AD-A143 943

A GUIDE FOR THE APPLICATION OF STATISTICAL ANALYSIS METHODS TO OPERATIONAL TEST AND EVALUATION DATA

THESIS

AFIT/GOR/OS/82S-1

Gregg M. Burgess
Captain USAF

AFIT/GOR/OS/82S-1

A GUIDE FOR THE APPLICATION OF STATISTICAL ANALYSIS METHODS TO OPERATIONAL TEST AND EVALUATION DATA

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air Training Command
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

by

Gregg M. Burgess, B.S.
Captain USAF

Graduate Operations Research

July 1982

Approved for public release; distribution unlimited.

Preface

The research which has resulted in this thesis was accomplished to assist those assigned the task of analyzing, or managing the analysis of, Test and Evaluation data. The guide was written as an appendix so that it can be easily extracted from the thesis. The guide was written so that analysts with very little mathematical background could use it effectively. The success of the thesis will be determined by how easy the guide is to use effectively. The orientation of the thesis also resulted in the inclusion of much introductory material, which will not be useful to all readers. In the attempt to write a very precise and technical guide in an unsophisticated language there are undoubtedly some innaccuracies, but that is the challenge of this kind of research.

I wish to express my gratitude to Lieutenant Colonel Ivy Cook, my advisor, and Lieutenant Colonel Richard Kulp, my reader. Both were very helpful in all phases of the research and the quality of this thesis is a reflection of their guidance and patience.

Contents

List of Figures

Abstract

This research effort was directed toward developing a practical, easy-to-use guide for statistical analysis of Test and Evaluation data. Ease of use and completeness have both been stressed in writing this guide. The guide is written for users with little mathematical background, however should be useful to all Test and Evaluation analysts and managers. The guide includes a discussion of the basic statistical terminology as well as a step-by-step analysis procedure. The analysis procedure is incorporated into a flow chart and includes discussions and examples of individual statistical tests. The statistical tests include parametric and nonparametric tests which can be applied to most test data types.

A GUIDE FOR THE APPLICATION OF STATISTICAL ANALYSIS METHODS TO OPERATIONAL TEST AND EVALUATION DATA

I. Introduction

Department of Defense (DOD) and Air Force directives and regulations require that the evaluation of weapon systems be carried out in such a way that subjective judgement is minimized and quantitative analysis is maximized (Ref 26:2). In accomplishing the quantitative analysis for Air Force test and evaluations, statistical methods have been the most widely and predominantly used. To effectively evaluate a weapon system, the Air Force analyst must bring to his task sufficient operational experience so that the proper issues are addressed and a working knowledge of statistical analysis so that the quantitative data can be analyzed correctly. Unfortunately, many test analysts are assigned to tests based solely on their operational experience, with little or no training with the analytical tools which are needed to accomplish their task. However, no primer or guide exists that would enable those with minimal mathematical background to perform the required statistical analysis satisfactorily.

Objective

The objective of the thesis is to provide a guide that will assist test analysts in the analysis and evaluation of Air Force weapon systems. The guide should be most useful to analysts with little or no mathematical expertise, but may be used as a quick reference by all analysts and managers in the field. The guide is general in nature so that the methods can be used in broad applications, however, not so general that it is not applicable to specific analyses. The guide was written in a step-by-step format so that a minimum of mathematical background is required of the user. The methods recommended in the guide are well known, well documented in referenced sources of various mathematical skill levels, easily applied, and readily available in computer statistical packages.

Background

As an entry level analyst (AFSC 2681), with no analytical or operational experience, I was assigned to the Tactical Air Command F-16 Multinational Operational Test and Evaluation Analysis Branch at Hill AFB, Utah and in Europe. Fully half of the operational analysts had no previous test experience and came from various operational and academic backgrounds. In the search for guides and manuals which would aid us in defining our responsibilities and the purpose of our job, I found only information which was too general and no information which provided guidance in all phases of the analysis of test data.

Conversations with other test analysts and my own experience highlighted the need for a guide which the inexperienced analyst could consult as an introduction to Operational Test and Evaluation (OT&E) and later as a reference for designing and carrying out test data analysis. The Air Force Test and Evaluation Center (AFTEC) was contacted to determine whether or not they viewed a guide of this sort as a worthwhile project which they would support as a thesis sponsor. AFTEC has sponsored the research in hopes that it can be published as an analyst's guide to statistical data analysis.

Scope

The coverage of statistical analysis methods is limited to those which have the most general applicability to operational testing. The methods recommended should provide at least one method of analysis for each type of objective and quantitative data encountered. Parametric and nonparametric statistical tests are used in the guide so that data samples which fail tests of key assumptions or have small sample sizes can be analyzed.

Limitations

The guide has been developed based on the review of twenty-six OT&E Final Reports and Test Plans. Thus some measures of effectiveness (MOE) may not have been identified or anticipated, so that no method is recommended here or a better method may be available.

The guide is by no means self-contained, that is, the analyst must have access to probability distribution tables, statistical references and computer support in some cases.

The guide was purposefully written with a low degree of mathematical complexity, and some advanced concepts were omitted or only referenced. It was felt that simplicity and generality contribute most to the guide's utility and this philosophy has been the over-riding determinant of what has been included.

There is no coverage given to the treatment of qualitative or subjective survey data. While this area may have substantial contributions for OT&E, the analytical methods are secondary to the proper development of the survey. Currently there is no guide available for developing a subjective analysis.

Table I

Test Reports Reviewed

AIM-9M Initial Operational Test and Evaluation, Final Report

AGM-88 (HARM) Air Force Preliminary Evaluation Report

AIM-7F Follow-on Operational Test and Evaluation, Final Report

Simplified Processing Station Phase I Follow-on Operational Test and Evaluation, Final Report

Advanced Medium STOL Transport Analysis Methodology

C-141B Initial Operational Test and Evaluation, Test Plan and Final Report

A-10 Initial Operational Test and Evaluation, Test Plan and Final Report

A-10 Follow-on Operational Test and Evaluation, Final Report

F-111 Initial Operational Test and Evaluation, Final Report

EF-111 Initial Operational Test and Evaluation, Test Plan and Final Report

F-15 Developmental Test and Evaluation Final Report

F-15 Initial Operational Test and Evaluation, Final Report

F-15 Follow-on Operational Test and Evaluation, Final Report

F-16 Initial Operational Test and Evaluation, Final Report

F-16 European Test and Evaluation, Test Plan and Final Report

F-16 Phase I Follow-on Operational Test and Evaluation, Test Plan, Detailed Analysis Products, and Final Report

F-16 Phase II Follow-on Operational Test and Evaluation, Test Plan, Detailed Analysis Products, and Final Report

F-16 Phase II Follow-on Operational Test and Evaluation Suitability, Test Plan and Final Report

II. Methodology

Initial research was conducted to determine if the proposed thesis had already been accomplished in some form. A search of Air Force literature and MITRE Corporation, DTIC, and DLSIE bibliography searches revealed a lack of general OT&E analysis methodologies. The Air Force (AFTEC), Army (OTEA) and Navy (OPTEVFOR) operational test centers were all contacted to determine whether they had a guide for the test analyst. A Navy publication (Ref 23) was the only document discovered which was written expressly for an analyst's use. This document, Analyst's Notebook, does not provide a step-by-step procedure, is of a very specific nature, and does not cover the entire analysis effort. All three services' operational testing centers expressed interest in this research as something which is necessary and unavailable.

The guide itself (Appendix A) has been written with three distinct objectives in mind:

1) an introduction to the purpose, nature and structure of Air Force OT&E
2) an introduction to statistical terms and methods which are commonly used in Air Force OT&E
3) a step-by-step guide for the analysis of OT&E data which enables and encourages the proper handling of design, analysis and interpretation by the analyst.

The first two objectives were accomplished after reviewing current Department of Defense and Air Force directives and regulations, related books and technical reports as well as a review of statisti l literature and the Final Reports from 17 recent Air Force tests (Table I).

The step-by-step method for the guide was developed into a flow chart to minimize the bulk of the guide and to provide a better graphical representation of the analysis procedure. Where possible the required calculations are included in the flow chart; however, due to the complexity of some steps, individual descriptions are also included in the guide with references, examples, and the correct program from both SPSS and BMD systems. The analysis procedure outlined in the guide includes the testing of the key assumptions of the tests prior to their use with an alternative path if an assumption is rejected. With the guide in this format, it is hoped that the important intermediate steps in the analysis will not be overlooked.

The flow chart was developed using some non-standard statistical notation in the interest of generality and to minimize the complexity of the flow chart itself. The flow chart serves the dual purpose of determining the type test and sample size in the design phase as well as determining the proper test in the analysis phase.

In addition to the flow chart, descriptions and examples of the statistical tests, the guide also provides the analyst with suggestions for reporting the results of the analysis, contacts within the Air Force for advice on all phases and a bibliography which lists the references in increasing order of mathematical complexity.

III. Results and Conclusions

The review of seventeen USAF Test Reports not only provided the basis for the scope and coverage of the guide (Appendix A), but also highlighted the need for such a guide. In all of the reports there was very little evidence of any statistical analysis being done at all. In the few reports which did report the results of the statistical analysis, the treatment indicated little understanding of the methods used. This problem, too little use of sound statistical analysis, may be the result of too few qualified test analysts or the reluctance of test management to include analyses which they feel are too complex or won't be understood. Whatever the case, a standardized, easy-to-use, easy-to-understand guide using basic statistical analysis methods could be used by both analyst and test manager alike to overcome the problem.

That easy-to-use guide is the result of this Air Force Test and Evaluation Center (AFTEC) sponsored research. The guide includes: a discussion of Air Force Test and Evaluation (T&E), a discussion of basic statistical terms and their relation to the test environment as well as a step-by-step guide for using statistical analysis on T&E data. Much of the background material could not be omitted or abbreviated if the guide is to be useful to all the prospective users; however the actual analysis section was reduced, by the use of a flow chart. The flow chart is helpful in that it presents all of the steps required in an analysis, some of which are often neglected. The flow chart will ensure that alternative methods are at least discovered if not used and should result in better analyses if followed correctly.

The Air Force T&E community is in need of a standardized method for analyzing test data; the attached guide should be a good beginning toward this end. Further work in the area of subjective evaluation, and other measures of effectiveness could be accomplished as well as the study of the applicability of specific tests to T&E data. Test and Evaluation is a difficult area in which to apply statistical analysis because the objective of the test must be weighed against the cost of testing; however it is also an area where good statistical analysis is needed. This guide may form the basis for an improvement in current T&E analysis practices.

The Test Plans and Test Reports which were reviewed and conversations with Air Force, Navy and Army test managers and analysts indicate that the conduct of Operational Tests does not satisfy the intent of the pertinent regulations concerning quantitative analysis of test data. The Operational Testing environment does not allow for significant input from the analysis function concerning experimental design or test conduct. This problem can not be corrected without significant changes in the role of test analysts and the attitudes of test managers. However the attitude of test managers may be altered by their education in statistical analysis methods. This guide may be as useful, to the analysts, in their bosses hands as it will be in their own.

To make this sort of step-by-step guide even more useful to test analysts it may be practical to develop an interactive software package which could be used to perform data analysis without accessing other statistical packages, e.g. BMD and SPSS, or performing any hand calculations. The development of such an analysis tool for mini-computers or larger systems would go a long way toward standardizing and improving current test data analysis.

The current state of USAF Operational Test data analysis is such that improvements must be made in order to:

1) improve the overall quality of the analyses,
2) standardize analysis procedures, and
3) develop a more knowledgeable and understanding test management, concerning test data analysis.

These changes should result in the development of better testing techniques and more confidence in test results. A guide or similar analysis tool is the first step in obtaining these improvements.

Recommend that this step-by-step guide be the basis for a standardized analysis procedure to be instituted by AFTEC and that further research be directed towards an interactive computer analysis system.

APPENDIX A

A GUIDE FOR THE STATISTICAL ANALYSIS OF OPERATIONAL TEST DATA

Contents

I. INTRODUCTION

The analysis of Operational Test and Evaluation (OT&E) data requires the mixing of operational knowledge and statistical theory. Applying theory to any practical problem is usually difficult; applying statistical theory to Test and Evaluation (T&E) analysis is no exception. The guide presents a step-by-step procedure which can be used for OT&E data analysis by the analyst and manager. It may be of most help to the new OT&E analyst who does not have an extensive analytical background. However, as a quick reference to the most common statistical tests, the guide can be useful to anyone involved in T&E.

The purpose of the guide is to present a brief, simple and complete plan of attack for analyzing the most common OT&E objectives. In the interest of brevity and simplicity it will be presented without much statistical detail but will concentrate on maximizing its usefulness. The guide is complete in the sense that it presents at least one method for dealing with each of the areas to be analyzed.

