in Test G.

Example: The test from the example in Section 8.2.2 of Volume 2 has been conducted (Test H). It resulted in 752,650 trials, 17 failures, and three pairs of consecutive failures. The specified relative precision was 30%. Analyze the test data.

Solution:

- Type, **star**.
- Type, **33** (the assigned code number), and press the return key.
- Type, **752650** (the number of trials), and press the return key.
- Type, **17** (the number of failures), and press the return key.
- Type, **3** (the number of pairs of consecutive failures), and press the return key.
- Type, **30** (the relative precision expressed as percent), and press the return key.

The following analysis of the test is listed:

To achieve your test objective, you must generate at least 27 more failures. After the test you will re-access this program to analyze the performance of your communication system. You will be asked to enter:

- **Your code number (it is 31),**
- **The total sample size,**
- **The total number of failures,**
- **The total number of pairs of consecutive failures.**

Example (continued): The second test has been conducted (now because the sample size is known, i.e., test G). It resulted in 2,249,012 additional trials, 50 additional failures, and 8 additional pairs of consecutive failures. Analyze the combined data from both tests.

Solution:

- Type, **star.**
- Type, **31** (the assigned code number), and press the return key.
- Type, **3001662** (i.e., the total number of trials is 752650 + 2249012), and press the return key.
- Type, **67** (i.e., the total number of failures is 17 + 50), and press the return key.
- Type, **11** (i.e., the total number of pairs of consecutive failures is 3 + 8), and press the return key.

The following analysis of the test is listed:

> **Your test resulted in an estimated failure rate of .22321E-04. You can be 90 percent confident that the true failure rate is between .17340E-04 and .28342E-04.**

The relative precision achieved is 24.6%, better than the specified 30%.

APPENDIX E: FORMULAS FOR ANALYSIS OF MULTIPLE TESTS

Performance parameters of data communication systems are often affected by one or more variable conditions. Those variable conditions that do not affect them are called factors. Therefore, an experiment is usually designed to determine whether and how the variable conditions affect the performance parameter. Any experiment design will require multiple tests at various levels of each variable condition. The subject of this analysis is to learn as much as possible about the performance parameter from the multiple tests.

Specifically, the analysis will determine if the trials from multiple tests come from the same population. If they do, the trials can be pooled, and the larger sample will usually provide more information about the performance parameter (i.e., the larger number of trials tends to cause the sampling variance to be smaller and the degrees of freedom to be larger - both of which contribute to a shorter confidence interval).[1] If the trials do not come from the same population, analysis will determine if the test means come from the same population. If they do, they will be pooled to form a larger sample (but smaller than if the trials could be combined). If the test means do not come from the same population, the level means will be pooled to form, yet, a smaller sample. However, there is no test to determine whether the level means come from the same population. This procedure is depicted in Figure 9.

The first section of this appendix introduces the mathematical model for the variation from trials, tests, and levels. The second section analyzes time parameters, and the third section analyzes failure probability parameters.

---

[1]However, the autocorrelation will have modifying effects on the amount of information as measured by precision:

- Negative Autocorrelation. If the autocorrelation is negative, both the sampling variance and the degrees of freedom will be larger than if it were zero.

- Positive Autocorrelation. If the autocorrelation is positive, both the sampling variance and the degrees of freedom will be smaller than if it were zero.

E.1 Linear Model for Analysis of Variance

E.1.1 The Linear Model

For a specified variable condition, assume the population mean is μ and there are three sources of variation: variation among levels, tests, and trials.[2] Assume the levels and the trials have been chosen randomly.

A. Levels

Suppose there are $l = 1, 2, \ldots, N''$ levels for the specified variable condition, and a_l is the variation in the lth level.

B. Tests

Suppose there are $m = 1, 2, \ldots, N_l$ tests in the lth level of the specified variable condition, and b_{lm} is the variation in the mth test in the lth level. The number of tests over the N" levels is

$$N' = \sum_{l=1}^{N''} N_l \,.$$

C. Trials

Suppose there are $n = 1, 2, \ldots, N_{lm}$ trials in the mth test in the lth level, and c_{lmn} is the variation in the nth trial in the mth test in the lth level. Assume that the variations in any two trials, c_{lmn} and $c_{lmn'}$, are stationary. The number of trials over N_l tests and N" levels is

$$N = \sum_{l=1}^{N''} \sum_{m=1}^{N_l} N_{lm} \,.$$

Now, assume that the variations are additive with equal variances (σ_a, σ_b, and σ_c), and let

$$x_{lmn} = \mu + a_l + b_{lm} + c_{lmn}$$

be the linear model.

---

[2]There is also variation from each factor (for that is the property of a factor). However, if the model for pooling included the variation from, say, n factors each trial would be used n times in the formulas.

E.1.2 The Linear Model and Hypothesis Tests

This linear model will be used with hypothesis tests in the next two sections to determine pooling of data for time parameters and failure probability parameters.

These sections will use hypothesis tests in the following ways:

- Trials. Determine if all trials can be pooled. The large degree(s) of freedom from this pooling provides the narrowest confidence interval. If they can be pooled,

$$\sigma_a = \sigma_b = 0.$$

- Test Means. If not, determine if means from the tests can be pooled. This degree(s) of freedom provides the next narrower confidence interval. If they can be pooled,

$$\sigma_a = 0.$$

- Level Means. If not, determine the confidence interval from the variation among the N" levels.

E.2 Time Parameters

Performance data from multiple tests of a time parameter can be considered to come from the same population if the population parameters can be considered to come from the same population.

The tests should come from populations that have approximately the same (but unknown) variance. Then the null hypothesis states that the tests are from populations with equal means. If the null hypothesis cannot be rejected, both the means and the variances are considered to be equal, and the tests are considered to come from a single population; hence, the trials can be pooled.

The sum of squares of normally distributed random variables have the chi-squared distribution, but since the population variance of this sum is unknown, this statistic cannot be compared with the chi-squared distribution to test the null hypothesis. However, the ratio of two such statistics is independent of the unknown variation. This ratio is the F statistic whose distribution is called the F distribution. The null hypothesis can be tested by comparing this statistic with a specified percentage point of the F distribution. If the statistic is less than the upper 5% point of the F distribution, the null hypothesis is accepted at the 5% level; the population means can be considered to be equal.

A random variable is said to have the F distribution with ν_1 and ν_2 degrees of freedom if its density function is given by

$$f_F(x|\nu_1,\nu_2) = \begin{cases} 0 & x \le 0 \\ \dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} \nu_1^{\nu_1/2}\nu_2^{\nu_2/2} x^{(\nu_1/2)-1}(\nu_1 x + \nu_2)^{-(\nu_1+\nu_2)/2} & x > 0 \end{cases} \qquad \text{(E-1)}$$

where Γ is the Gamma function. This density function is shown in Figure E-1 for three pairs of degrees of freedom.

E.2.1 Pool the Trials

It is assumed that the delays are approximately normally distributed. Further it is assumed that the standard deviations of the delays from the tests are equal and denoted by σ. Under these two assumptions, the delays can be considered to come from the same population if the means from each test are also equal. That is, the delays from N′ tests can be combined if there is no reason to reject the null hypothesis that the means of the N′ populations are equal:

$$H_0: w_1' = \ldots = w_{N'}' .$$

Pooling the trials will be discussed for delays, user fraction of delays, and rates.

A. Delays

Suppose w_{lmn} is the nth trial in the mth test in the lth level of a variable condition. Consider the following two statistics:[3]

$$A = \sum_{l=1}^{N''}\sum_{m=1}^{N_l} N_{lm}(w_{lm.} - \overline{w})^2, \qquad \text{(E-2)}$$

and

$$B = \sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}} (w_{lmn} - w_{lm.})^2 \qquad \text{(E-3)}$$

[3]The following notation, such as $w_{lm.}$, indicates that w_{lmn} has been averaged over the subscript, n.

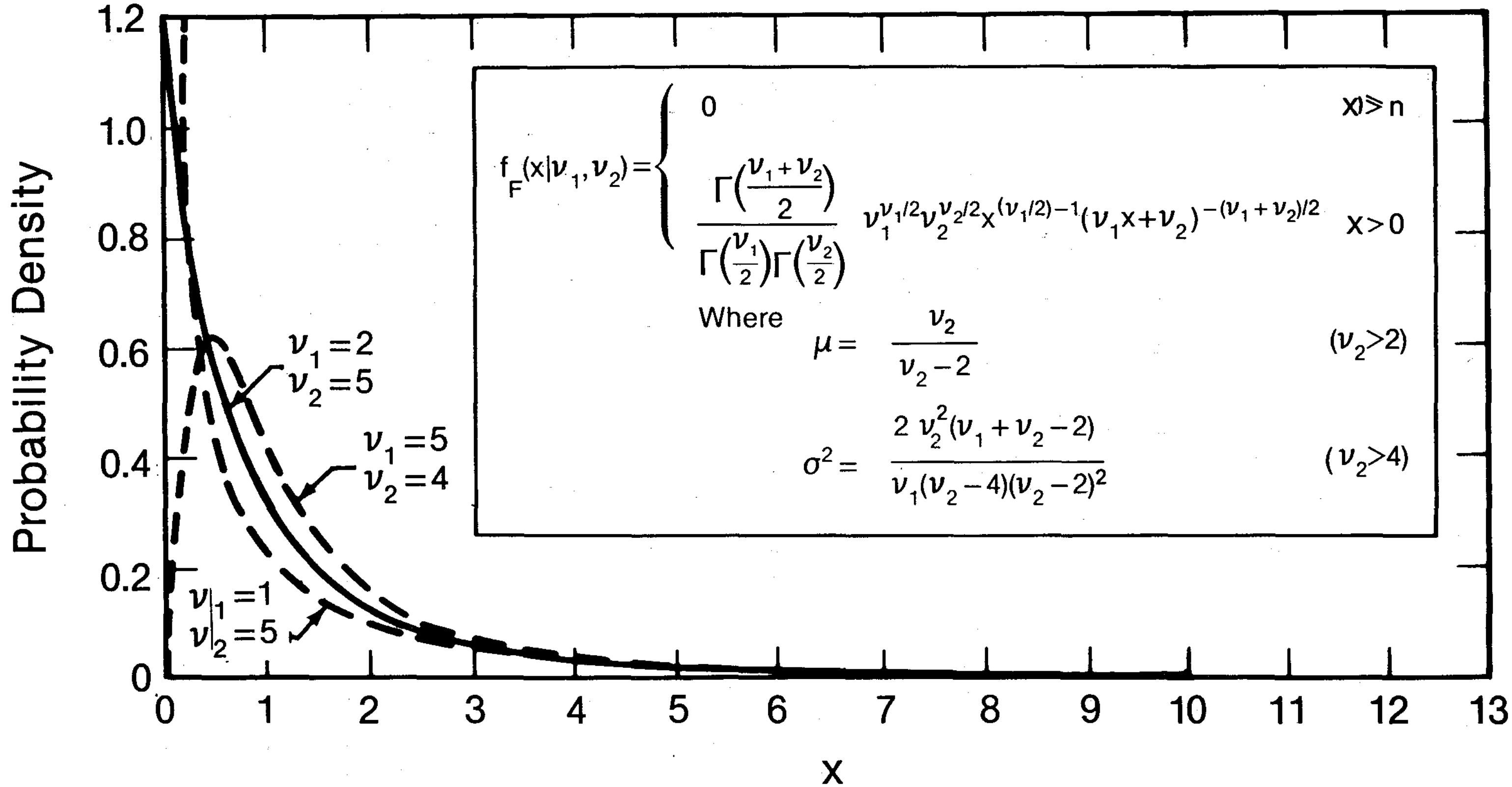

Figure E-1. F density function for three pairs of degrees of freedom.

where

$$N = \sum_{l=1}^{N''} \sum_{m=1}^{N_l} N_{lm} , \tag{E-4}$$

$$\overline{w} = \frac{1}{N} \sum_{l=1}^{N''} \sum_{m=1}^{N_l} \sum_{n=1}^{N_{lm}} w_{lmn} , \tag{E-5}$$

and

$$w_{lm.} = \frac{1}{N_{lm}} \sum_{n=1}^{N_{lm}} w_{lmn} .$$

The statistic, A, is the sum of squares of deviations of the N′ means from the overall mean, $\overline{w}$; it is the sum of squares <u>among</u> tests. The statistic, B, is the sum of squares of deviations of the trial values from the N′ means; it is the sum of squares <u>within</u> tests.

Since these independent statistics are the sum of squares of normally distributed random variables, the statistics, $A/(N' - 1)$ and $B/(N - N')$, have the chi-squared distribution with $N' - 1$ and $N - N'$ degrees of freedom, respectively. Since the variance is unknown, the chi-squared test cannot be used for either statistic to test the null hypothesis. However, the ratio of the two statistics,

$$F = \frac{A/(N' - 1)}{B/(N - N')} , \tag{E-6}$$

is a statistic that is independent of the unknown variance, and its distribution under H_0 is known. Under the assumption of independence among trials, its distribution is the F distribution with $\nu_1 = N' - 1$ and $\nu_2 = N - N'$ degrees of freedom. However, the effective degrees of freedom must be modified by dependence that may exist among trials. This is accomplished by multiplying F by the factor

$$\frac{N/c_N^2(\overline{\rho}_1) - N'}{N - N'} \tag{E-7}$$

where, in this case,

$$\overline{\rho}_1 = \frac{\sum_{l=1}^{N''}\sum_{m=1}^{N_l}(N_{lm}-1)\overline{\rho}_{1lm.}}{N-N'} , \tag{E-8}$$

$$\overline{\rho}_{1lm.} = \frac{\sum_{n=1}^{N_{lm}-1}(w_{lmn}-w_{lm.})\cdot(w_{lm,n+1}-w_{lm.})}{\sum_{n=1}^{N_{lm}}(W_{lmn}-w_{lm.})^2} ,$$

and $c_N^2(\overline{\rho}_1)$ is defined in equation (A-1) (where $\overline{\rho}_1$ replaces r_1).

If the F statistic is less than the 5% point of the F distribution, the trials can be considered to come from the same population. The estimate of the mean delay is $\overline{w}$, and its confidence limits are

$$\left.\begin{matrix} w_U \\ w_L \end{matrix}\right\} = \overline{w} \pm t_{N_0-1,\alpha}\,\overline{\sigma_{\overline{w}}} \tag{E-9}$$

where

$$\overline{w} = \frac{1}{N}\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}} w_{lmn} ,$$

$$\overline{\sigma}_w^2 = \frac{1}{N-1}\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}}(w_{lmn}-\overline{w})^2 , \tag{E-10}$$

and

$$\overline{\rho}_1 = \frac{\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}-1}(w_{lmn}-\overline{w})(w_{lm,n+1}-\overline{w})}{\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}}(w_{lmn}-\overline{w})^2} . \tag{E-11}$$

Then

$$\overline{\sigma}_{\overline{w}}^2 = \frac{\overline{\sigma}_w^2}{N}\cdot c_N^2(\overline{\rho}_1) , \tag{E-12}$$

and

$$N_0 = N/c_N^2(\overline{\rho}_1) . \tag{E-13}$$

First-order Markov dependence affects these confidence limits in two ways. It affects the variance of $\overline{w}$, and it affects the effective degrees of freedom. The confidence interval is increased in both ways if $\overline{\rho}_1 > 0$. Since autocorrelation can exist between trials, it affects these confidence limits (from pooled trials), but it should not affect the confidence limits of the pooled test means or pooled level means (as determined in sections E.2.2 and E.2.3).

B. User Fractions of Delays

The F statistic uses delays as the basis for pooling user fractions of delays (e.g., Access Times as the basis for pooling User Fraction of Access Times). If trials can be pooled, the unbiased estimate of the mean and its confidence limits are obtained as in subroutine **ftest-r**.

Specifically, for the user fractions of delays,

$$v = t/w.$$

Then

$$\overline{t} = \frac{1}{N}\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}} t_{lmn} , \tag{E-14}$$

$$\overline{\sigma}_t^2 = \frac{1}{N-1}\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}} (t_{lmn} - \overline{t})^2 , \tag{E-15}$$

$$\overline{\rho}_1(\overline{t}) = \frac{\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}-1} (t_{lmn} - \overline{t})(t_{lm,n+1} - \overline{t})}{\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}} (t_{lmn} - \overline{t})^2} , \tag{E-16}$$

$$\overline{\sigma}_{\overline{t}}^2 = \frac{\overline{\sigma}_t^2}{N} \cdot c_N^2[\overline{\rho}_1(\overline{t})] , \tag{E-17}$$

$$\overline{\sigma}_{tw} = \frac{1}{N-1}\sum_{l=1}^{N''}\sum_{m=1}^{N_l}\sum_{n=1}^{N_{lm}} (t_{lmn} - \overline{t})(w_{lmn} - \overline{w}) , \tag{E-18}$$

and

$$\overline{\sigma}_{\overline{t}\overline{w}} = \frac{\overline{\sigma}_{tw}}{N} \cdot c_N^2 \{ [\overline{\rho}_1(\overline{t}) \cdot \overline{\rho}_1(\overline{w})]^{1/2} \} \tag{E-19}$$

where $\overline{\rho}_1(\overline{w})$ is defined as in equation E-16. Now,

$$\overline{v} = \frac{\overline{t}}{\overline{w}} \cdot \left(1 + \frac{\overline{\sigma}_{\overline{t}\overline{w}}}{\overline{t}\overline{w}} - \frac{\overline{\sigma}_{\overline{w}}^2}{\overline{w}^2} \right) , \tag{E-20}$$

and

$$\overline{\sigma}_{\overline{v}}^2 = \overline{v}^2 \cdot \left(\frac{\overline{\sigma}_{\overline{t}}^2}{\overline{t}^2} + \frac{\overline{\sigma}_{\overline{w}}^2}{\overline{w}^2} - 2 \frac{\overline{\sigma}_{\overline{t}\overline{w}}}{\overline{t}\overline{w}} \right) . \tag{E-21}$$

Then,

$$\left. \begin{matrix} v_U \\ v_L \end{matrix} \right\} = \overline{v} \pm t_{N_0 - 1, \alpha} \overline{\sigma}_{\overline{v}} \tag{E-22}$$

where

$$N_0 = N / c_N^2 \{ [\overline{\rho}_1(\overline{t}) \cdot \overline{\rho}_1(\overline{w})]^{1/2} \} .$$

C. **Rates**

The F statistic uses delays as the basis for pooling rates (e.g., Input/Output Times as the basis for pooling User Information Bit Transfer Rates). If trials can be pooled, the unbiased estimate of the mean and its confidence limits are obtained as in subroutine **ftest-r**.

The equations for rates are identical to those for user fractions of delays, except that b (Section A.1.3) replaces t, and r (Section A.1.3) replaces v.

E.2.2 Pool the Test Means

If the trials cannot be considered to come from the same population, determine whether the test means come from the same population. If so, this pooling would have the next smaller degrees of freedom, and, therefore, the next larger confidence interval.

Formulate the null hypothesis that the N" level means are equal. That is,

$$H_0: \ w_1'' = w_2'' = \ldots = w_{N''}'' .$$

It is assumed that trials are dependent, but test means are independent.

Pooling the test means will be discussed for delays, user fraction of delays, and rates.

A. Delays

Consider the following two statistics:

$$A = \sum_{l=1}^{N''} N_l \, (w_{l..}' - \overline{w}')^2, \qquad \text{(E-23)}$$

and

$$B = \sum_{l=1}^{N''} \sum_{m=1}^{N_l} (w_{lm.} - w_{l..}')^2 \qquad \text{(E-24)}$$

where

$$N' = \sum_{l=1}^{N''} N_l \, ,$$

$$\overline{w}' = \frac{1}{N'} \sum_{l=1}^{N''} \sum_{m=1}^{N_l} w_{lm.} , \qquad \text{(E-25)}$$

and

$$w_{l..}' = \frac{1}{N_l} \sum_{m=1}^{N_l} w_{lm.} \ . \qquad \text{(E-26)}$$

The statistic, A, is the sum of the squares of deviations of the N" level means from the test mean, $\overline{w}'$; it is the sum of squares among levels. The statistic, B, is the sum of squares of the deviations of the test means from the N" level means, $w'_{l..}$; it is the sum of squares within levels.

The F statistic is

$$F = \frac{A/(N'' - 1)}{B/(N' - N'')} \ . \qquad \text{(E-27)}$$

If the F statistic is less than the 5% point of the F distribution, the test means can be considered to come from the same population. The estimate of the mean delay is $\bar{w}'$, and its confidence limits are

$$\left.\begin{matrix} w_U' \\ w_L' \end{matrix}\right\} = \bar{w}' \pm t_{N'-1,\alpha}\, \bar{\sigma}_{\bar{w}'} \qquad \text{(E-28)}$$

where

$$\bar{w}' = \frac{1}{N'}\sum_{l=1}^{N''}\sum_{m=1}^{N_l} w_{lm.}\,,$$

$$\bar{\sigma}^2_{w'} = \frac{1}{N'-1}\sum_{l=1}^{N''}\sum_{m=1}^{N_l} (w_{lm.} - \bar{w}')^2\,, \qquad \text{(E-29)}$$

and

$$\bar{\sigma}^2_{\bar{w}'} = \frac{\bar{\sigma}^2_{w'}}{N'}\,. \qquad \text{(E-30)}$$

B. **User Fraction of Delays**

The F statistic uses delays as the basis for pooling user fractions of delays (e.g., Access Times as the basis for pooling User Fraction of Access Times.)