The guide is based on recent Final Reports and Test Plans from the T&E of current USAF weapon systems. In the reports reviewed, three basic types of measures of effectiveness (MOE) were identified. Other MOE's may be available, but since these three types will provide a solid base, the guide is limited to them: proportions, averages, and measures of dispersion.

II. THE PHILOSOPHY OF TESTING

Military Test and Evaluation has developed as a major step in the systems acquisition process "... to identify, assess and reduce the acquisition risk, to evaluate operational effectiveness and operational suitability, and to identify any deficiencies in the systems." (Ref 26:2) Air Force T&E is divided into two major categories: OT&E and Development Test and Evaluation (DT&E). Analysis and test conduct are very similar for both, however, they have separate purposes. While this guide concentrates on OT&E, many of the applications can be made to DT&E directly.

Operational Test and Evaluation is that testing "... conducted to estimate a system's operational effectiveness and operational suitability, identify needed modification, and provide information on tactics, doctrine, organization and personnel requirements." (Ref 27:3)

Development Test and Evaluation is that "... testing and evaluation used to measure progress, verify accomplishment of development objectives, and to determine if theories, techniques, and material are practicable; and if systems or items under development are technically sound, reliable, safe, and satisfy specifications." (Ref 26:30)

OT&E differs from DT&E in that OT&E tests systems which have progressed through the procurement process to the point where the system can actually function in an operational environment. OT&E is the first opportunity for the using command to critically evaluate the system. This can often lead to an adversary relationship with the engineering and procurement functions which execute the development phases.

The importance of OT&E to an effective procurement process cannot be overstated. In August 1967 the President's Scientific Advisory Committee (military aircraft panel) stated "... we believe most of our previous failures to be prepared for wars we fight would have been thoroughly exposed had an adequate program of testing and evaluation existed." (Ref 31:8) The necessity for honest, effective testing has become clearer in recent years due to the ever increasing cost of military weapon systems and the ever increasing threat.

The analyst's responsibility in the OT&E environment is to carry the test beyond the realm of subjective judgment, by providing the methods and techniques for quantifying and analyzing system performance. "Quantitative data must be used, to the maximum extent practical, to show that the major objectives have been met. Subjective judgment, relative to system performance, must be minimized." (Ref 26:2)

The analyst is responsible to the Test Director for the design of the test, which should produce quantifiable results that can withstand the scrutiny of operational and scientific experts. Additionally, the analyst is responsible for analyzing the data collected during the test. In carrying out the analysis function, the analyst must above all else maintain impartiality. Interpretation of the results of any analysis is often done by someone other than the analyst. To enable others to do this correctly, the analyst should provide an analysis which is objective and repeatable. The statement of all the assumptions is required as well as any other possible limitations. Basically, the analyst's function is to translate a complex question into a number of options which have been enumerated and can be compared. The analyst will be responsible for presenting an analysis that aids the decision maker; however, final responsibility for any decision always rests with the decision maker alone.

The budget constraint will most often be felt because of its affect upon the amount of data which can be collected. Whether testing aircraft or munitions, there is usually a substantial expenditure associated with each sortie or trial. A limited budget reduces either the number of objectives which can be answered or the number of trials devoted to a single objective. The budget constraint will almost always be in effect, so that the decision must be made between limiting the objectives or limiting replications. The objectives of any test should be scoped down so that each area is covered adequately.

The lack of measurement instrumentation is often a constraint due to the cost of making it available or because of integration time requirements. To perform a test with quantifiable results will often require special equipment, like video tape recorders (VTR) or time, space, position instrumentation (TSPI), to provide the necessary data. A lack of test instrumentation will result in less accurate quantitative data; however, instrumented test systems alone are not always enough. Quite often support systems or threat simulators must also be instrumented to make any of the data useable. Test Instrumentation must be a well planned and coordinated effort to be effective, but the rewards are often worth the additional effort.

"OT&E should be accomplished in an environment as operationally realistic as possible." (Ref 27:3). Often the desire to test under operationally realistic conditions is in direct conflict with the collection of quantitative data. In some cases, test instrumentation cannot function in realistic conditions and the need for realism must be weighed against the need for quantifiable data.

Perhaps a more basic question than "What are the conditions under which the data is collected?" or "What data should be collected?" is the question "How accurate is the data?" Data accuracy can be affected by test design but more often is affected by the data collection process. Usually an analyst's responsibility, data collection must be performed in the most accurate way. Analyzing OT&E data is very difficult because of the operational influences; however, the job becomes nearly impossible to do well if excessive error is introduced at the point where data is collected, reviewed, or processed.

The main tools for data collection are test instrumentation and manual data forms. Before any data is collected, the objective(s) for which it will be used should be identified and an analysis technique should be specified. Data collection forms should be designed to minimize the possibility of human error by limiting their content to essential items and by proper formatting. Simple forms will reduce the possibility of incorrect data collection. The forms should be self-explanatory and easy to follow. Proper design of forms can allow the operators to complete forms and eliminate the need for dedicated data collectors to conduct interviews.

Subjective judgments contribute much to current OT&E final reports. The reasons for this include the lack of adequate instrumentation, unquantifiable test objectives and the lack of accurate quantitative data. The use of subjective data analysis has been neglected for the most part in USAF tests. Subjective survey data can never replace quantifiable results; however, when analyzed with the proper analysis techniques, it can be used to lend support to the subjective judgment.

III. STATISTICS AND TESTING

The use of statistical analysis methods is not required in Air Force testing; however, when compared with other techniques, such as decision analysis or cost-benefit analysis, it becomes clear that statistical analysis has many advantages. Statistical methods are familiar to many decision makers. They are well documented, easy to interpret and easy to apply. The majority of OT&E analysis has been done using statistical methods and there is a lot of information in technical reports on specific applications.

To effectively use statistical analysis a "correct" technique must be found; this technique must be applied correctly; and the results must be interpreted correctly. OT&E is not an easy area to apply statistical methods because of the presence of so many uncontrollable and unmeasureable variables. Creativity and caution must both be applied in equal amounts in all phases of the analysis. Creativity - because many times OT&E analysis is completely original, has not been documented, or at the very least, is a new application of an old method. Caution must be exercised because OT&E data analysis is more than an art. It is also a science, with some strict rules which cannot be broken if the analysis is to withstand inspection by professional statisticians.

The objective of the analysis must be translated from the language of the Test Plan into a statistical hypothesis.

This hypothesis (H_O), called the <u>null hypothesis</u>, made in the form of a mathematical formula, can be tested, and, if rejected, an <u>alternate hypothesis</u> (H_A) which is the exact complement can be accepted. For example:

H_o: Circular Error Probable (CEP) $\geq$ military specification (e.g. 100')

H_A: CEP < military specification.

If the stated hypothesis (H_O) is rejected statistically, then H_A must be accepted.

Choosing the proper statistical model for the combination of objective and data characteristics is the hardest step in the procedure and can most affect the results. The first criterion in choosing a model is data dependent, but subject to the analyst's decision. Models are of two basic types: parametric and nonparametric. <u>The parametric models</u> and the associated tests are based on relatively specific assumptions concerning characteristics of the population from which the data is drawn. The assumptions include: measurability of the data, shape of the population probability distribution (e.g., normal distribution) and the number of populations being sampled. The effectiveness of the parametric tests is directly related to the validity of these assumptions. However, some of the tests still function relatively well with slight violations to the assumptions concerning the shape of the sampled distribution. This is especially true of tests of averages.

The nonparametric tests are sometimes referred to as distribution-free because they do not contain strict assumptions concerning the probability distribution of the sampled population. The nonparametric tests do make assumptions of measurability and population characteristics, but they are more general in nature and are applicable more often. Because of the relaxed assumptions, the nonparametric tests will not identify statistical differences as readily as the parametric tests (Ref 9:32-33); however, this is true only when the assumptions of the parametric tests hold. In cases where the assumptions of the parametric test do not hold, the results of parametric tests are unpredictable. Thus the acceptance or rejection of a statistical hypothesis may be caused by the failure of the assumptions rather than the validity or invalidity of the hypothesis being tested. For further discussion of nonparametric statistics see (Ref 9:30-34) and (Ref 10:91-93).

Another important use for nonparametric statistics in the OT&E environment is in its applicability to qualitative data. Qualitative or subjective data can usually be collected in terms of rankings or categories. The advantage here is that subjective data is inexpensive to collect, usually results in large samples and can be used to support quantitative results to evaluate an unquantifiable objective. Multivariate analysis methods, such as multiple regression and discriminant analysis have also been applied to subjective data with success outside of OT&E. Any analysis of subjective data is very dependent upon the content and format of the collection form. The design of subjective surveys is nearly a science in itself and should not be undertaken carelessly.

IV. STATISTICAL TERMS

The application of statistical theory to test data is most often done with classical statistical hypothesis testing and its associated techniques. A clearly stated objective is the cornerstone of a good analysis, but in OT&E, just as important is the measure of effectiveness (MOE). The MOE will be used to determine if the objective has been met and, as the name implies, must be measurable. The scale of measurement can be:

1) Nominal or Categorical - the least preferred type of data, e.g. aircraft type (F-4, F-16); air combat outcome (win, loss, draw).

2) Ordinal or Ranked - e.g. altitude (low, medium, high), subjective evaluation (inferior to current systems, as good as current systems, superior to current systems).

3) Interval - the distance between observable values is consistent, however the zero is not an absolute value, e.g. temperature in F^{o} or C^{o}; IQ or SAT scores.

4) Ratio - the data is interval and an absolute zero exists which makes ratios comparable, e.g. distance (feet), probability of kill, normal acceleration (g's), velocity (knots), temperature in ^{o}K.

A further distinction in the measurability of data which concerns only ratio and interval measures is the notion of discrete and continuous data. Continuous data should be able to take on all values between the extremes whereas discrete data can take on only a finite or countably infinite, e.g. all integers, number of values For further discussion of the scales of measurement see (Ref 9:21-30).

How well an objective can be analyzed will be determined by the characteristics of the MOE. The MOE must be readily measured in an unambiguous way and should relate directly to the objective. The more that is known about the probabilistic nature of the MOE, the easier it will be to determine the correct analytical model. If nothing or little is known about the MOE, it will usually be best to assume that it is normally distributed, if it is continuous, until shown to be otherwise.

In statistical applications the data that is collected is termed a sample and is assumed to be a random selection from an underlying population. This population is often assumed to have a specific statistical distribution, e.g. normal , which is the basis of the parametric tests in this guide. The term "statistical confidence" is one that is common to nearly all applications and is represented by the term $(1-\alpha)$ x100%. The significance level, α, is the probability of rejecting a stated hypothesis (H_o) when it is true, called a type I error. Confidence is important when H_o has been rejected, as it gives an indication of how likely it is that you have drawn the correct conclusion. Because the collected data is a random sample of a larger population, it is very likely there will be some discrepancy between the value of a sample parameter and its "true" value.

Generally α is chosen to be .01 or .05 which results in a 99% or 95% confidence level, respectively. It must be noted that as confidence goes up the probability of accepting H_o goes up, regardless of whether H_o is true or not. The probability of accepting H_o when it is indeed false, a <u>type II error</u>, is denoted by β. For a given sample of data, if the probability of rejecting H_o when it is true (α) goes down, the probability of accepting H_o when it is false (β) goes up. Both α and β are affected by the amount of data collected; as sample size goes up, β will drop for a given α, or α can be lowered while β is unaffected. Since sample size is constrained due to costs in OT&E, reasonable values of β are not possible at low significance levels (α). When dealing with "small" samples, it is best to set α as high as is comfortable, without loss of credibility. In operational tests, α as high as .2 can be used. For further discussion of significance and power see (Ref 7:389-390), (Ref 3:284-288).

When using statistical methods to analyze OT&E data, there are three distinct areas of interest:

1) <u>estimation of a MOE</u>, e.g. CEP, Probability of Kill (P_K);

2) determination of a <u>confidence interval</u>, i.e. finding the interval about an estimate which has a known probability of including the "true" mean value of the MOE;

3) <u>hypothesis testing</u>, i.e. comparing sample values of two or more MOE's or comparing one value against a known or assumed value.

<u>Hypothesis testing</u> is the main statistical tool used in OT&E data analysis. The hypotheses under consideration can take on many different forms. Two of the most common are:

1) hypotheses concerning population parameters, e.g.

a) H_o: CEP $\geq$ 100 feet

H_A: CEP < 100 feet

b) H_o: MTBF (component A) = MTBF (component B)

H_A: MTBF (component A) $\neq$ MTBF (component B)

2) hypotheses concerning the relationships among a response variable (MOE) and one or more independent variables, e.g.

Given: $y = B_o + B_1X_1 + B_2X_2 + e$

y : detection range

X_1: target altitude

X_2: target airspeed

B_o, B_1, B_2: coefficients

e : random error term

H_o: $B_1=B_2=0$

Detection range is not a linear function of target altitude or airspeed.

H_A: $B_1\neq0$ or $B_2\neq0$ or $B_1\neq0$ and $B_2\neq0$

Detection range is a linear function of either target altitude, target airspeed or both.

The estimation of a population parameter from a sample of data can be carried out by common mathematical equations. In this guide these parameters include:

1) sample mean ($\bar{x}$), an estimate of the population mean (μ);

2) sample variance (s^2), an estimate of population variance (σ^2);

3) and sample probability of success ($\hat{p}$), an estimate of the population probability of success (p).

The term confidence interval applies to the range of values about a sample estimate which has a known probability ($1-\alpha$) of including the true population value. Because the data sample is a random selection, a confidence interval can be calculated which takes into account the randomness. The larger the confidence level $(1-\alpha)\times100\%$, the wider the interval will be.

To determine the necessary sample size for a given confidence level, the "maximum error of the estimate", E , must also be known. The maximum error (E) is the largest acceptable difference between the estimate and the population value which is measurable. For a given significance level (α), the sample size (n) must be increased if the maximum error is decreased.

Review of Statistical Terms

MOE: measure of effectiveness

H_0: stated or null hypothesis

H_A: alternate hypothesis

α: significance level - probability of a type I error

Confidence level: $(1-\alpha)x100\%$

Type I error: reject H_o when it is true

β: probability of type II error

Power: $1 - \beta$

Type II error: fail to reject H_o when it is false

n: sample size

E: maximum error of the estimate

x_i: ith sample observation

$\bar{x}$: sample mean, an estimate of μ, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

μ: population mean

s^2: sample variance, an estimate of σ^2, $s^2 = (\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)/(n-1)$

σ^2: population variance

$\hat{p}$: sample probability of success, an estimate of p, $\hat{p}$ = (# of successes)/n

p: population probability of success

Relationship

- For constant n, an increase in α decreases β, and vice versa
- For constant β, an increase in n decreases α

- For constant α, an increase in n decreases β
- As confidence level increases, the confidence interval is wider for fixed n.

V. MEASURES OF EFFECTIVENESS

Recent OT&E reports have made use of three MOE types:

1) Proportions

2) Averages

3) Measures of Dispersion

These types of MOEs will adequately measure most objectives to be tested; however, in no way do they comprise all of the possibilities. Analysis of only these MOE types will be covered in this guide.

Proportions

Examples of measures which are proportions are:

1) Probability of kill (P_k)

2) Percent of total launches that are successful

3) Attrition rate

4) Number of targets sighted vs. number present

Proportions can be binomial or multinomial measures depending on the number of possible outcomes of a single trial. The binomial trial has only two outcomes possible, e.g. success or failure and kill or no kill, while the multinomial trial can have more than two outcomes, e.g. win, lose, draw and kill, no kills, killed. The MOE associated with binomial trials will generally be the probability associated with one of the outcomes, whereas in the multinomial case the MOE could be the probability of one or more outcomes as well as the ratio between outcomes or groups of outcomes. The sum of all probabilities of all outcomes must add up to one in both the binomial and multinomial experiments. The assumptions which most affect tests of proportions are the assumptions that the probability of an outcome remains constant and that the trials are independent, which means that the outcome of a trial is not affected by the outcomes of previous trials. Both distributions are fully specified by the total number of observations, n, and p_i, the probability of the ith outcome, for all i.

Tabled probabilities are available for the binomial distribution for n up to 20 and for p a multiple of .05. When testing below n of 20, a linear interpolation of the tabled values is acceptable (Ref 15:183), but the best approach with small n is to compute the probability exactly. When n is greater than 20, the normal approximation is acceptable for p between .1 and .9 and the Poisson approximation can be used when p or 1-p is less than .1.

Binomial

two outcomes (success/failure)

p:	the probability of success
q=1-p:	the probability of failure
n:	total number of observations
y:	total number of success
$\hat{p} = \frac{y}{n}$:	estimate of p

Exact: $P(y=k) = \binom{n}{k} p^k q^{n-k}$ (probability that y=k)

$$P(y \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i}$$

$$\binom{n}{k} = \frac{n!}{(n-k)!\ k!}$$

Normal Approximation: $Z = \dfrac{k \pm \frac{1}{2} - np}{\sqrt{npq}}$

If $Z \geq 0$, $P(y \leq k) = F(Z)$ (Ref 15:127)

If $Z < 0$, $P(y \leq k) = 1-F(-Z)$ (Ref 15:127)

The $\pm \frac{1}{2}$ is a continuity correction to be used if $n \leq 100$.

Use $-\frac{1}{2}$ when $k<np$ and $+\frac{1}{2}$ otherwise.

Poisson Approximation: $\lambda = np$, $x' = k$

$P(y \geq k) = T(x', \lambda)$ (Ref 15:213)

$P(y \leq k) = 1-T(x'+1, \lambda)$ (Ref 15:213)

Multinomial

k possible outcomes

n: total number of observations

p_i: the probability of the ith outcome

$$\sum_{i=1}^{k} p_i = 1$$

y_i = total number of observations of outcome i

$$\hat{p}_i = \frac{y_i}{n}$$

Averages

Some examples of MOE's which are <u>averages</u>, or <u>means</u>, are:

1) average detection range
2) mean miss distance (MMD)
3) mean time between failures (MTBF).

These measures are the most common to OT&E, and the statistical tests are generally known. Two basic approaches are covered in dealing with averages. The distribution of the measure can be assumed 1) to be normal - that is, continuous, unimodal and bell-shaped symmetrically about the mean or 2) to be non-normal. Nonparametric tests must be used if the population is assumed to be non-normal or distribution-free. The proper parametric or nonparametric test is determined by the number of samples to be tested. Linear regression and analysis of variance (ANOVA) can be used, with data assumed to be normal, to determine the relationship of the MOE to other factors or independent variables.

Normal Distribution

unimodal: one "hump" or "peak"

continuous: data point y_i can take on all values

symmetric: the probability function is a mirror image about the mean (μ)

interval: $-\infty$ to ∞, the true normal distribution cannot be absolutely restricted from taking on <u>any</u> value although the probability of taking on values in extreme regions is essentially zero.

$x \geq 0$

$$P(y_i \leq x) = F(x) \qquad \text{(Ref 15:127)}$$

$x < 0$

$$P(y_i \leq x) = 1-F(-x) \qquad \text{(Ref 15:127)}$$

Measures of Dispersion

The primary measures of dispersion are circular error probable (CEP), range error probable (REP) and deflection error probable (DEP). These three measures are all based on the assumption that the data points are distributed normally on one (REP,DEP) or both axes (CEP). If this is not found to be the case, the mean miss distance (MMD) can be used as another measure. The CEP measure assumes that the dispersion along both the range and deflection axes is normally distributed with equal variances, that is, the pattern is circular normal. The REP and DEP measures were developed to account for dispersion patterns which are elliptical in shape. Circular error probable can be thought of as the radius of the circle which encompasses half of the population's points. Because CEP is the most accepted measure of dispersion for bombing or navigation accuracy, it will often be required even if the pattern is not circular. When computing CEP, REP, or DEP, the mean impact point (MPI) should be used and this position should be reported with respect to the target center or desired mean impact point. REP and DEP are similar to CEP in that they give borders in which 50% of the impacts should occur on one axis.

x_i: lateral miss distance

y_i: longitudinal miss distance

MPI: $\bar{x}$, $\bar{y}$; mean point of impact

$$REP = .6745 \sqrt{s_y^2}$$

$$DEP = .6745 \sqrt{s_x^2}$$

For non-circular patterns a 50% circle can be computed using these formulas:

$$k = s_x / s_y$$

If $k \geq .3$, $CEP = .6152\, s_x + .5640\, s_y$

If $k \leq .3$, $CEP = (.82k + .007)\, s_x + .6745\, s_y$

*NOTE: These formula assume $s_y > s_x$; if not, reverse s_x, s_y in all formulas.

When dealing with circular patterns the 50% circle can be computed as:

$$CEP = 1.1774 \sqrt{\frac{s_x^2 + s_y^2}{2}}$$

A more generalized equation, making use of the Rayleigh distribution, for the probability distribution of the radial miss distance, $\sqrt{X^2+Y^2}$, is:

$$P(r \leq R) = 1 - (\tfrac{1}{2})^{-R^2/CEP^2}$$

r: radial miss distance

R: a specific radius

For a complete discussion of CEP, REP, and DEP see (Ref 28).

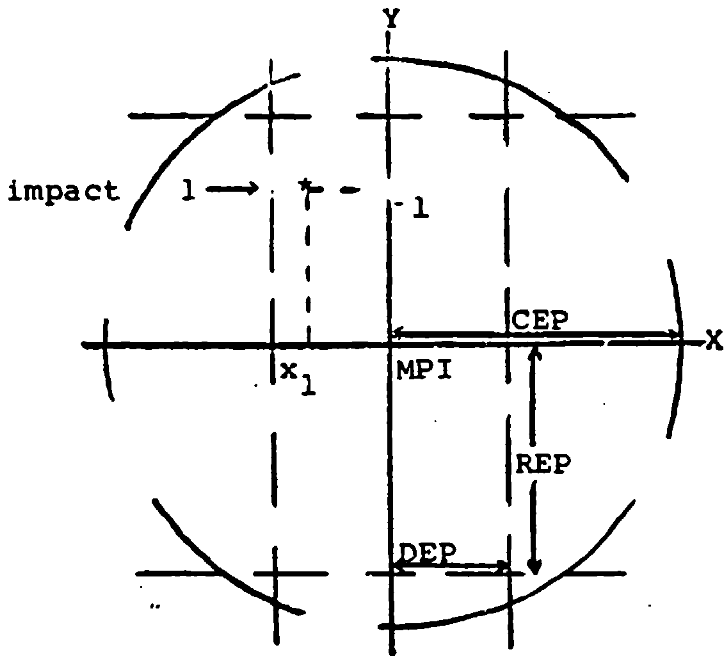


Figure A-1. Measures of Dispersion

VI. DATA ANALYSIS

The procedure which OT&E analysts have used to proceed from the planning phase to writing the final report is not something that is easily standardized. However, every analysis has certain aspects which are always present. The objective is to provide a procedure which is general enough to cover all aspects of OT&E analysis and yet be specific enough to be used as a ready reference by all analysts.

The basic steps in an analysis are:

A. SPECIFICATION

1) Determine the objective of the analysis.

2) Determine the MOE that will best quantify the results of the analysis and best relates to the objective.

B. EXPERIMENTAL DESIGN

1) State the hypothesis (H_o) for which rejection is desired and the alternative (H_A) which satisfies the objective.

2) Determine what level of confidence ($(1-\alpha)x100\%$) is required.

3) Determine which statistical test is most suited to the MOE and stated hypothesis (H_o).

4) Determine the sample size required based on the significance level (α) and the maximum error of the estimate (E). (Not always possible)

C. DATA COLLECTION

1) Collect the data, following the experimental design closely, and keeping the assumptions of the selected test in mind.

D. ANALYSIS

1) If possible, test each of the assumptions of the chosen test with the collected data.

2) Using the results of the test of assumptions, determine which test is best suited for the data.

3) Apply the statistical test at the confidence level originally stated.

4) Interpret the results of the test, stated in terms of rejecting or failing to reject the stated hypothesis (H_o).

E. REPORTING RESULTS AND CONCLUSIONS

A. Specification

Determination of the objective and the MOE are completed with the publication of the Test Plan. Three criterion should be applied in the decision:

1) operational relevance

2) measurability

3) quantifiability.

B. Experimental Design

The design of the experiment is the most critical phase of the analysis in terms of getting the most out of the resources expended. The number of replications, or sorties, of each kind should be based on analytical, not operational considerations. This cannot be done without a complete design, accomplished before the data is collected. Once the MOE, hypothesis and confidence level are determined the appropriate statistical test must be selected.

1. Proportions

The analysis of MOE's which are proportions can be accomplished in three basic ways. If the experiment is indeed binomial with two outcomes and a constant probability of success (p), the exact binomial or an approximation can be used. If the probability of success is changing, linear regression can be used to determine the behavior of p as a function of other variables. In the case of multinomial experiment (more than two outcomes), the Chi-square contingency table can be used to determine equality of p_i across samples.