In the previous section we assumed that the test mean of

$$v_{lm.} = t_{lm.}/w_{lm.}$$

would be a biased estimator of v. Therefore, equation E-20 was derived to provide an unbiased estimate. In this section it is assumed that the test mean of

$$v_{lm.} = t_{lm.}/w_{lm.}$$

is not biased. Then

$$\overline{v}' = \frac{1}{N'}\sum_{l=1}^{N''}\sum_{m=1}^{N_l} v_{lm.} \ , \tag{E-31}$$

$$\overline{\sigma}^2_{v'} = \frac{1}{N'-1}\sum_{l=1}^{N''}\sum_{m=1}^{N_l} (v_{lm.} - \overline{v}')^2 \ , \tag{E-32}$$

and

$$\overline{\sigma}^2_{\overline{v}'} = \frac{\overline{\sigma}^2_{v'}}{N'} \ . \tag{E-33}$$

Then the estimate of the mean and its confidence limits are

$$\left.\begin{matrix} v'_U \\ v'_L \end{matrix}\right\} = \overline{v}' \pm t_{N'-1,\alpha}\, \overline{\sigma}_{\overline{v}'} \ . \tag{E-34}$$

C. **Rates**

The F statistic uses delays as the basis for pooling rates (e.g., Input/Output Times as the basis for pooling User Information Bit Transfer Rates). If test means can be pooled, the unbiased estimate of the mean and its confidence limits are obtained as in subroutine **ftest-r**.

The equations for rates are identical to those of user fractions, except that $b_{lm.}$ (equation A-13) replaces $t_{lm.}$, and $r_{lm.}$ (equation A-13) replaces $v_{lm.}$.

E.2.3 Pool the Level Means

If the test means cannot be considered to come from the same population, we are forced to use confidence limits based on the pooled tests from each level. In this case there is no hypotheses test.

Pooling the level means will be discussed for delays, user fraction of delays, and rates.

A. Delays

The level means are assumed to be independent (as are the test means). The estimate of the delays is w", and its confidence limits are

$$\left.\begin{matrix} w''_U \\ w''_L \end{matrix}\right\} = \overline{w}'' \pm t_{N''-1,\alpha}\,\overline{\sigma}_{\overline{w}''} \tag{E-35}$$

where

$$\overline{w}'' = \frac{1}{N''}\sum_{i=1}^{N''} w'_{i..} \;, \tag{E-36}$$

$$\overline{\sigma}^2_{w''} = \frac{1}{N''-1}\sum_{i=1}^{N''} (w'_{i..} - \overline{w}'')^2 \;, \tag{E-37}$$

and

$$\overline{\sigma}^2_{\overline{w}''} = \frac{\overline{\sigma}^2_{w''}}{N''} \;. \tag{E-38}$$

B. User Fraction of Delays

For user fractions of delays, the estimate of v" is assumed to be unbiased because v' is assumed to be unbiased. Hence,

$$\left.\begin{matrix} v''_U \\ v''_L \end{matrix}\right\} = \overline{v}'' \pm t_{N''-1,\alpha}\,\overline{\sigma}_{\overline{v}''} \tag{E-39}$$

where

$$v'_{i..} = \frac{1}{N_1}\sum_{m=1}^{N_1} v_{im.} \;, \tag{E-40}$$

$$\overline{v}'' = \frac{1}{N''}\sum_{i=1}^{N''} v'_{i..} \;, \tag{E-41}$$

$$\overline{\sigma}^2_{v''} = \frac{1}{N''-1}\sum_{i=1}^{N''} (v'_{i..} - \overline{v}'')^2 \;, \tag{E-42}$$

and

$$\overline{\sigma}^2_{\overline{v}''} = \frac{\overline{\sigma}^2_{v''}}{N''} \;. \tag{E-43}$$

C. **Rates**

The equations for rates are identical to those for user fraction of delays, with r" replacing v".

E.3 Failure Probability Parameters

The performance data from multiple tests of a failure probability parameter can be considered to come from the same population if the population parameters can be considered to come from the same population. Our Markov model has two population parameters, p and λ, where p is the probability of a failure and λ is the probability of a failure given that a failure occurred in the previous trial.

E.3.1 Pool the Trials

If the first-order Markov chain is the model, both of its parameters, p and λ, must pass the hypothesis test. Both parameters are proportions that are binomially distributed. It is not necessary to follow the notation of the mathematical model for x_{lmn} in this subsection; assume there are k tests.

A. **Hypothesis Test Applied to p**

Suppose random samples from i = 1, 2, ..., k independent binomial distributions with population means, $p_1, \ldots, p_k$, yield estimates, $\overline{p}_1, \ldots, \overline{p}_k$, from test sizes, $n_1, \ldots, n_k$. The standard deviation of each estimate is

$$\sigma_{\overline{p}_i} = \sqrt{\frac{p_i q_i}{n_i}} \cdot c_{n_i}(\overline{\rho}_{1i}) \ .$$

If the samples are sufficiently large, the distribution of each standardized random variable,

$$X_i = \frac{p_i - \overline{p}_i}{\sigma_{p_i}},$$

can be approximated by the standard normal distribution. Since

$$\sum_{i=1}^{k} X_i^2$$

is known to have the chi-squared distribution with k - 1 degrees of freedom,

$$\chi^2_{k-1}(p) = \sum_{i=1}^{k}\left(\frac{p_i - \bar{p}_i}{\sigma_{\bar{p}_i}}\right)^2 .$$

The chi-squared density function is

$$f_c(x|k) = \begin{cases} 0 & x \le 0 \\ \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} \cdot x^{\frac{k}{2}-1} e^{-\frac{k}{2}} & x > 0 . \end{cases} \qquad \text{(E-44)}$$

where Γ is the Gamma function. Figure E-2 shows the chi-squared density function for k - 1 degrees of freedom for k = 1, 4, 10, and 20. Figure E-3 shows the acceptance and rejection intervals for the distribution with k = 4 degrees of freedom and the $\alpha = 0.05$ significance level.

Formulate the null hypothesis that the k population proportions are equal. That is,

$$H_0: p_1 = \ldots = p_k .$$

Since p is the only unknown parameter of the binomial distribution, accepting the null hypothesis is equivalent to accepting the hypothesis that the samples are from the same population.

By virtue of the null hypothesis, estimate the common value of p_i, say P, by the proportion of pooled outcomes,[4]

$$\bar{P} = \frac{S}{N} ,$$

where

$$S = \sum_{i=1}^{k} s_i , \text{ and } N = \sum_{i=1}^{k} n_i .$$

[4]In this section, upper case letters usually represent pooled outcomes, not random variables.

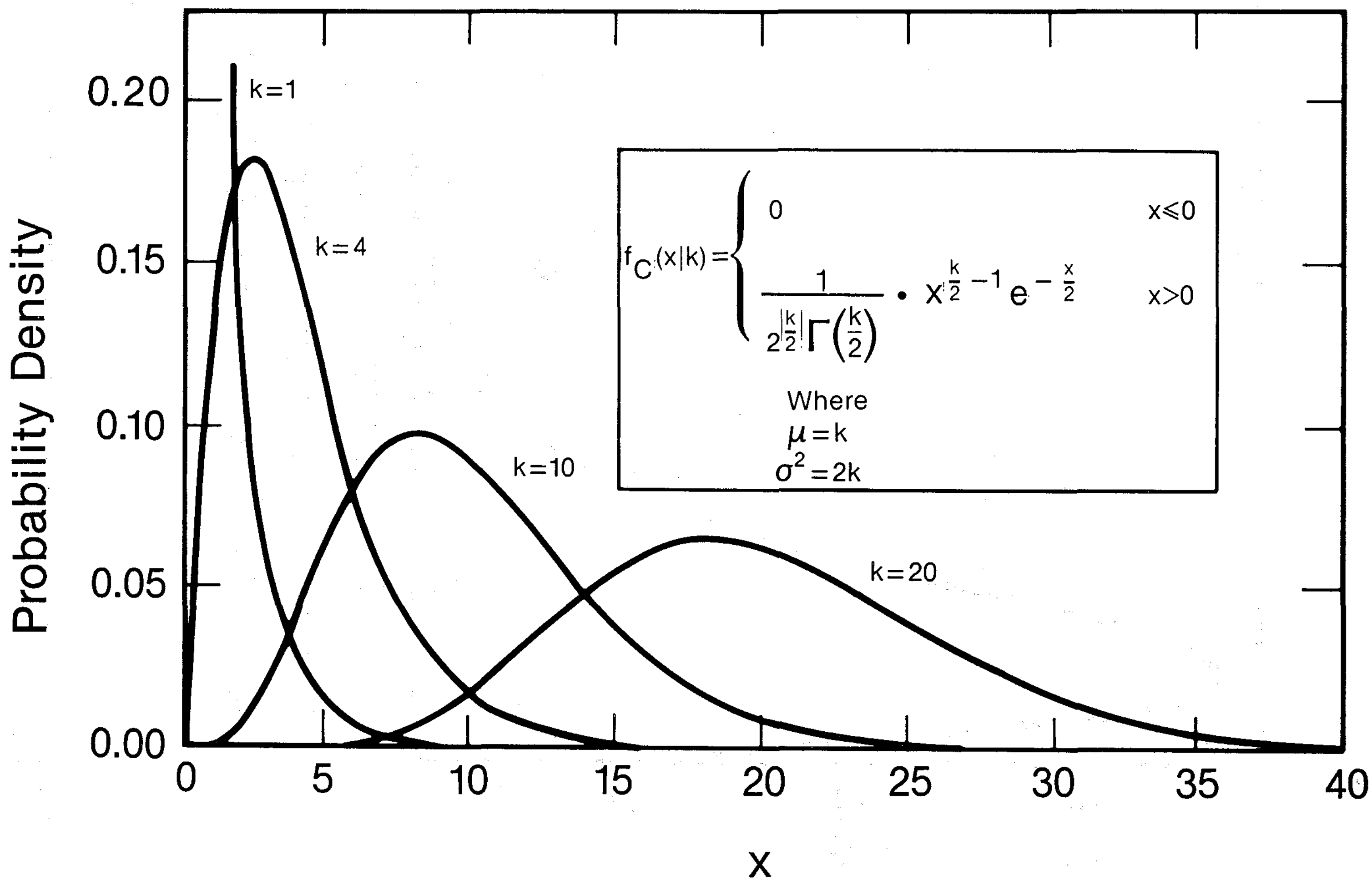

Figure E-2. Chi-squared density function for four degrees of freedom.

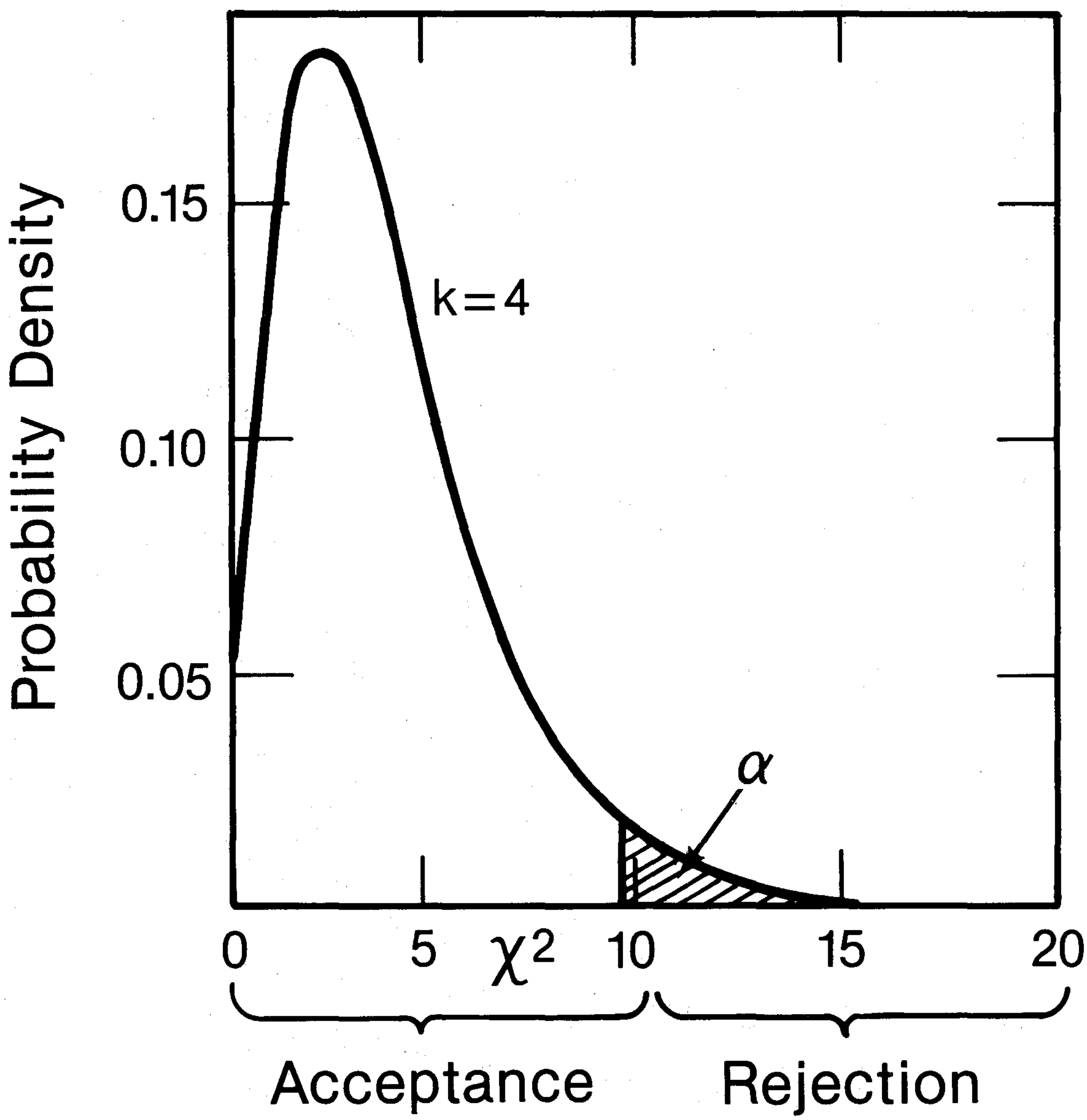

Figure E-3. Acceptance and rejection intervals for a chi-squared density function.

Also replace $\sigma_{\bar{P}_i}$ with its estimator, $\bar{\sigma}_{\bar{P}_i}$ (determined in E-46). Now,

$$\chi^2_{k-1}(p) = \sum_{i=1}^{k} \left(\frac{p_i - \bar{P}}{\bar{\sigma}_{\bar{P}_i}} \right)^2 \tag{E-45}$$

has the chi-squared distribution with k - 1 degrees of freedom. The degrees of freedom have been reduced by one due to the single constraint,

$$\bar{P} = \frac{S}{N} .$$

The standard deviation of each p_i under H_0 is estimated by

$$\bar{\sigma}_{\bar{P}_i} = \sqrt{\frac{\bar{P}\bar{Q}}{n_i}} \cdot c_{n_i}(\bar{P}_1) \tag{E-46}$$

where $\bar{Q} = 1 - \bar{P}$, and $\bar{P}_1$ (in this case, P is upper case ρ) is determined below.

Estimate each λ_i by the relative frequency of pooled outcomes,[5]

$$\Lambda^* = \frac{\left(\frac{1}{k}\sum_{i=1}^{k} r_i\right) \Big/ \left(\frac{1}{k}\sum_{i=1}^{k} (n_i - 1)\right)}{\left(\frac{1}{k}\sum_{i=1}^{k} s_i\right) \Big/ \left(\frac{1}{k}\sum_{i=1}^{k} n_i\right)}$$

$$= \frac{\frac{1}{k}\sum_{i=1}^{k} n_i}{\frac{1}{k}\sum_{i=1}^{k} (n_i - 1)} \cdot \frac{\frac{1}{k}\sum_{i=1}^{k} r_i}{\frac{1}{k}\sum_{i=1}^{k} s_i} = \frac{N}{N - k} \cdot \frac{R}{S} \tag{E-47}$$

where

$$R = \sum_{i=1}^{k} r_i . \tag{E-48}$$

[5]Even though both p_i and λ_i are estimated, λ_i is not a parameter of the chi-squared distribution; hence, only one degree of freedom is lost.

Now,

$$P_1 = \frac{\Lambda^* - \overline{P}}{\overline{Q}} \ . \tag{E-49}$$

B. **Hypothesis Test Applied to λ**

Similarly, suppose that random variables from k binomial distributions with population proportions, λ_1, λ_2, ..., λ_k, have yielded the estimates λ_1^*, λ_2^*, ..., λ_k^*, from tests of size s_1, s_2, ..., s_k, respectively. These estimates are defined in equation A-22 where they are labelled $\overline{\lambda}$. The probability that s_i failures result in r_i pairs of consecutive failures is given by[6]

$$f_B(r_i|\lambda_i, s_i) = \binom{s_i}{r_i}\lambda_i^{r_i}(1 - \lambda_i^{s_i-r_i}) \ . \tag{E-50}$$

Approximately the mean of each estimate, λ_i^*, is λ_i, and the standard deviation is

$$\overline{\sigma}_{\lambda_i} = \frac{\lambda_i(1 - \lambda_i)}{s_i} \cdot c_{s_i}(0) \ . \tag{E-51}$$

Then, similar to the formula for p,

$$\chi^2_{k-1}(\lambda) = \sum_{i=1}^{k}\left(\frac{\lambda_i^* - \lambda_i}{\overline{\sigma}_{\lambda_i}}\right)^2 \tag{E-52}$$

has approximately the chi-squared distribution with k-1 degrees of freedom.

Formulate the null hypothesis

$$H_0 : \lambda_1 = \ldots = \lambda_k \ .$$

Then, by virtue of the null hypothesis, replace each λ_i with Λ^*, the estimator of λ. Also replace each $\sigma\lambda_i$ with its estimator,

[6]Because the model is only first order Markov, the pairs of failures (not trials) are asymptotically independent and their autocorrelation is zero.

$$\overline{\sigma}_{\lambda_i^*} = \sqrt{\frac{\Lambda^*(1 - \Lambda^*)}{s_i}} \ . \qquad \text{(E-53)}$$

Now,

$$\chi^2_{k-1}(\lambda) = \sum_{i=1}^{k}\left(\frac{\lambda_i^* - \Lambda^*}{\overline{\sigma}_{\lambda_i^*}}\right)^2 \qquad \text{(E-54)}$$

has the chi-squared distribution with k - 1 degrees of freedom.

However, since each sample size, s_i, is probably quite small (especially compared to each n_i), it may be more appropriate to compare the sampling distribution to the population distribution by testing the hypothesis that the test variance is compatible with the population variance. This test can be obtained as a slight modification of the chi-squared test for contingency tables. Hence, test the statistic (the modified binomial index of dispersion),

$$\chi^2_{k-1}(\lambda) = \sum_{i=1}^{k} \frac{s_i(\lambda_i^* - \Lambda^*)^2}{\Lambda^*(1 - \Lambda^*)} \qquad \text{(E-55)}$$

which has the chi-squared distribution with k-1 degrees of freedom. If R = 0, then $\Lambda^* = 0$. In this case, the chi-squared statistic for λ cannot be computed. But, perhaps it needn't be since the trials may be statistically independent and may be modeled by the binomial distribution.

C. Summary

To test the hypotheses, the chi-squared statistics for both p and λ are compared with the chi-squared distribution. If neither χ^2_{k-1} (p) nor χ^2_{k-1} (λ) exceeds the 5% point of the distribution with k-1 degrees of freedom, there is no reason to reject the two hypotheses.

If either χ^2_{k-1} (p) or χ^2_{k-1} (λ) exceeds the 5% point of the chi-squared distribution for k-1 degrees of freedom, there is a question that the data from the k tests should be combined. In this case, the chi-squared test can be applied again after omitting data from one or more tests that are thought to cause rejection of the hypothesis. (Of course, at least two samples must remain.)

Estimate the failure probability and its confidence limits for the pooled trials exactly as in Appendix A, substituting

$$\begin{aligned} &\overline{P} \quad \text{for} \quad \overline{p}\ , \\ &\Lambda^* \quad \text{for} \quad \lambda\ , \\ &\overline{P}_1 \quad \text{for} \quad \overline{\rho}_1\ , \\ &R \quad \text{for} \quad r\ , \\ &S \quad \text{for} \quad s\ , \text{ and} \\ &N \quad \text{for} \quad n\ . \end{aligned} \qquad \text{(E-56)}$$

E.3.2 Pool the Test Proportions

If the trials cannot be pooled, determine if the test proportions can be pooled. Formulate the null hypothesis that the N" level proportions are equal. That is,

$$H_0 \ : \ p_1 = p_2 = \ldots = p_{N''} \ .$$

It is assumed that test proportions are independent, hence the equivalent null hypothesis for λ needn't be posed. The number of failures and trials in the mth test of the lth level is s_{lm} and N_{lm}. Since the proportions p_{lm} cannot be assumed to be normally distributed (even approximately) the following transformation should be used:[7]

$$w = \frac{1}{2}\left[\sin^{-1}\left(\frac{s_{lm}}{N_{lm}+1}\right)^{\frac{1}{2}} + \sin^{-1}\left(\frac{s_{lm}+1}{N_{lm}+1}\right)^{\frac{1}{2}}\right] \cdot \frac{180°}{\pi} \qquad \text{(E-57)}$$

(Bishop et. al., 1975, p. 367) where $l = 1, \ldots, N''$, and $m = 1, \ldots, N_l$. If p denotes the population proportion and if the probability of failure in different samples is independent, then $w_{lm.}$ has an asymptotically normal distribution with mean $\sin^{-1}\sqrt{p}$ and variance $(4N_{lm})^{-1}$.

[7]Some calculators and computers multiply the inverse sine by the factor 180°/π to show the value in degrees, hence this factor may be unnecessary in the following equation.