2. Averages

The analysis of MOEs which are averages can be carried out using either parametric or nonparametric (distribution-free) statistical tests. Both methods have certain advantages and disadvantages; however, the choice should be based solely on the actual or expected characteristics of the data. For a complete discussion of the subject see (Ref 9: 30-34), (Ref 10:91-93).

The disadvantages of nonparametric statistical tests include:

a) When the assumptions of the parametric tests are met, for a given sample size and significance level nonparametric tests are often not as powerful as the best parametric test, that is, the probability of failing to reject an erroneous H_o is greater.

b) Required sample sizes cannot be estimated accurately for any nonparametric test.

c) Nonparametric tests are not as sensitive to small deviations in some distribution measures, e.g. shape, position.

d) Nonparametric tests are less common, there are fewer references, and there are fewer people well acquainted with their use.

The <u>advantages of nonparametric statistical tests</u> include:

a) Nonparametric tests have fewer and less restrictive assumptions which make them applicable to a wide variety of data.

b) Nonparametric tests can be used to compare measures from differing distributional types.

c) Nonparametric tests can be used with ranked (ordinal) and categorical (nominal) data.

d) Nonparametric tests can be used when nothing is known about the underlying distribution.

e) Nonparametric statistics often have relatively simple mathematical derivations and the development is more easily understood.

f) Nonparametric statistical tests must be used unless the population distribution is known to be normal (Ref 9:32).

While it might seem that nonparametric statistics have more general applicability, the parametric tests presented here will generate accurate results even with minor deviations from some assumptions. The choice of a particular test within each category, parametric or nonparametric, is very similar in that the choice is usually based on the number of samples. The parametric tests used here will all assume a normal distribution.

The one and two sample tests of means are based on comparing sample means and determining the probability that they are statistically different. The analysis of variance (ANOVA) and regression methods are based on the linear relationship between a dependent variable and other independent variables.

3. Measures of Dispersion

Use of CEP, REP, DEP and variance measures are generally confined to the analysis of weapons delivery or navigation accuracy. The analysis of means can be applied to accuracy scores; however, the standard practice is the use of CEP. The CEP is calculated about the midpoint of the dispersion pattern (MPI). The relative position of the MPI to the target can be considered system bias.

With no prior knowledge of system accuracy or dispersion patterns it will be best to assume a non-circular pattern and use the adjusted CEP formula for elliptical patterns. While the delivery of free-fall ordnance from jet aircraft will almost always result in elongated patterns, navigation accuracy and guided weapons may in fact have circular distributions.

The difference between the MPI and the target will only be true system bias when all other errors, (e.g. aiming error) have been accounted for and corrected.

4. Determining Sample Size Required (N_r)

In many cases it is not possible to determine how many samples will be required for a given combination of test, significance level (α) and maximum error of the estimate (E). When planning on the use of nonparametric statistics there are no sample size calculations available. However, for some parametric tests the required sample size (N_r) is easily calculated.

<u>TEST</u>	<u>N_r</u>
One sample test of proportion	$N_r = \hat{p}(1-\hat{p})\ [W_a/E]^2$
	W_a=x: F(x)=1-a (Ref 15:127)
	$a = a_1/2$
	E: maximum error of the estimate
	$\hat{p}$: an estimate of p (If no estimate is available use $\hat{p}=\frac{1}{2}$)
One sample test of a mean	$N_r = (\hat{s} \cdot W_{F,n}/E)^2$
	$W_{F,n} = t$ (Ref 15:283)
	$F = 1-a_{\frac{1}{2}}$
	$n = \alpha$
	$\hat{s}$ = an estimate of σ
	Note: If σ^2 is known substitute σ for $\hat{s}$
Two sample test of means	$N_{r_i} = \hat{s}_i(\hat{s}_1+\hat{s}_2)(W_{F,n}/E)^2$
	$W_{F,n} = t$ (Ref 15:283)
	$F = 1-a_{\frac{1}{2}}$
	$n = \alpha$
	Note: If σ_i^2 is known substitute σ_i for $\hat{s}_i$

C. Data Collection

The process of collecting data for OT&E is carried out in many ways and is not always an analyst's responsibility. The analyst should be given some input because the quality of any analysis is only as good as the data upon which it is based. For this reason the analyst should do all that can be done to see that data collection is carried out properly. This can be accomplished by participating in the collection and by verifying the quality of all data by inspection.

D. Analysis

The data analysis can be divided into three basic tasks:

1) testing the assumptions
2) testing the stated hypothesis (H_0)
3) interpreting the results .

1) Testing the assumptions

The test of assumptions are those tests which determine whether the test selected during the experimental design phase is in fact the proper test. Unlike the testing of hypotheses in the next section, the hypothesis when testing assumptions will be that the assumption is valid. Therefore assumption tests will not detect some deviations from the assumptions.

Assumption	Test	References
Randomness	Runs Test (pp A-35)	(9:52-58) (14:34-38)
Proportions		
Constant p	Chi-Square goodness-of-fit Test (pp A-52)	(9:42-47) (10:190-198)
Averages		
Normality	Lilliefors Test (pp A-57)	(10:357)
	Chi-Square goodness-of-fit Test (pp A-52)	(9:42-47) (10:190-198)
Known variance	Chi-Square Test (pp A-43)	(5:75) (7:396)
Equality of two variances	F Test (pp A-43)	(5:90) (7:400)
Equality of k variances	Levene's Test	(5:253)
	Bartlett's Test	(5:252) (7:510)
Measures of Dispersion		
Normality	Lilliefors Test (pp A-57)	(10:357)
MPI = Target	F Test (pp A-43)	(7:400) (27:2)
X-Y Correlation	t Test	(27:2)

2) Hypothesis Testing

Tests of hypotheses can be formulated in many ways; in this guide the most general cases will be used to ensure broad applicability.

a) <u>One sample case</u>

(1) H_o: $|Q-Q^*| \leq E$

H_A: $|Q-Q^*| > E$

(2) H_o: $Q \leq Q^* + E$

H_A: $Q > Q^* + E$

(3) H_o: $Q+E \geq Q^*$

H_A: $Q+E < Q^*$

These hypotheses allow comparison of a population measure or parameter, Q, against a hypothesized value, Q^*, with a meaningful difference, E. The difference, $E \geq 0$, is the max error of the estimate when the objective is to detect any significant difference. Case (1) is called a <u>two-tailed test</u> because differences are detected when the sample measure is significantly different, larger or smaller, than the hypothesized value, Q^*.

b) <u>Two sample case</u>

(4) H_o: $|Q_1 - Q_2| \leq E$

H_A: $|Q_1 - Q_2| > E$

(5) H_o: $Q_1 \leq Q_2 + E$

H_A: $Q_1 > Q_2 + E$

Note: estimate of $Q_1 \geq$ estimate of Q_2

These hypotheses allow the comparison of two populations' measures, Q_1 and Q_2. By setting E=0 the hypotheses reduce to a direct comparison of the measures: (4) H_o: $Q_1=Q_2$, (5) H_o: $Q_1 \leq Q_2$.

c) <u>More than two sample case</u>

(6) H_o: $Q_1=Q_2=\ldots Q_k$ or

H_A: At least one of the measures is not equal to all others

(7) H_o: $B_1=B_2=\ldots B_k=0$

H_A: At least one of the slope coefficients, B_i, is not equal to zero.

In hypothesis (6) the comparison is made among all sample measures for equality. Hypothesis (7) pertains only to tests where linear regression is used.

d) <u>Hypotheses for testing assumptions</u>

(8) H_o: The assumption holds

H_A: The assumption does not hold.

Hypotheses for tests of assumptions are constructed so that the burden of proof is on rejecting the assumption. Unlike other testing, the hypothesis is not made so that the analyst wants to reject H_o.

The <u>acceptance region</u> is the range of values of the test statistic, T, which <u>do not</u> support rejection of the H_o. The acceptance region differs with the type of hypothesis and the particular test employed.

Reporting Results

The final step in this guide is the reporting of conclusions based on the results of the statistical test. The interpretation of the results must be left to the tester; however, the conclusions reached from the tests can be generalized. It should be remembered that no statistical test has ever proved anything; they do lead the tester to conclusions concerning the probability that the stated hypothesis (H_0) is or is not true.

VII. STEP-BY-STEP GUIDE

The following step-by-step guide will cover the analysis of test data from the initial tests of assumptions to the reporting of the conclusions. Much of the procedure is carried out in a flow chart which may be confusing at first use. To overcome any potential problems it is best to work through the overall example (pp A-82) before using the guide to analyze a real problem.

Step I:

Graph the collected data to detect any obvious trends.

Step II:

Select a_1, the significance level for the test of hypothesis for the MOE.

Select a_2, the significance level for the test of assumptions.

Step III:

Test each sample for randomness

H_o: The sample is a random sample

H_a: The sample is not a random sample

RUNS TEST

Proportion measures: assign "+"s to one, or a group of, outcome(s); assign "-"s to all other outcomes.

Other measures: assign a "+" to each observation greater than the mean, assign a "-" to all other observations.

n_1 = no. of +'s $\quad$ n_2 = no. of -'s run: "a succession of identical symbols which are followed and preceded by different symbols or no symbol at all" (Ref 9:52)

Arrange the +'s and -'s in the order that the corresponding data was collected. Count the number of runs.

T = the number of runs

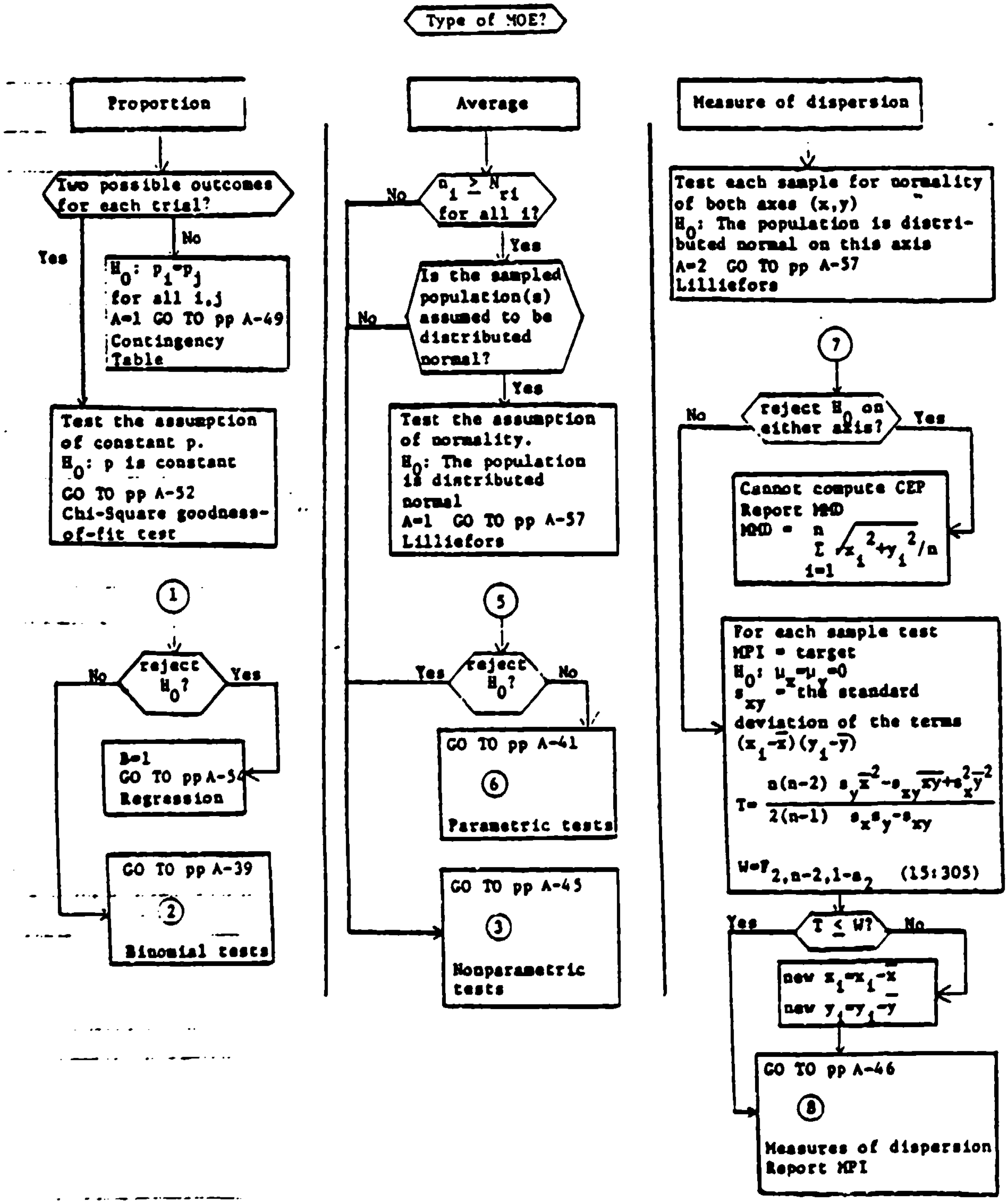
$L = U_{n_1,n_2,a}$ (Ref 15:415)

$U = U_{n_1,n_2,1-a}$

$a = a_2/2$

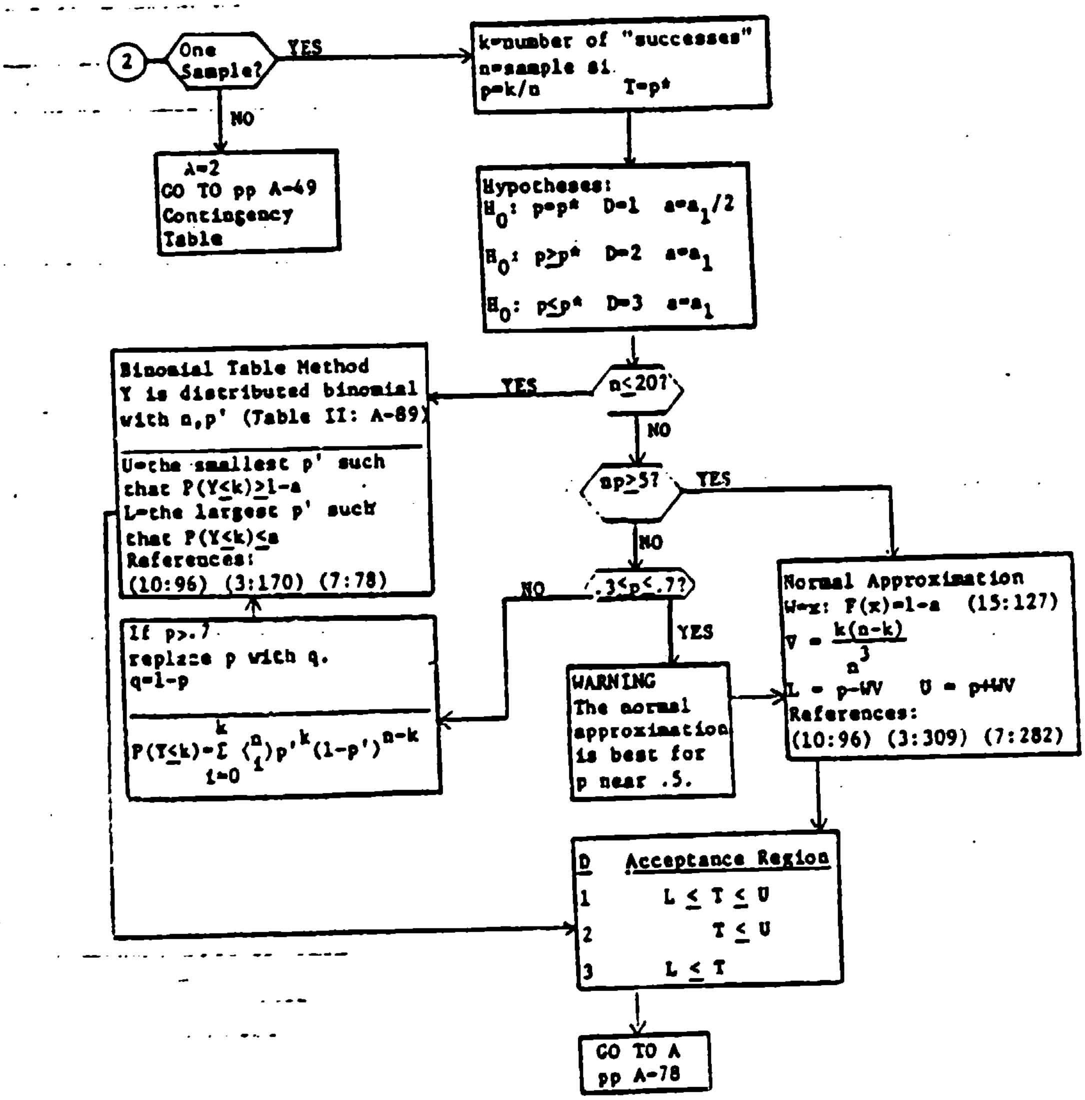
Acceptance region: $L \leq T \leq U$

If the H_0 is rejected the data should be discarded and new data should be collected. If this is not done the statistical tests are no longer valid and may lead to erroneous conclusions.



Type of MOE?
Proportion
Average
Measure of dispersion
Two possible outcomes for each trial?
Yes
No
H_0: $p_i = p_j$ for all i,j
A=1 GO TO pp A-49
Contingency Table
Test the assumption of constant p.
H_0: p is constant
GO TO pp A-52
Chi-Square goodness-of-fit test
1
reject H_0?
No
Yes
B=1
GO TO pp A-54
Regression
GO TO pp A-39
2
Binomial tests
$n_i \geq N_{ri}$ for all i?
No
Yes
Is the sampled population(s) assumed to be distributed normal?
No
Yes
Test the assumption of normality.
H_0: The population is distributed normal
A=1 GO TO pp A-57
Lilliefors
5
reject H_0?
Yes
No
GO TO pp A-41
6
Parametric tests
GO TO pp A-45
3
Nonparametric tests
Test each sample for normality of both axes (x,y)
H_0: The population is distributed normal on this axis
A=2 GO TO pp A-57
Lilliefors
7
reject H_0 on either axis?
No
Yes
Cannot compute CEP
Report MMD
$MMD = \sum_{i=1}^{n} \sqrt{x_i^2 + y_i^2}/n$
For each sample test MPI = target
H_0: $\mu_x = \mu_y = 0$
s_{xy} = the standard deviation of the terms $(x_i - \bar{x})(y_i - \bar{y})$
$T = \frac{n(n-2)}{2(n-1)} \frac{s_y \bar{x}^2 - s_{xy} \overline{xy} + s_x \bar{y}^2}{s_x s_y - s_{xy}}$
$W = F_{2,n-2,1-a_2}$ (15:305)
$T \leq W$?
Yes
No
new $x_i = x_i - \bar{x}$
new $y_i = y_i - \bar{y}$
GO TO pp A-46
8
Measures of dispersion
Report MPI

BINOMIAL TESTS



Binomial Confidence Interval

$$P(p_L \le p \le p_u) = 1-a$$

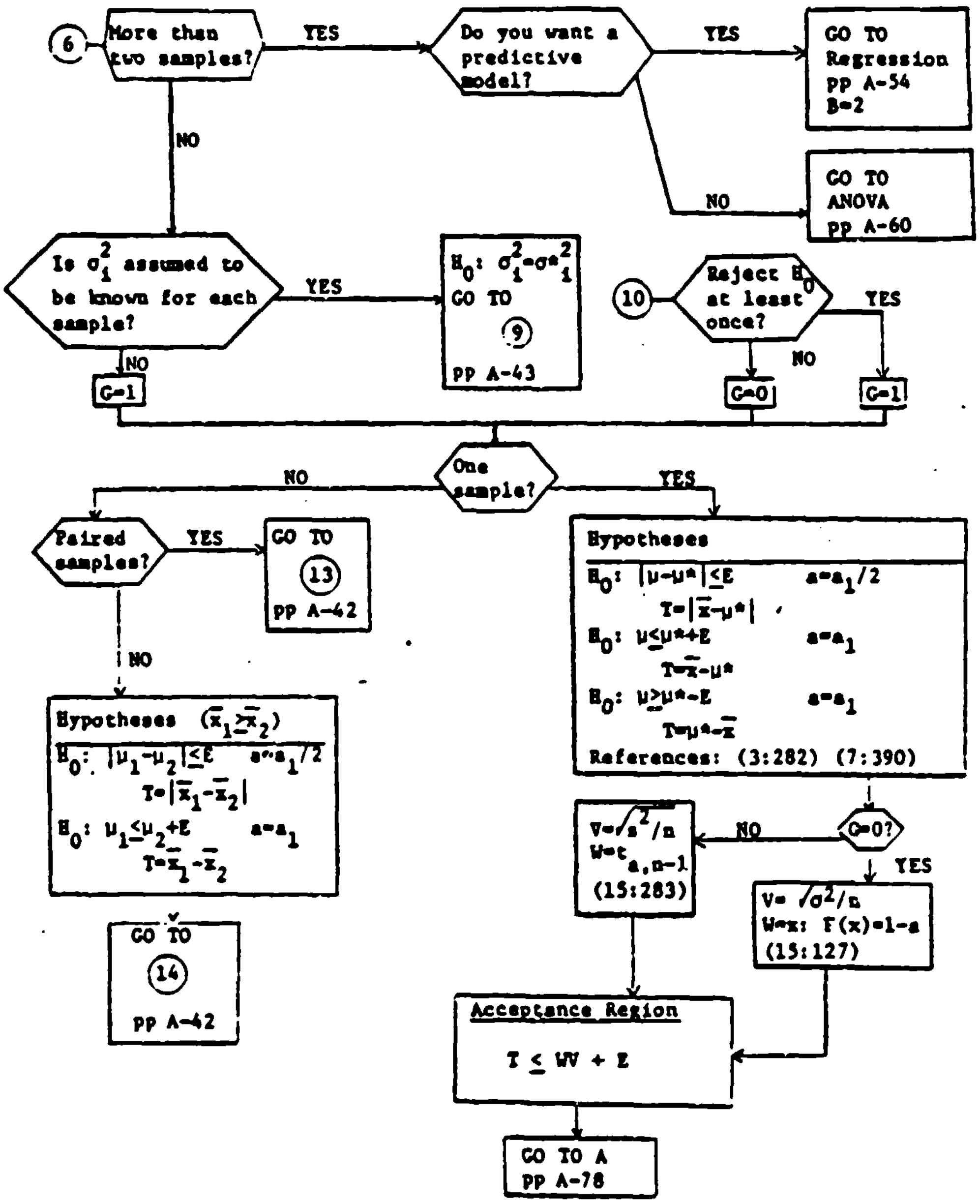
$$p_L = \frac{k}{k+(n-k+1)F_{1-a/2,2(n-k+1),2k}}$$

$$p_u = \frac{(k+1)F_{1-a/2,2(k+1),2(n-k)}}{(n-k)+(k+1)F_{1-a/2,2(k+1),2(n-k)}}$$

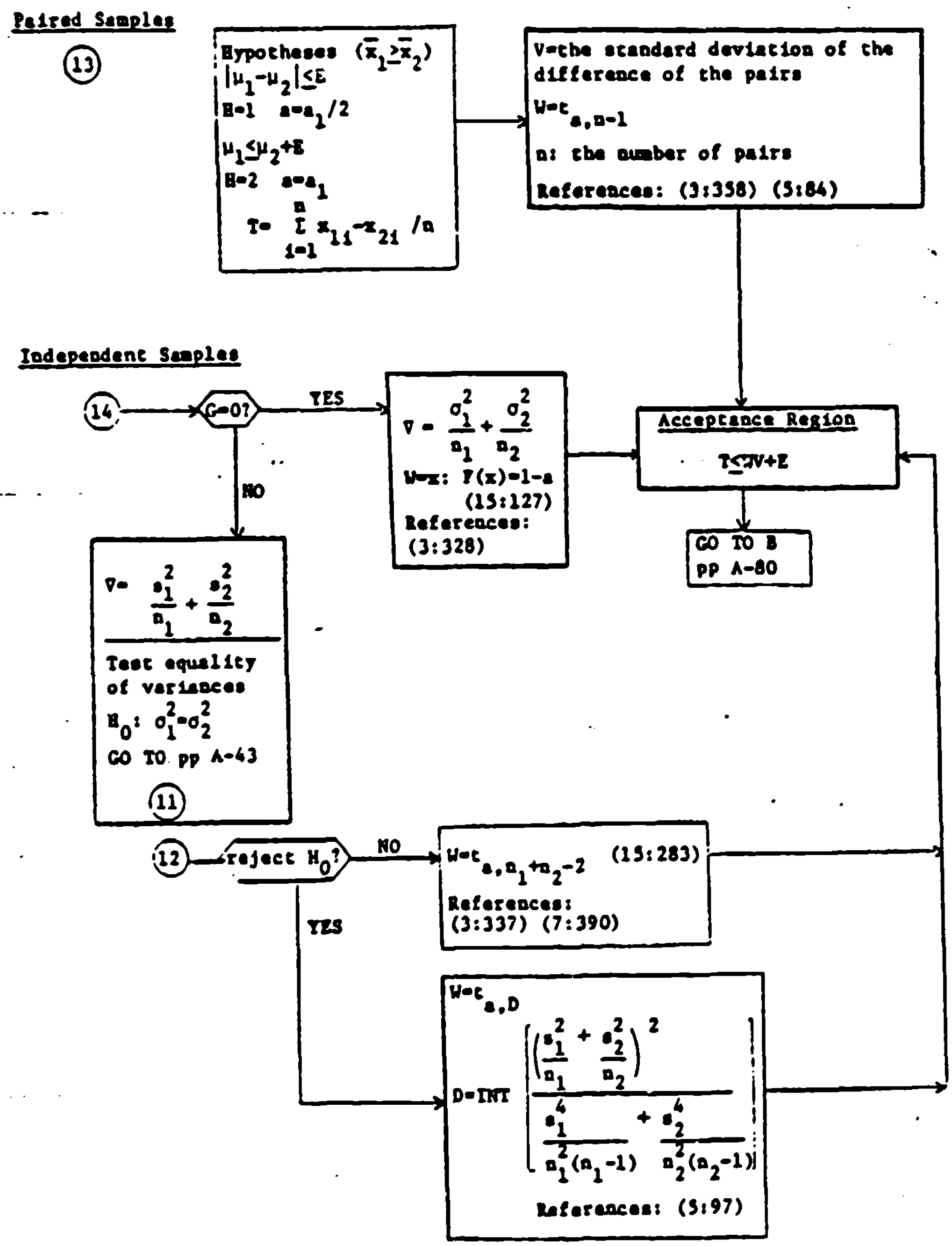
F_{a,n_1,n_2} is the tabled F value from (15:305)