Consider the following two statistics:

$$A = \sum_{l=1}^{N''} N_l (w'_{l..} - \overline{w}')^2, \tag{E-58}$$

and

$$B = \sum_{l=1}^{N''} \sum_{m=1}^{N_l} (w_{lm.} - w'_{l..})^2 \tag{E-59}$$

where

$$N' = \sum_{l=1}^{N''} N_l , \tag{E-60}$$

$$\overline{w}' = \frac{1}{N'} \sum_{l=1}^{N''} \sum_{m=1}^{N_l} w_{lm.}, \tag{E-61}$$

and

$$w'_{l..} = \frac{1}{N_l} \sum_{m=1}^{N_l} w_{lm.} . \tag{E-62}$$

The statistic A is the sum of the squares of deviations of the N" level means from the test mean, $\overline{w}'$; it is the sum of squares among levels. The statistic B is the sum of squares of the deviations of the test means from the N" means, $w'_{l..}$; it is the sum of squares within levels. The F statistic is

$$F = \frac{A/(N'' - 1)}{B/(N' - N'')} . \tag{E-63}$$

The estimator of the mean is $\overline{w}'$, and its confidence limits are

$$\left.\begin{matrix} w'_U \\ w'_L \end{matrix}\right\} = \overline{w}' \pm t_{N'-1,\alpha}\, \overline{\sigma}_{\overline{w}'} \tag{E-64}$$

where

$$\overline{w}' = \frac{1}{N'} \sum_{l=1}^{N''} \sum_{m=1}^{N_l} w_{lm.}, \tag{E-65}$$

$$\overline{\sigma}^2_{w'} = \frac{1}{N'-1}\sum_{l=1}^{N''}\sum_{m=1}^{N_l} (w_{lm.} - \overline{w}')^2 \ , \qquad \text{(E-66)}$$

and

$$\overline{\sigma}^2_{\overline{w}'} = \frac{\overline{\sigma}^2_{w'}}{N'} \ , \qquad \text{(E-67)}$$

Finally, the estimate of the mean and its confidence limits must be retransformed. The estimate of the failure probability is

$$\overline{p}' = \sin^2\left(\overline{w}' \cdot \frac{\pi}{180°}\right) \ , \qquad \text{(E-68)}$$

and its confidence limits are

$$\left.\begin{matrix} p'_U \\ p'_L \end{matrix}\right\} = \sin^2\left[(\overline{w}' \pm t_{N'-1,\alpha}\, \overline{\sigma}_{\overline{w}'}) \cdot \frac{\pi}{180°}\right] . \qquad \text{(E-69)}$$

E.3.3 Pool the Level Proportions

If the test proportions cannot be considered to come from the same population, we are forced to use confidence limits based on the pooled samples from each level. At this point there is no hypothesis test. The estimator of the means is w", and its confidence limits are

$$\left.\begin{matrix} w''_U \\ w''_L \end{matrix}\right\} = \overline{w}'' \pm t_{N''-1,\alpha}\, \overline{\sigma}_{\overline{w}''} \qquad \text{(E-70)}$$

where

$$\overline{w}'' = \frac{1}{N''}\sum_{l=1}^{N''} \overline{w}'_{l..} \ , \qquad \text{(E-71)}$$

$$\overline{\sigma}^2_{w''} = \frac{1}{N''-1}\sum_{l=1}^{N''} (w'_{l..} - \overline{w}'')^2 \ , \qquad \text{(E-72)}$$

and

$$\overline{\sigma}^2_{\overline{w}''} = \frac{\overline{\sigma}^2_{w''}}{N''} . \qquad \text{(E-73)}$$

Finally, the estimate of the mean and its confidence limits must be retransformed. The estimate of the failure probability is

$$\overline{p}'' = \sin^2\left(\overline{w}'' \cdot \frac{\pi}{180°}\right) , \qquad \text{(E-74)}$$

and its confidence limits are

$$\left.\begin{matrix} p_U'' \\ p_L'' \end{matrix}\right\} = \sin^2\left[(\overline{w}'' \pm t_{N''-1,\alpha}\, \overline{\sigma}_{\overline{w}''}) \cdot \frac{\pi}{180°}\right] . \qquad \text{(E-75)}$$

E.4 References

Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland (1975), *Discrete Multivariate Analysis: Theory and Practice* (The MIT Press, Cambridge, MA).

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, 294pp. (NTIS Order number PB 89-117519/AS)

APPENDIX F: FLOWCHARTS FOR ANALYSIS OF MULTIPLE TESTS

This appendix is a set of flowcharts for each subroutine from **star** that analyzes multiple tests. Figure F-1 is a flow chart that shows the relationship of the subroutines of **star**. The flowcharts of each subroutine are listed alphabetically. In these figures, diamonds indicate decisions, rectangles indicate arithmetic operations, and parallelograms indicate input (output is omitted).

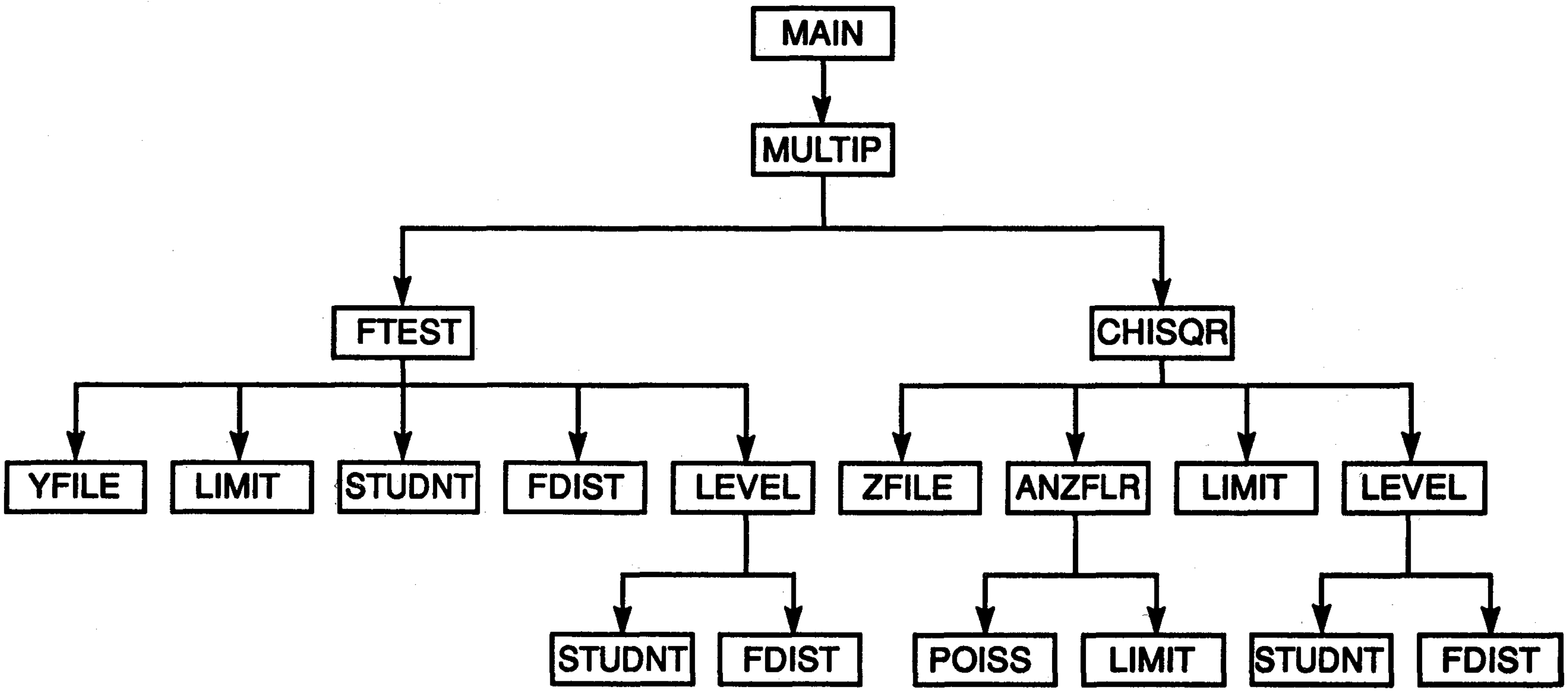

Figure F-1. Flowchart of relationship among subroutines of **star**.

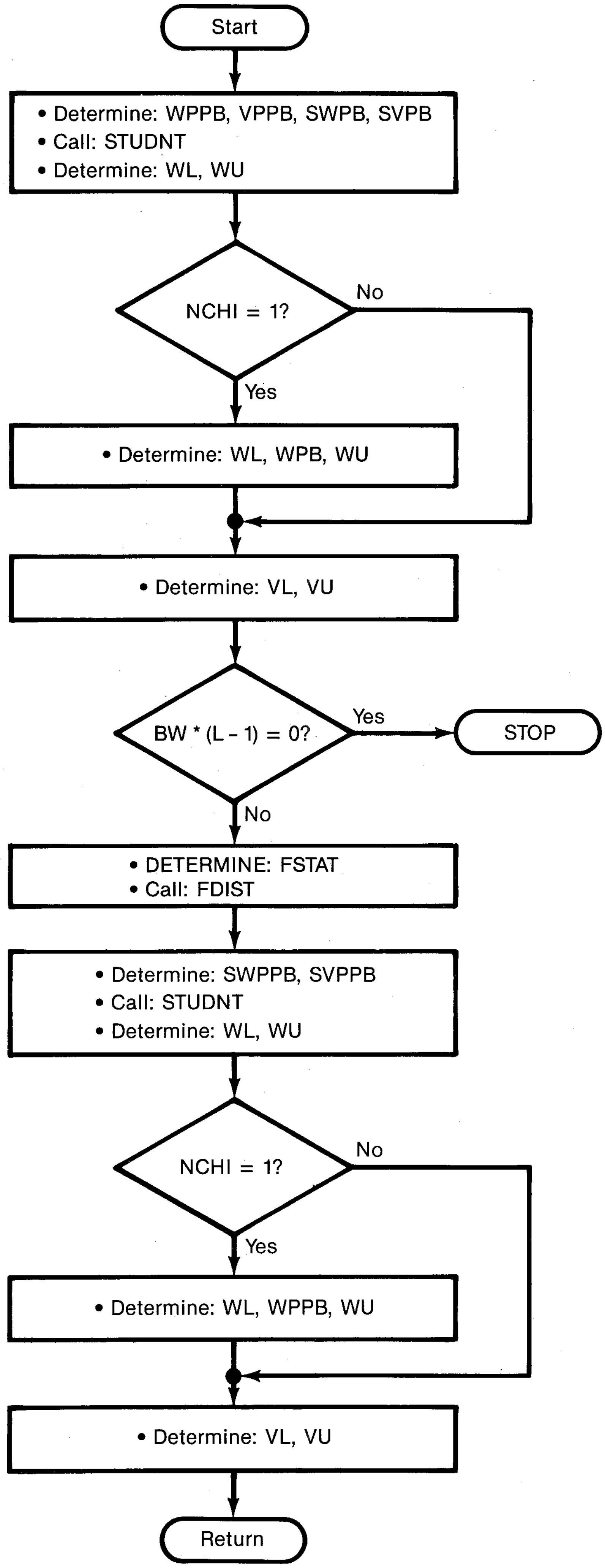

Figure F-5. Flowchart of subroutine level.