Ref: (17)

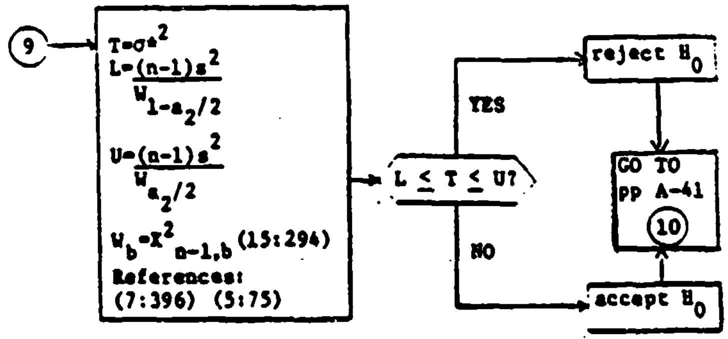
PARAMETRIC TESTS



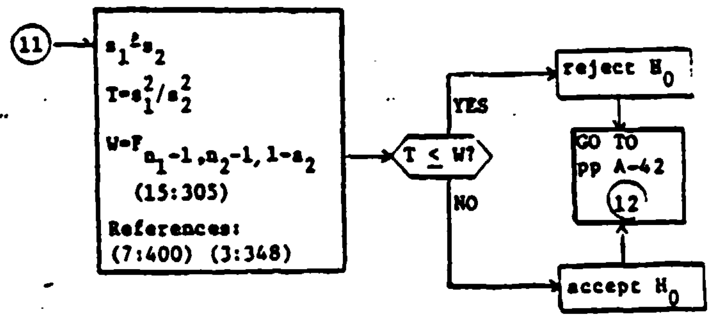
PARAMETRIC TESTS Continued



Test of a Known Variance



Test of Equality of Two Variances



Confidence Intervals

CI about a point estimate

$$P(\overline{X}-W^*V \leq \mu \leq \overline{X} + W^*V) = 1-a$$

If σ^2 is known: $W^* = X: F(X) = 1-a/2$

If σ^2 is unknown: $W^* = t_{a/2,n-1}$

CI about a difference

$$P(\overline{X}_1 - \overline{X}_2 - W^*V \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + W^*V) = 1-a$$

Paired Data

$D_i = (X_{1i}-X_{2i})$; $i = 1,2,\ldots n$

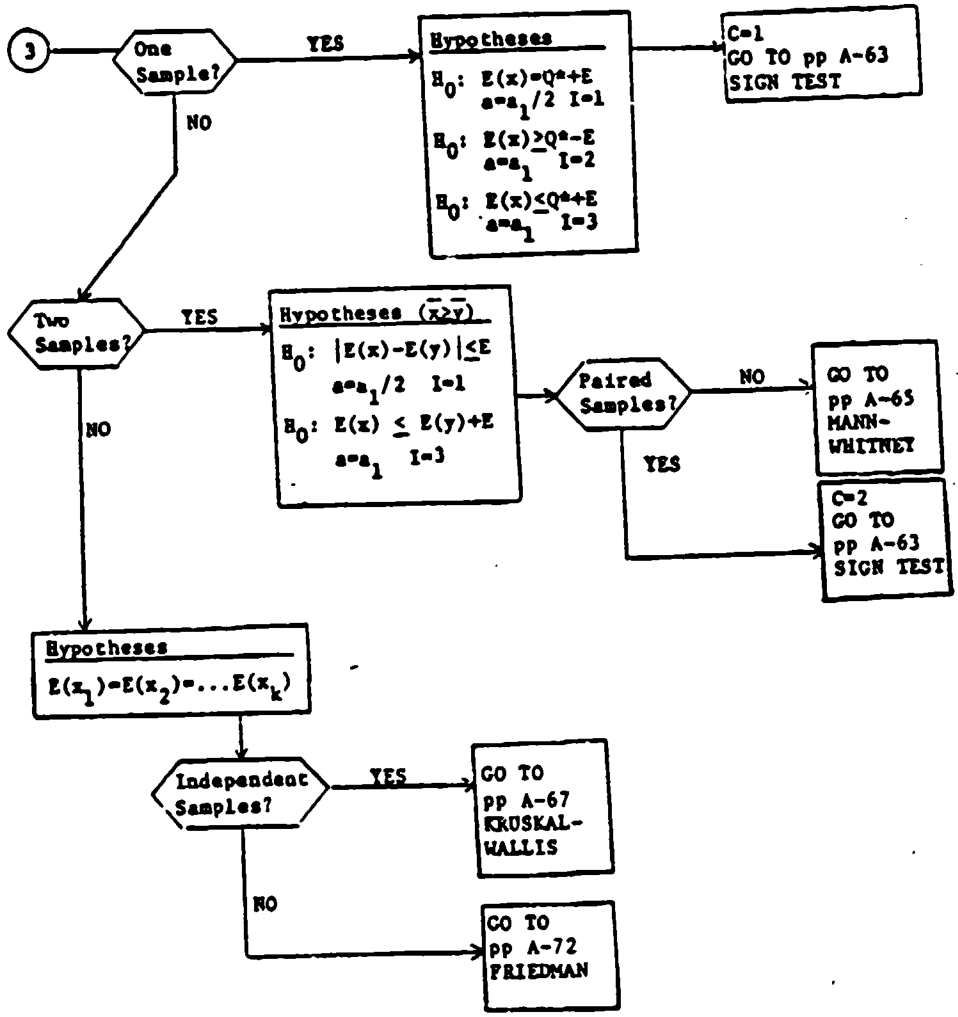
If σ_D^2 is known: $W^* = X: F(X) = 1-a/2$

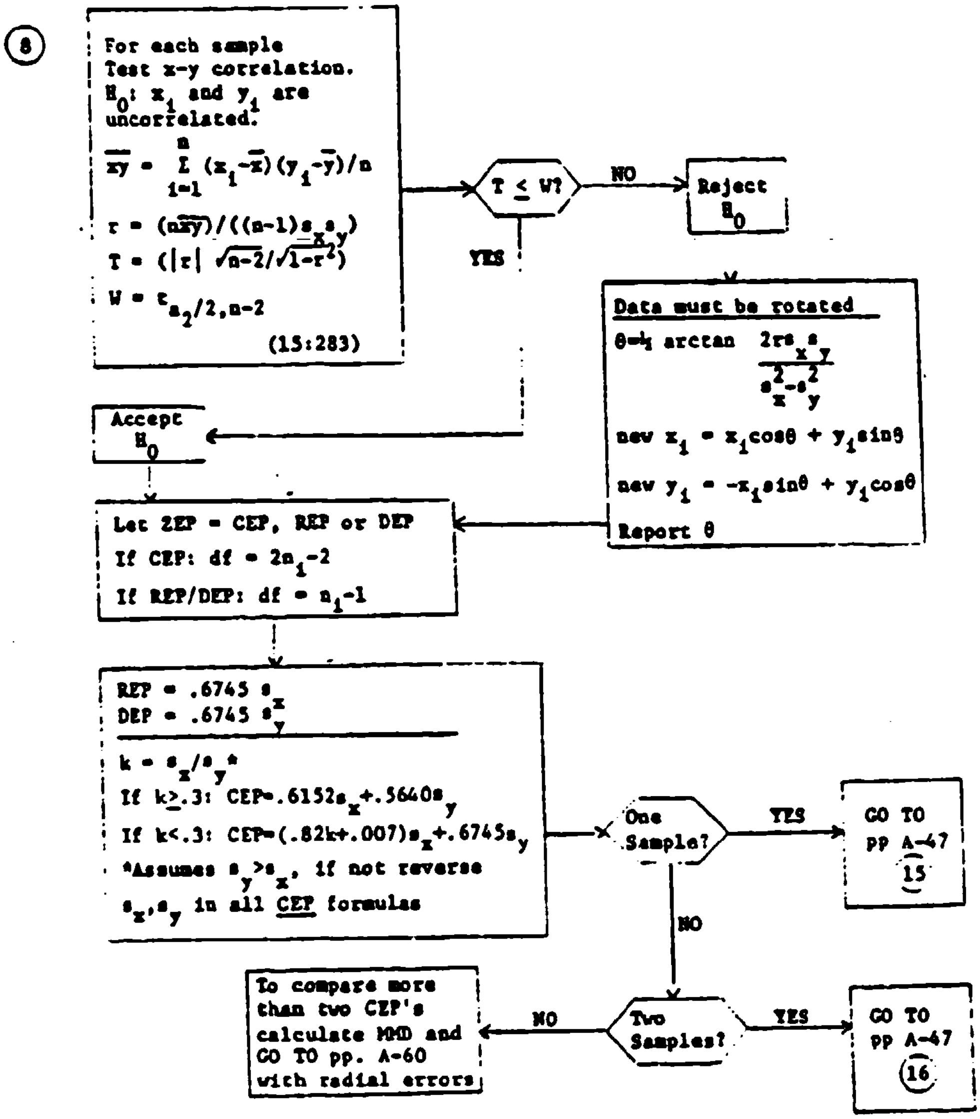
If σ_D^2 is unknown: $W^* = t_{a/2,\ n-1}$

Independent Samples

If $\sigma^2_{\overline{X}_1-\overline{X}_2}$ is known: $W^* = X: F(X) = 1-a/2$

If $\sigma^2_{\overline{X}_1-\overline{X}_2}$ is unknown: $W^* = t_{a/2,\ n_1+n_2-2}$


3
One Sample?
YES
NO
Hypotheses
H0: E(x)=Q*+E a=a1/2 I=1
H0: E(x)≥Q*-E a=a1 I=2
H0: E(x)≤Q*+E a=a1 I=3
C=1
GO TO pp A-63
SIGN TEST
Two Samples?
YES
NO
Hypotheses (x̄≥ȳ)
H0: |E(x)-E(y)|≤E a=a1/2 I=1
H0: E(x) ≤ E(y)+E a=a1 I=3
Paired Samples?
NO
YES
GO TO pp A-65 MANN-WHITNEY
C=2 GO TO pp A-63 SIGN TEST
Hypotheses
E(x1)=E(x2)=...E(xk)
Independent Samples?
YES
NO
GO TO pp A-67 KRUSKAL-WALLIS
GO TO pp A-72 FRIEDMAN


8
For each sample
Test x-y correlation.
H_0: x_i and y_i are uncorrelated.
$\overline{xy} = \sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})/n$
$r = (n\overline{xy})/((n-1)s_x s_y)$
$T = (|r| \sqrt{n-2}/\sqrt{1-r^2})$
$W = t_{a_2/2,n-2}$
(15:283)
$T \leq W$?
NO
Reject H_0
YES
Accept H_0
Data must be rotated
$\theta = \frac{1}{2}\arctan \frac{2rs_x s_y}{s_x^2 - s_y^2}$
new $x_i = x_i\cos\theta + y_i\sin\theta$
new $y_i = -x_i\sin\theta + y_i\cos\theta$
Report θ
Let 2EP = CEP, REP or DEP
If CEP: df = $2n_i-2$
If REP/DEP: df = n_i-1
REP = .6745 s_x
DEP = .6745 s_y
$k = s_x/s_y$*
If $k \geq .3$: CEP=.6152s_x+.5640s_y
If $k < .3$: CEP=(.82k+.007)s_x+.6745s_y
*Assumes $s_y > s_x$, if not reverse s_x, s_y in all CEP formulas
One Sample?
YES
GO TO pp A-47 15
NO
Two Samples?
YES
GO TO pp A-47 16
NO
To compare more than two CEP's calculate MMD and GO TO pp. A-60 with radial errors

15 → **Hypotheses**

H_0: $ZEP = Q^*$
$R=1$ $a=a_1/2$

H_0: $ZEP \geq Q^*$
$R=2$ $a=a_1$

H_0: $ZEP \leq Q^*$
$R=3$ $a=a_1$

→

$T = (Q^*)^2$

$U = ZEP^2 \dfrac{df}{W_a}$

$L = ZEP^2 \dfrac{df}{W_{1-a}}$

$W_b = \chi^2_{b,df}$ (15:294)

↓

R	Acceptance Region
1	$L \leq T \leq U$
2	$L \leq T$
3	$T \leq U$

→ GO TO A pp A-78

16 → estimate of $ZEP_1 \geq$ estimate of ZEP_2

H_0: $ZEP_1 = ZEP_2$
$a=a_1$

↓

TESTING CEP?

NO →

$T = \dfrac{ZEP_1^2}{ZEP_2^2}$

$W = F_{n_1-1,n_2-1,1-a_1}$

(15:305)

→ GO TO B pp A-80

YES ↓

$k \leq .3$?

YES → GO TO pp A-76 Siegel-Tukey

NO →

$T = \dfrac{n_1(CEP_1)^2(n_2-1)}{n_2(CEP_2)^2(n_1-1)}$

$W = F_{2n_1-2,2n_2-2,1-a_1}$

(15:305)

→ GO TO B pp A-80

Upper Confidence Limit for CEP

$$P\left(CEP \leq (\text{estimated CEP}) \frac{(1+k^2)(n-1)}{W_a}\right) = a$$

$W = X^2_{df,a}$

a: the desired confidence level

$df = INT\left[(1+k^2)(n-1)\right]$

Reference (28:6)

Contingency Table

The contingency table is used to test hypotheses concerning the probability of the occurrence of particular outcomes between populations. This is of special use in the multinomial experiment where the hypotheses is that the probability of outcome j is equal among all populations for all outcomes. The test is generalized for "r" populations and "c" outcomes. For further discussion see (Ref 10:153-158).

	outcome 1	outcome 2 ...	outcome c	total
population 1	0_{11}	0_{12}	0_{1c}	n_1
population 2	0_{21}	0_{22}	0_{2c}	n_2
⋮				
population r	0_{r1}	0_{r2}	0_{rc}	n_r
total	t_1	t_2	t_c	N

0_{ij} = the number of times outcome j is observed in the ith population.

t_j = the column total of 0_{ij}'s for all i

n_i = the row total of 0_{ij}, for all j

If A=1: C= the number of possible outcomes or groups of outcomes, r= the number of samples to be compared.

If A=2: C=2 (Binomial), r = the number of samples to be compared.

Hypotheses

H_o: $P_{1j} = P_{2j} = \ldots P_{rj}$ for all j

The probability of outcome j is equal across all populations for all outcomes.

H_A: At least two of the populations (P_{1j}, P_{2j} ... P_{rj}) are not equal within an outcome.

Test

The test statistic, T, is calculated as:

$$T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij})^2 N}{c_i n_i} - N$$

$$W = \chi^2_{F,n} \qquad \text{(Ref 15:294)}$$

$$F = 1-a_1$$

$$n = (r-1)(c-1)$$

Acceptance Region

$T \leq W$

GO TO Reporting Results - pp. A-78

Example:

Determine whether or not the probability of kill (P_k) is affected by missile type (AIM-9J, AIM-9L, AIM-9M) at $a_1 = .05$.

Hypothesis:

H_o: $P_k(9J) = P_k(9L) = P_k(9M)$

H_A: At least two of the missiles have different P_k's

Data:

Outcome of Trial

Missile Type	Kill	No-Kill	Total
9J	10	8	18
9L	13	4	17
9M	14	4	18
Total	37	16	53

Test:

$$T = \frac{10^2 \cdot 53}{37 \cdot 18} + \frac{13^2 \cdot 53}{37 \cdot 17} + \frac{14^2 \cdot 53}{37 \cdot 18} + \frac{8^2 \cdot 53}{16 \cdot 18} + \frac{4^2 \cdot 53}{16 \cdot 17} + \frac{4^2 \cdot 53}{16 \cdot 18} - 53$$

$T = 2.635$

$W = X^2_{F,n}$ (Ref 15:294)

$F = 1 - .05 = .95$

$n = (3-1)(2-1) = 2$

$W = 5.991$

$T \leq W$ therefore we fail to reject H_o at $a_1 = .05$.

There is no statistical evidence to support a claim that missile type affects P_k at $a_1 = .05$.

Chi-Square Goodness-of-Fit Test

The Chi-Square goodness-of-fit test is used to determine if a binomial experiment has a constant probability of success across time. This test is similar to the contingency table however there is only one row. Each cell should contain five samples or more. The data must be ordered by time and divided into cells. The observed frequency (0_i) is equal to the number of successes in cell i. The expected frequency (E_i) is equal to $\hat{p}$, the total successes divided by total observations, times the number in cell i.

n_i = number in cell i

0_i = successes observed in cell i

$$E_i = \frac{k}{n} \cdot n_i$$

k = total observed successes

n = total sample size

m = no. of cells

Hypothesis

H_o: p is constant across time, i.e. $p_1=p_2=\ldots p_m$

H_A: p is not constant across time

Test

$$T = \sum_{i=1}^{m} (0_i-E_i)^2/E_i$$

$$W = X^2_{F,n} \qquad \text{(Ref 15:294)}$$

$$F = 1-a_2$$

$$n = m-1$$

Acceptance Region

$T \leq W$

Return to 1 pp. A-38

Linear Regression

Linear regression is very similar to ANOVA, indeed ANOVA is a special case of regression. Linear regression attempts to determine the linear relationship between the dependent variable (MOE) and any number of independent variables which affect its value. The technique estimates the coefficients (B_i) of the linear equation ($y=B_o + B_1X_1 + B_2X_2$) which best fits the data. The "best" line is defined as the one for which the sum of the distances squared (SSE) from the observed values of the dependent variable (y) to the estimated line ($\hat{y}$) is minimized. The measure of worth of the model is R^2 which is a function of how much the SSE for the estimated line ($\hat{y}$) differs from the SSE about the line ($\bar{y}$). An R^2 close to zero indicates little additional explanation of the dependent variable by the independent variables. It should be noted that R^2 will always increase when variables are added and that a better measure of explanatory power is the adjusted R^2 ($\bar{R}^2$) which is adjusted for the number of independent variables in the model. Linear regression routines are available on computerized statistics packages and are necessary to complete the analysis. Regression can be accomplished using ordinal or nominal variables; however, this requires modifications to the basic model. For further references for applying regression see (Ref 21:21-272), and (Ref 20:34-208).

Example

Objective: Determine the relationship between radar detection range and target altitude.

MOE: radar detection range.

Model:

$$y = B_0 + B_1 x_1 + e$$

y: radar detection range

x_1: target altitude (in 100 x feet)

e: random or unexplained error

Hypothesis:

$$H_0: B_0 = \bar{y} + e$$

$$B_1 = \emptyset$$

$$H_A: B_1 \neq \emptyset$$

Test:

T = overall F - from printout

$W = F_{m,n,1-a_1}$ (Ref 15:305)

m = regression degrees of freedom - from printout

n = error degrees of freedom - from printout

Acceptance Region:

$T \leq W$

If B=1: y=p x_1=time

If B=2: y=MOE x_i=independent variables

The regression technique is capable of determining the relationship with more than one independent variable, however, the graphical relationship is not easy to show. Conclusions made from the regression results should be restricted to the range of values which are within the limits of the data collected. Along with the relationship between the variables, the explanatory power of the model ($\bar{R}^2$) should be reported and can be interpreted as the percent of variance in the dependent variable explained by the model.

BMD: programs - P1R
P2R

(Ref 16:375-417)

SPSS: program - regression

(Ref 22:370-367)

Lilliefors Test

The Lilliefors test is used to determine if a sample could have come from a normal distribution. Initially the sample must be ordered, the observed cumulative distribution at X_i is equal to the number of data points less than or equal to X_i divided by n, the sample size. The expected cumulative frequency is found in standard normal tables after standardizing the data. The test statistic, T, is equal to the maximum difference, for any x_i, between F_{obs} and F_{exp}. The test statistic, T, is compared to the Lilliefors test statistic, (Ref 10:357).

Hypothesis:

H_o: The sample is distributed normal

H_A: The sample is not distributed normal

Test:

$$F(x_i)_{obs} = \frac{\text{number of } x_j \le x_i}{n}$$

$$F(x_i)_{exp} = F\ \frac{x_i - \mu}{T^{\sigma}}$$ (Ref 15:127)

$$T = \max \left| F(x_i)_{obs} - F(x_i)_{exp} \right|$$

$$W = T_{n,p}$$ (Table III, pp A-101)

n = sample size

$p = 1\text{-}a_2$

Acceptance Region:

$T \leq W$: A=1 Return to 5 pp. A-38

A=2 Return to 7 pp. A-38

Example:

An analysis of component reliability uses the mean time between failures (MTBF) as the MOE. Test the assumptions of normality.

Data:

Failure	2	3	4	5	6	7	8	9	10	11
Time	41.2	88.5	134.7	179.8	224.7	271.9	321.7	358.4	415.6	469.8
TBF	41.2	47.3	46.2	45.1	44.9	47.2	49.8	36.7	57.2	54.2

y:TBF $\bar{y}=47.0$ $s^2=34.7$ $s=5.9$ $n=10$ $a_2=.1$

Test:

H_o:TBF is distributed normal with $\mu=47.0$ and $\sigma^2=34.7$.

Ordered Observations	36.7	41.2	44.9	45.1	46.2	47.2	47.3	49.8	54.2	57.2
Observed Cumulative Frequency	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
F(y-47.0/5.9) (Ref 15:127)	.040	.163	.360	.373	.446	.514	.524	.683	.890	.958
T=|observed-F|	.060	.037	.060	.027	.054	.086	.179	.117	.010	.042

$T = .179$

$W = T_{n,p}$ (Ref 10:463)

$n = 10$

$p = 1 - .1 = .9$

$W = .239$

$T \leq W$ fail to reject H_o at a_2=.1, therefore there is no evidence which supports the claim that the data is not normally distributed at a_2=.1.

ANOVA

Analysis of variance will be most useful to compare more than two means simultaneously. The samples which are tested can be defined as levels of factors or treatments, and ANOVA can determine which factors and/or levels represent significant effects on the overall mean. Experimental design is of great importance, if ANOVA is to be employed effectively. For a complete discussion of ANOVA and its design see (Ref 18) and (Ref 5:194-333).

Example:

Objective: Determine the main determinants of radar detection.

MOE: radar detection range

Question: Is radar detection affected either positively or negatively by aircraft altitude or radar pulse repetition frequency (PRF)

Model:

$$y = \mu + A + P + e$$

y: radar detection range

μ: the overall mean detection range

A: the affect of altitude on y

P: the affect of PRF on y

e: random or unexplained error

Hypothesis: There is no affect due to altitude or PRF on radar detection range.

Test:

$T = F$ - from printout

$W = F_{m,n,1-a_1}$ (Ref 15:305)

m = degrees of freedom for factors

n = degrees of freedom for error

Acceptance Region

$T \leq W$

This ANOVA model, known as two-way, can be used to detect the effect of different levels, e.g. high or low, of each factor or the effect of interaction between the factors, altitude and PRF. Interaction between factors is the effect of combinations of factors not already explained by the factors alone (e.g. AP, the interaction between altitude and PRF). Interaction terms should not be included in the model unless the interaction can be interpreted. In general interaction beyond two terms is not reported. To carry out any ANOVA requires at least one observation in each cell, however, it should be noted that interaction cannot be tested without replications within cells.