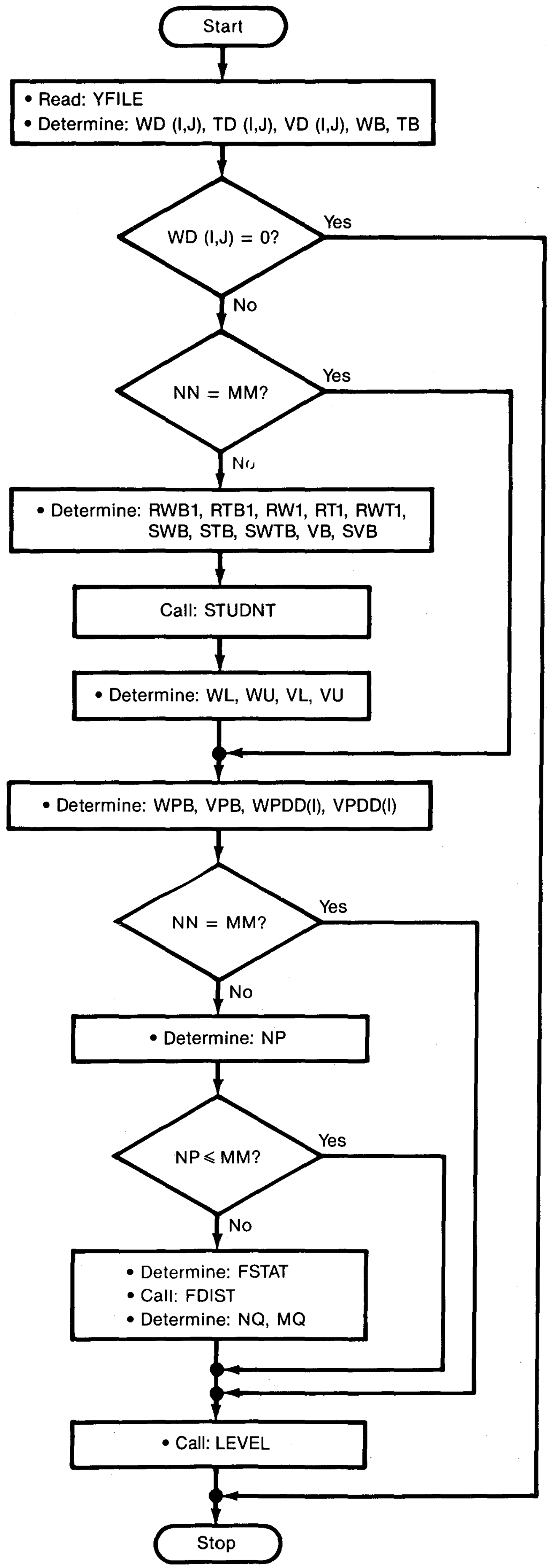

Figure F-4. Flowchart of subroutine ftest.

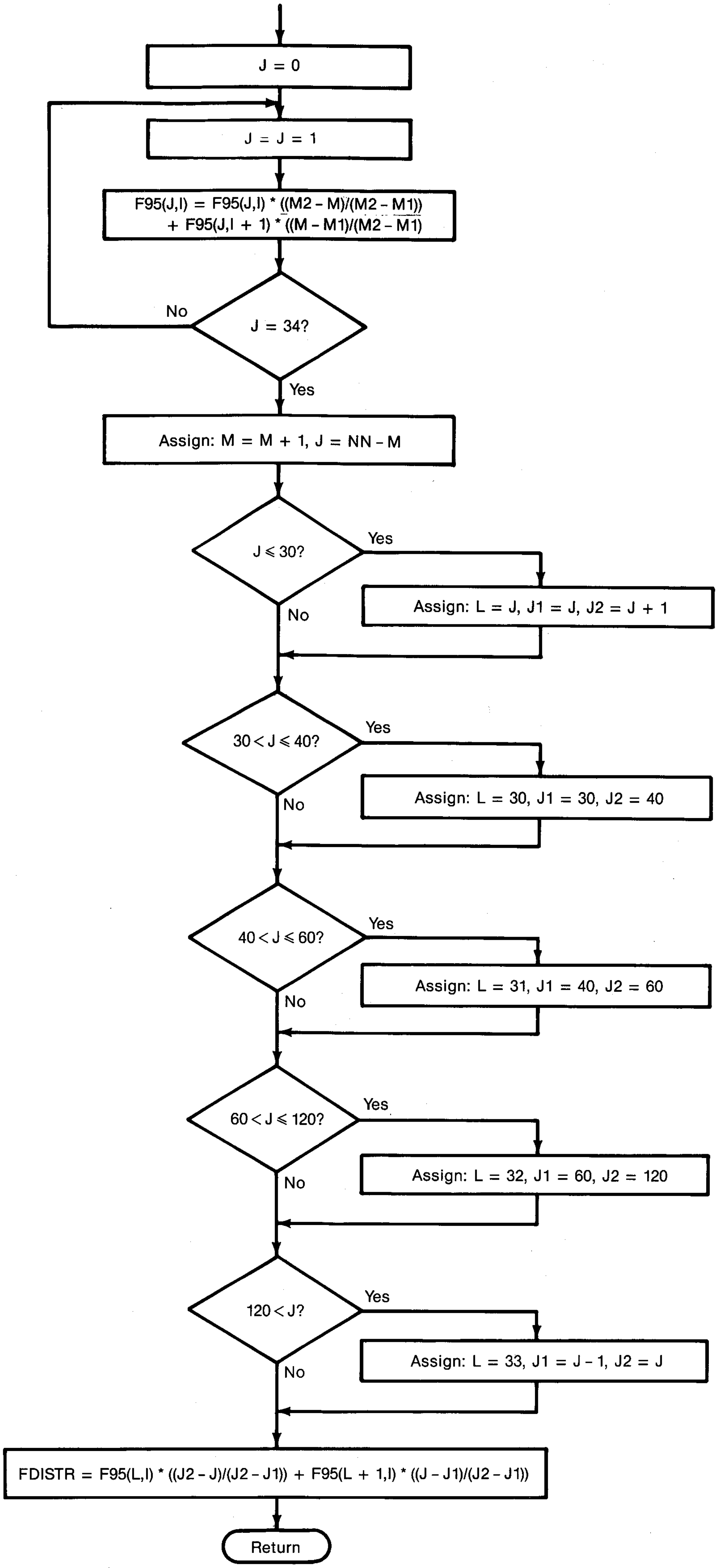

Figure F-3. (Part 3) Flowchart of subroutine fdist.

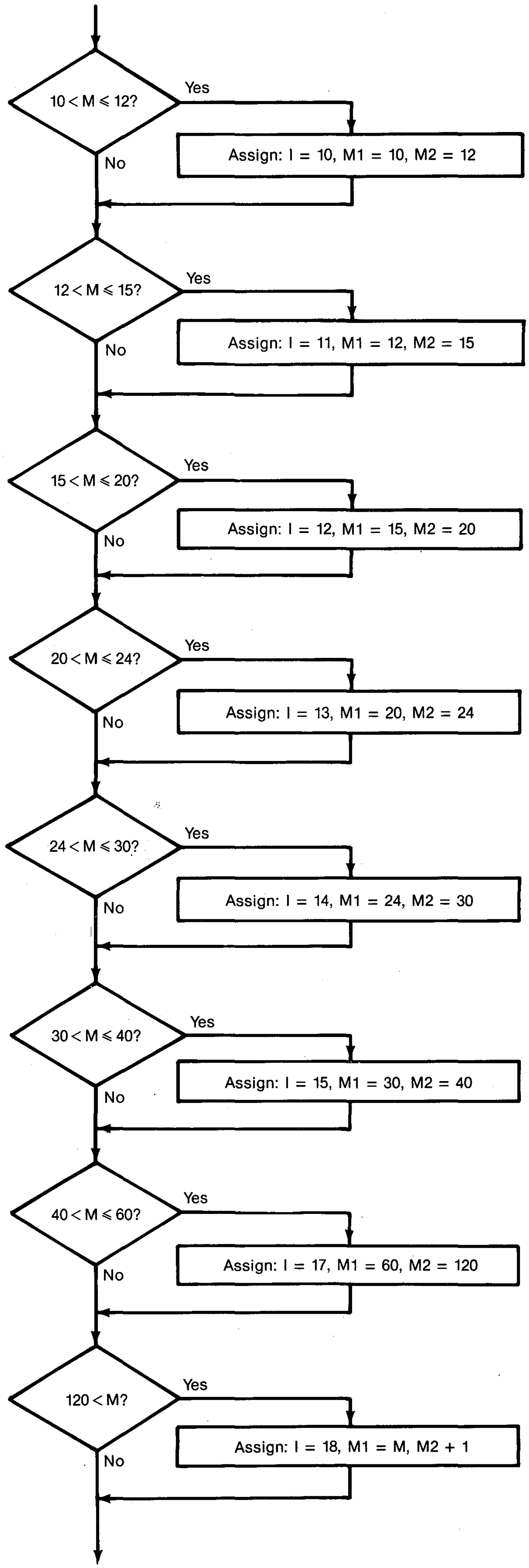

Figure F-3. (Part 2) Flowchart of subroutine fdist.

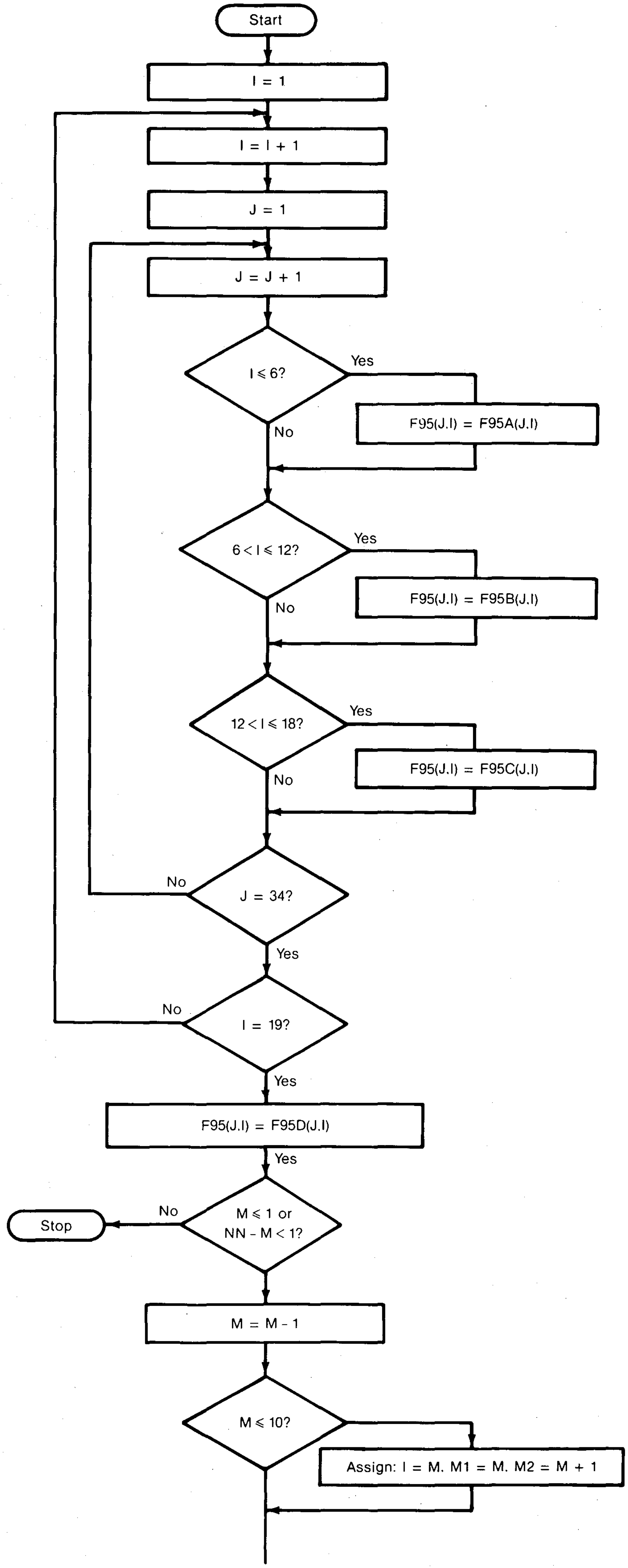

Figure F-3. (Part 1) Flowchart of subroutine fdist.

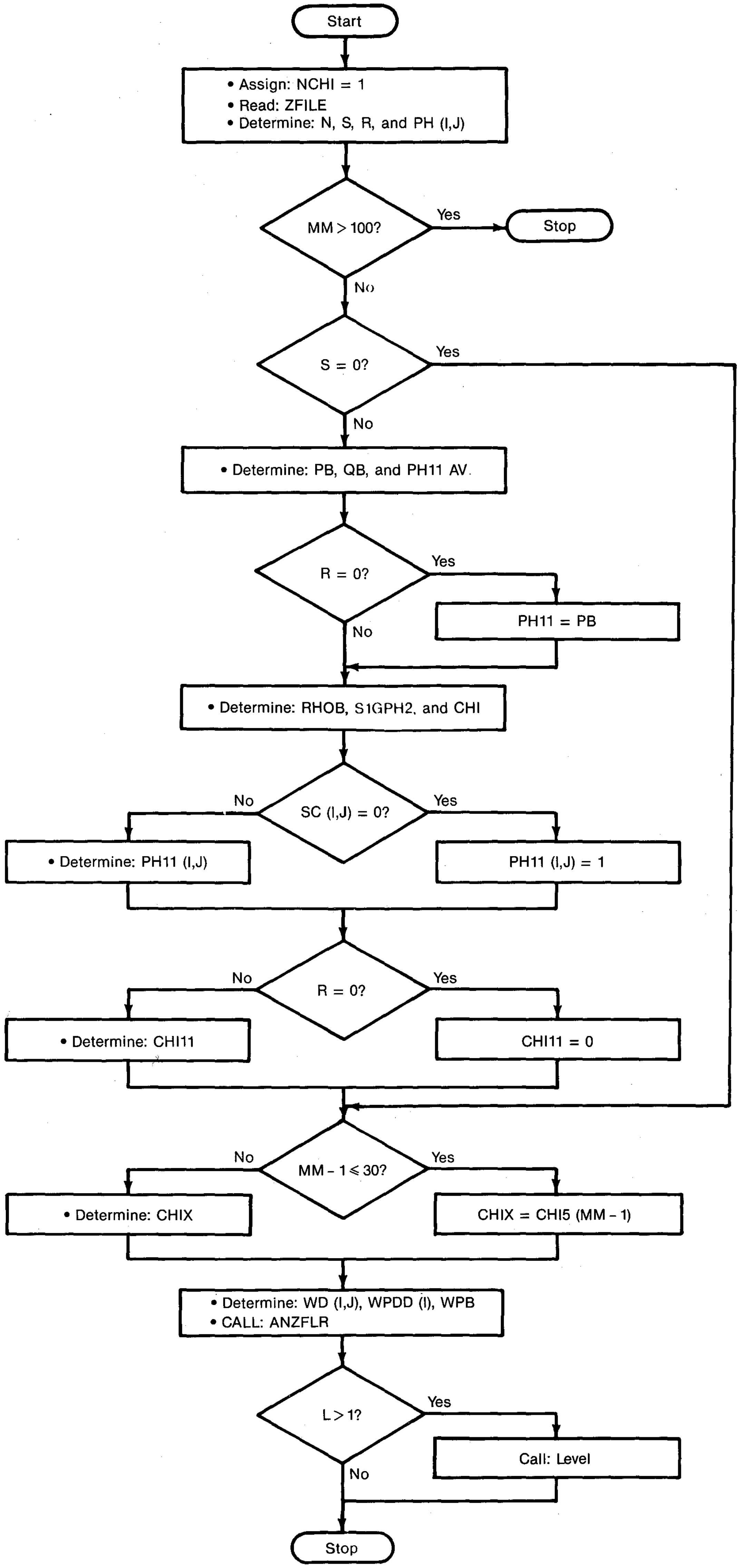

Figure F-2. Flowchart of subroutine chisqr.

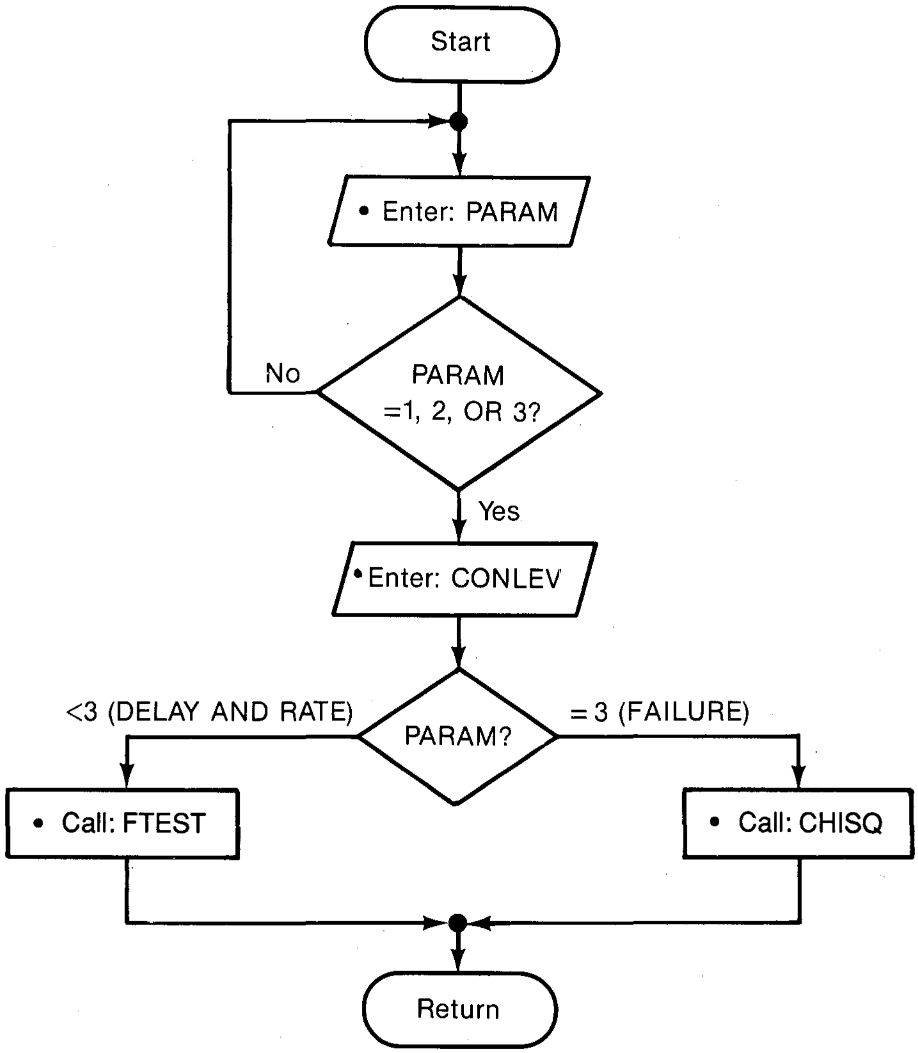

Figure F-6. Flowchart of subroutine **multip**.

APPENDIX G: SHELL SCRIPT IMPLEMENTATION OF ANALYSIS OF MULTIPLE TESTS

Analyses of multiple tests are implemented by the **delay**, **rate**, and **fail** shell scripts. Each of these shell scripts

- generates a file (**star.input**) containing the keyboard responses to prompts issued by **star**,
- estimates the sample means and standard deviations of individual tests,
- calls **star** to estimate the performance parameter and its 95% confidence limits, and
- generates a summary of the analysis and writes it to a file.

Input to a shell script consists of

- a file (**log.wrk**) that contains a line of identification of each test to be analyzed,
- a pair of arguments that specify the performance parameter and a variable condition (selected to test the null hypothesis that level means are equal), and
- one or more files that contain the relevant performance data.

The **log.wrk** file is created prior to executing the shell script. The lines constituting **log.wrk** are extracted from one of two files of test identifications, called **log.acc** and **log.xfr** (depending upon whether the parameters come from access-disengagement tests or user information transfer tests, respectively), by the UNIX^tm **grep** utility. For example, the lines from Thursday (a level of the variable condition, Day of the Week) in file **log.acc** can be stored in file **log.wrk** by typing

grep thu log.acc > log.wrk

G.1 Delay Parameters

The performance data files for delays are listed in Table G-1 where **nnnn** is the test number and the suffix specifies the communication function containing the performance parameters.[8] For each trial in the test, these files list the performance time and the user portion of performance time.

---

[8]These files are created by the shell scripts **time-a** and **time-x** when the tests are conducted. See Appendix C.

Table G-1. Delay Parameters and the Name of the Files That Contain Their Performance Data

DELAY PARAMETERS	FILE
Access Time & User Fraction of Access Time	nnnnac
Block Transfer Time & User Fraction of Block Transfer Time	nnnnb2
Source Destination Time & User Fraction of Source Disengagement Time	nnnnd1
Destination Disengagement Time & User Fraction of Destination Disengagement Time	nnnnd2

The UNIX^{tm} **sort** utility is invoked to group the tests in **log.wrk** according to levels of the specified variable condition; sorted lines are written to a temporary file (**log.tl**). Various UNIX^{tm} utilities are then used to generate a file (**star.input**) that contains the keyboard responses to prompts issued by **star**. Successive lines in **star.input** contain

- the code (40) that specifies multiple test analysis,
- the code (1) that specifies a delay parameter,
- the code (2) that specifies the 95% confidence level, and
- the number of levels of the specified variable condition (determined by **delay**).

These lines are followed by groups of lines corresponding to the different variable condition levels. The successive lines for a particular level contain

- the number of tests at that level, and
- the names of the files that contain the performance data for each test at a level (one name per line).