The ANOVA technique is accomplished by comparisons between row and column means and the overall mean. Simple ANOVA can be accomplished with the help of a hand calculator. However, most statistical software packages have ANOVA routines which will make the job easier. It should be noted that these computerized routines can often only be used to analyze fixed effect models without some manipulation of the output. A factor is fixed if the levels which are specified are the only levels of interest and no interpolations or extrapolations will be made. Consult the references already listed for the procedure to be applied to random or mixed effect models.

BMD Programs - P1V
P2V
P3V

(Ref 16:523-602)

SPSS Program - ANOVA

(Ref 22:398-433)

Go To Reporting Results - C. pp. A-81

Sign Test

The sign test is a nonparametric test which can be used to test one average against a hypothesized value or to compare two averages.

Hypothesis: Bring hypothesis (I) from flow chart

Test:

C=1 (one sample)

I=1, I=3

T=no. of observations, x_i, $> Q^*+E$

I=2

T=no. of observations, x_i, $> Q^*-E$

C=2 (two samples)

T=no. of $x_i > y_i + E$

$n \leq 20$

$U=n-W$

$L=W$

$W=S_{n,a}$ (Ref 15:398)

$n > 20$

$U=n/2 + W(\sqrt{n/4})$

$L=n/2 - W(\sqrt{n/4})$

$W=X$: $F(X)=1-a$ (Ref 15:127)

Acceptance Region:

I=1	I=2	I=3
$L \leq T \leq U$	$L \leq T$	$T \leq U$

Example: Determine whether a maintenance task is easier to accomplish using process B than process A. Ten maintenance personnel are asked to perform the task using each process and then to state their preference. Assign a "+" to those preferring process A and a "-" to those who prefer process

S, a_1=.05.

Hypothesis:

H_o: P(+) $\geq$ P(-)

i.e. process A is preferred to process B.

H_A: P(+) < P(-)

Data:

Three people preferred process A.

Seven people preferred process B.

n=10 $a=2a_1/2=.1$

Test: (two sample) (I=2)

T = 3

L = 1 (Ref 15:398)

L $\leq$ T fail to reject H_o at a_1 = .05

There is no evidence to support a claim

that process B is preferred to process A.

Go To Reporting Results - pp. A-78

Mann-Whitney Test

The Mann-Whitney test is a nonparametric test for comparing the average values of two independent samples (Ref 10:215-223).

Sample 1	Sample 2
x_{11}	x_{21}
x_{12}	x_{22}
$\vdots$	$\vdots$
x_{1n}	x_{2m}

$N=n+m$

$R(x_{ij})$: the rank, from 1 to N, assigned to each observation in the combined samples. A rank of 1 is assigned to the smallest observed values and N to the largest, tied observations should each be given the average rank.

Hypothesis: bring hypothesis (I) from flow chart

Test:

$$T = \sum_{i=1}^{n} R(x_{1i})$$

$$U = n(N+1)-W$$

$$L = W$$

$$W = U_{m,n,a} \qquad \text{(Ref 15:406)}$$

Acceptance Region:

I=1	I=3
$L \leq T \leq U$	$L \leq T$

Example:

Determine whether the average detection range (nm) of radar A is different from that of radar B, a_1=.1.

Hypothesis:

H_o: $|E(A)-E(B)| \leq 0$, i.e. there is no difference in the detection ranges of radar A and radar B.

Data: (ordered)

Radar	B	B	A	B	B	A	A	B	A	B	A	A	B	A	B	A
Detection range	37	38	38	39	41	42	42	42	43	44	45	46	47	49	49	51
Rank	1	2.5	2.5	4	5	7	7	7	9	10	11	12	13	14.5	14.5	16

m=8 n=8 N=16

Tests:

T=2.5 + 7 + 7 + 9 + 11 + 12 + 14.5 + 16 = 79

$W = U_{m,n,a}$ (Ref 15:406)

W = 19

U = 8(16+1) - 19 = 117

L = 19

$L \leq T \leq U$ therefore fail to reject H_o at a_1=.1, there is no evidence to support a claim that the detection range of radar A is significantly different from radar B at a_1=

Go To Reporting Results - B. pp. A-80

Kruskall-Wallis Test

The Kruskall-Wallis test is an extension of the Mann-Whitney test for two samples to more than two samples (Ref 10:229). The test is used to test the hypothesis that k independent populations are identical against the alternative that at least one population tends to yield larger values than at least one other population. The samples are assumed to be independent; that is, the test was carried out so that outcomes in separate samples are unrelated. If this is not the case the Friedman Test for related samples should be used (pp. A-72).

Sample 1	Sample 2	...	Sample k
x_{11}	x_{21}		x_{k1}
x_{12}	x_{22}		x_{k2}
$\vdots$	$\vdots$		$\vdots$
x_{1n_1}	x_{2n_2}		x_{kn_k}

$$N = \sum_{i=1}^{k} n_i$$

$R(x_{ij})$: the rank from 1 to N assigned to the combined sample observations. A rank of 1 is assigned to the smallest observation and N to the largest, tied observations should each receive the average rank.

$$R_i = \sum_{j=1}^{n_i} R(x_{ij})$$

Hypothesis:

H_o: $E(x_1) = E(x_2) = \ldots E(x_k)$

The k populations are identically distributed.

H_A: At least one of the populations tends to yield larger observations than at least one of the others.

Test:

The test statistic T is calculated as:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} R(x_{ij})^2 - \frac{N(N+1)^2}{4}$$

$$T = \frac{1}{S^2} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}$$

$W = X^2_{n\ ,F}$ (Ref 15:294)

$n = k-1$

$F = 1-a_1$

Acceptance Region:

$T \leq W$

Go To Reporting Results - C. pp. A-81

Pairwise Comparison:

If, and only if, the H_o is rejected pairwise comparisons can be made to detect individual differences between two samples.

$$T_1 = \left|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right|$$

$$W_1 = (t_{n,F})\ \left(S^2\ \frac{N-1-T}{N-k}\right)^{\frac{1}{2}}\ \left(\frac{1}{n_i} + \frac{1}{n_j}\right)^{\frac{1}{2}}$$

$t_{n,F}$ from (Ref 15:283)

$n = N-k$

$F = 1 - a_1/2$

A significant difference exists if $T_1 > W_1$

Example:

Determine if there is a significant difference in the results of air combat using three different tactics, the MOE is average exchange ratio, (ER=blue force "killed"/red force "killed"), a_1=.1.

Hypothesis:

H_o: The expected outcome of tactic A (E(A)) is equal to the expected outcome of tactics B and C, i.e. E(A) = E(B) = E(C)

Data:

	Tactic A		Tactic B		Tactic C	
	ER	Rank	ER	Rank	ER	Rank
	.20	1	.33	5.5	.40	8.5
	.25	2	.33	5.5	.50	12
	.30	3	.40	8.5	.60	15.5
	.33	5.5	.50	12	.60	15.5
	.33	5.5	.50	12	.67	17
	.50	12	.50	12	.75	18
R_i		29		55.5		86.5
n_i		6		6		6

$N = 18$

$$S^2 = \frac{1}{18-1} \left(1^2+2^2+3^2+4(5.5^2)+2(8.5^2)+5(12^2)+2(15.5^2)+17^2+18^2\right) - \frac{18(19)^2}{4}$$

$$= 27.559$$

$$T = \frac{1}{27.559} \left(\frac{29^2}{6} + \frac{55.5^2}{6} + \frac{86.5^2}{6}\right) - \frac{18(19)^2}{4}$$

$$= 10.01$$

$$W = X^2_{k-1,1-a_1}$$

$k = 3$

$W = 4.61$

$T > W$, therefore reject H_o at $a_1=.1$, there is evidence that exchange rate changes with tactics at $a_1=.1$. There is a 10% (a_1) chance the difference detected is due to random chance.

Determine which tactics of the three differ.

<u>A vs. B</u>

$$T_1 = \left|\frac{29}{6} - \frac{55.5}{6}\right| = 4.417$$

$$W_1 = 1.753 \left(27.559 \frac{(18-1-10.01)}{18-3}\right)^{1/2} \left(\frac{1}{6} + \frac{1}{6}\right)^{1/2}$$

$$= 3.627$$

$T_1 > W_1$ therefore tactic A and B are significantly different.

<u>B vs. C</u>

$$T_1 = \left|\frac{55.5}{6} - \frac{86.5}{6}\right| = 5.167$$

$W_1 = 3.627$

$T_1 > N_1$ therefore tactic B and C are significantly different.

<u>A vs. C</u>

$$T_1 = \left|\frac{29}{6} - \frac{86.5}{6}\right| = 9.583$$

$W_1 = 3.627$

$T_1 > W_1$ therefore tactic A and C are significantly different.

Friedman Test

The Friedman test, like the Kruskall-Wallis test, is used to determine which treatments, if any, in an experiment cause significantly different measured values of the response variable or MOE. The Friedman test is used for related samples, that is, samples which lack independence. The samples are considered to be related when the test is carried out in such a way that more than one sample is measured at the same time, or some other factor causes the relationship (i.e. same pilot) that is some uncontrolled factor may influence measurements between samples (Ref 10:299-308).

		Treatment			
		1	2	...	k
Block	1	x_{11}	x_{12}		x_{1k}
	2	x_{21}	x_{22}		x_{2k}
	.				
	.				
	.				
	b	x_{b1}	x_{b2}		x_{bk}

Treatment: The factor which is to be tested, with k levels.

Block: The environmental conditions which are controlled in the experiment, (e.g. aircraft tail number, trial).

$R(x_{ij})$; The rank from 1 to k which is assigned to each x_{ij} within a block, a rank of 1 is assigned to the smallest observation and k to the largest. In case of a tie assign the average rank to both observations.

$$R_j = \sum_{i=1}^{b} R(x_{ij})$$

The blocks are ranked from 1 to b according to the range of values within a block (maximum x_{ij} - minimum x_{ij}). Rank 1 is assigned to the block with the smallest range.

Q_i = the rank assigned to block i.

$$S_{ij} = Q_i \left(R(x_{ij}) - \frac{k+1}{2}\right)$$

$$S_j = \sum_{i=1}^{b} S_{ij}$$

k: The number of treatments

b: The number of blocks

<u>Hypothesis</u>:

H_o: The treatments have equal effects on the response variable or MOE.

H_A: At least one of the treatments tends to yield larger observations than at least one other treatment.

<u>Test</u>:

$$A = \sum_{i=1}^{b} \sum_{j=1}^{k} R_{ij}^2 \qquad B = \frac{1}{b} \sum_{j=1}^{k} R_j^2$$

$$T = \frac{(b-1)B}{A-B}$$

$W = F_{m,n,1-a_1}$ (Ref 15:305)

$m = b-1$

$n = k-1$

Acceptance Region:

$T \leq W$

Go To Reporting Results - C. pp. A-81

Example:

Determine which radar modification results in the longest radar detection range; all trials involve three aircraft simultaneously, one with each modification, $a_1=.1$.

Hypothesis:

H_o: Modification A, B, and C are equally effective in terms of their effect on detection range.

Data:

Modification

Trial	A	rank	B	rank	C	rank
1	23	3	22	1.5	22	1.5
2	27	2	28	3	26	1
3	24	1.5	24	1.5	27	3
4	22	1	24	2	26	3
5	23	1	25	2	26	3
6	21	1	22	2	24	3
7	29	1	30	2	33	3
8	28	2	27	1	31	3
		12.5		15		20.5

Test:

$$A = 8(3^2) + 6(2^2) + 4(1.5^2) + 6(1^2) = 111$$

$$B = \frac{1}{8}(12.5^2 + 15^2 + 20.5^2) = 100.188$$

$$T = \frac{(8-1)(100.188)}{(111-100.188)} = 64.86$$

$$W = F_{m,n,1-a_1}$$

$$m = b-1 = 7$$

$$n = k-1 = 2$$

$W = 9.35$ (Ref 15:305)

$T > W$ therefore reject H_o at $a_1 = .1$, there is evidence that the modifications to the radar do not result in equally effective radar performance.

Siegel-Tukey Test

The Siegel-Tukey test is performed in a manner very similar to the Mann-Whitney test and uses identical statistical tables. The test compares the equality, or inequality, of the variances from two populations. When using this test the sample variance should be computed using the radial miss distance (RMD) from the MPI (mean point of impact), ($RMD_i = \sqrt{x_i^2 + y_i^2}$)

Sample 1	Sample 2
x_{11}	x_{21}
x_{12}	x_{22}
$\vdots$	$\vdots$
x_{1n}	x_{2m}

$N = n+m$

The two samples should be combined and ordered as for the Mann-Whitney test, however, the ranks, $R(x_{ij})$, should be assigned so that rank 1 is given to the smallest value, rank 2 to the largest, rank 3 to the second largest, rank 4 to the second smallest, rank 5 to the third smallest, and so