An example of a **star.input** file for a Block Transfer Time analysis is shown in Figure G-1. Key procedures in generating such a file include counting the number of levels of the variable condition, counting the number of tests at each level, and constructing the names of the files that contain relevant delay data. Several temporary files (not shown) are utilized in these procedures.

```
40
1
2
3
1
775s1
6
823s1
815s1
835s1
858s1
876s1
811s1
4
997s1
928s1
952s1
978s1
```

Figure G-1. Example of file star.input.

delay calls another shell script (**mean-dev**) that estimates, for each test, the mean and standard deviation of the performance time and the user fraction of performance time. Input to **mean-dev** consists of a list of the relevant "stan" files (in **stan.list**) and the performance times in those delay data files. Output from **mean-dev** is written to a temporary file (**s2.out**). **delay** next calls **star** to conduct the delay analysis. Input/output redirection is used so that keyboard responses to **star** prompts are supplied by the **star.input** file, and output is written to a temporary file (**s2.tmp**). **delay** concludes by editing the temporary files output by **mean-dev** and **star** to produce a summary of results which is written to the file **star.out**.

G.2 Rate Parameters

Rate performance data for a given test are contained in the file **thrput**.[9] For each trial (i.e., Transfer Sample), this file lists the Input/Output Time, the user portion of performance time, and the number of bits successfully transferred. The file (**star.input**) that contains the keyboard responses to prompts issued by **star** is identical to that used by **delay** except the parameter code is 2 instead of 1.

rate calls a subordinate shell script (**meanrdev**) that estimates, for each test, the Input/Output Time, User Information Bit Transfer Rate, and User Fraction of Input/Output Time.

G.3 Fail Parameters

The performance data files for failure probabilities are listed in Table G-2 where the suffix specifies the communication function containing the performance parameters.[10] Each line of a file for a particular communication function contains (in addition to the test number) the number of trials in the test, the number of failures, and the number of pairs of consecutive failures for each failure outcome associated with the function. Note that the first two characters of the parameter argument in the **fail** command are the same as the

---

[9]This file is created by the shell script **time-x** when the test is conducted. See Appendix C.

[10]These files are created by the shell scripts **fail-a** and **fail-x** when the test is conducted. See Appendix C.

Table G-2. Failure Probability Parameters and the Name of the Files That Contain Their Performance Data

FAILURE PROBABILITY PARAMETERS	FILE
Access Failures Access Denial Access Outage Incorrect Access	**failac**
Bit Transfer Failures Incorrect Bit Extra Bit Lost Bit	**failb1**
Block Transfer Failures Incorrect Block Extra Block Lost Block	**failb2**
Transfer Denial	**failb3**
Source Disengagement Denial	**faild1**
Destination Disengagement Denial	**faild2**

first two characters in the name of the corresponding failure performance data file. This feature is utilized by the **fail** shell script to select the appropriate summary file.

fail calls the UNIX^{tm} **sort** utility to group the tests in **log.wrk** according to levels of the specified variable condition, and sorted lines are written to the temporary file **log.tl**. **fail** then calls a subordinate shell script (**relate**) to extract, for each test listed in **log.tl**, the corresponding failure summary record from the appropriate file. The extracted lines are written to the file **fail.list**. The number of trials, the number of failures, and the number of pairs of consecutive failures corresponding to the specified parameter are obtained from these lines by another subordinate shell script (**faildev**) and the results are written to the file **failin**. **faildev** also estimates, for each of the specified tests, the failure probability and the standard deviation. These results are recorded in the temporary file **fail.stats**.

Several UNIX^{tm} utilities utilize data in the **log.tl** and **failin** files to generate a file (**star.input**) that contains the keyboard responses to prompts issued by **star**. Successive lines in **star.input** contain

- **the code (40) that specifies multiple test analysis,**
- the code (3) that specifies a failure probability parameter,
- the code (2) that specifies the 95% confidence level,
- the number of levels of the specified variable condition (determined by **fail**), and
- the estimate of the conditional probability of a failure, given that a failure occurred in the previous trial (**0.8**). This estimate is required only if the total number of failures is 0 or 1. This probability directly affects the autocorrelation of lag 1, which directly affects the upper confidence limit. This rather conservative value can be altered at the discretion of the experimenter.

These lines are followed by groups of lines corresponding to the different variable condition levels. The successive lines for a particular level contain

- the number of tests at that level,
- the relevant numbers of trials, failures, and pairs of consecutive failures in each test at the particular level (one test per line), and
- the value -30 in place of performance statistics (to inform the input routine that all data for the level have been entered).

Procedures for counting the number of levels of variable condition and the number of tests at each level are the same as those used in **delay**.

fail next calls **star** to analyze the failures. I/O redirection is used so that keyboard responses to **star** prompts are supplied by the **star.input** file, and output is written to a temporary file (**s2.tmp**). **fail** concludes by editing the temporary files output by **faildev** and **star** to produce a summary of results which is written to the file **fail.out**.

APPENDIX H: OPERATOR IMPLEMENTATION OF ANALYSIS OF MULTIPLE TESTS

Multiple tests can be analyzed to accomplish any of the four recommended statistical analyses: estimation, tests of acceptance, tests of comparison, and tests to determine if a variable condition is a factor.

Although NTIA software provides convenient analysis of multiple tests through shell scripts that use UNIX[tm] utilities and prepared files of performance data, it is important to provide operator implementation because the operator can directly use files of any performance data.

Figure H-1 shows the decisions required to analyze multiple tests of delays, rate, and failure probability.

From **/usr/data/5a**, execute **star** by typing

star

- Enter the code number 40.
- Select the performance parameter type.
- The 95% confidence level is the only confidence level available for multiple tests, so enter 2.

The following two sections show how multiple tests of time and failure probability parameters, respectively, are analyzed.

H.1 Time Parameters

The procedures to analyze multiple tests of delays and rates are similar. Performance data for time parameters can be entered by file only. The first field is the performance time, the second field is the user portion of performance time, and the third field is the number of elements transferred (usually bits). This third field needn't be completed for analysis of delays,

Prepare one file for each test.

but must be completed for analysis of rates. The format is 2F8.3,F8.0. The last line contains -30 in each field as an end of file indicator.

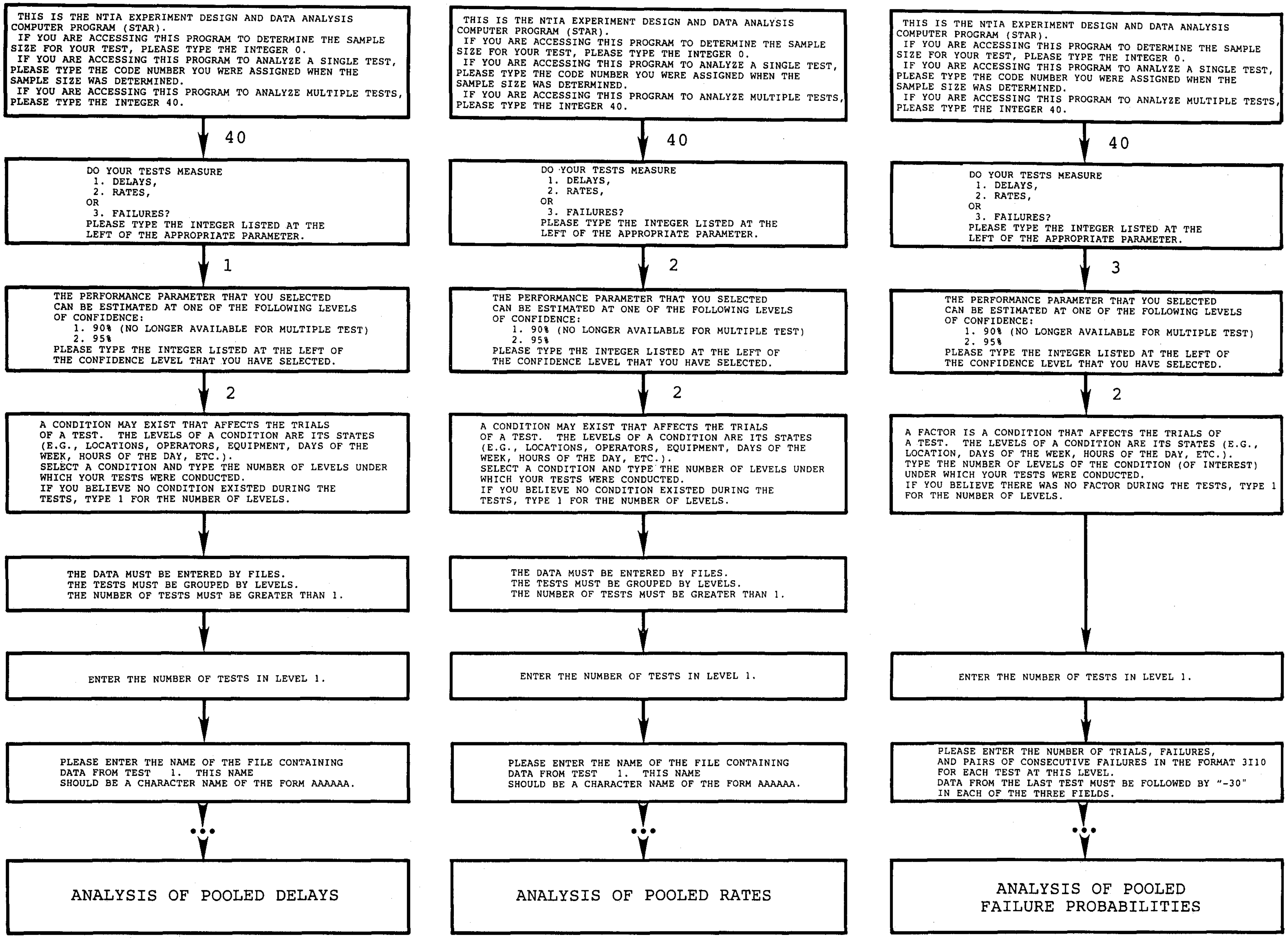

Figure H-1. Example from operator implementation of analysis of multiple tests.

H.2 Failure Probability Parameters

Performance data for failure probability parameters can be entered by keyboard only. Each line contains the number of trials, failures, and pairs of consecutive failures. The format is 3I13. The performance data from the last test at each level must be followed by -30 in each field.

Enter the performance data.

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 (5)		

4. TITLE AND SUBTITLE
Performance Evaluation of Data Communication Services: NTIA Implementation of American National Standard X3.141, Volume 5. Data Analysis

5. Publication Date
August 1995

6. Performing Organization Code
NTIA/ITS.N3

7. AUTHOR(S)
Martin J. Miles

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

9. Project/Task/Work Unit No.

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 analyze a performance parameter from a single test, and from multiple tests conducted at selected combinations of levels of variable conditions.

Single and multiple tests can be analyzed in any of four ways: estimation with known precision, acceptance tests, comparison tests, and tests to determine if a variable condition is a factor. Formulas for these analyses are provided. The formulas are incorporated in an interactive FORTRAN program that can be implemented by either a shell script or an operator. In all cases, dependence between trials is estimated by a first-order Markov chain. Performance parameters are analyzed from multiple tests by, first, pooling trials of tests, then means of tests, and, finally, means of levels of a selected variable condition.

16. Key Words *(Alphabetical order, separated by semicolons)*

acceptance test; American National Standards; analysis of variance; comparison test; data communication systems; dependent trials; estimation; factors; performance measurements; precision

17. AVAILABILITY STATEMENT	18. Security Class. *(This report)*	20. Number of pages
☒ UNLIMITED.	Unclassified	165
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. *(This page)* Unclassified	21. Price: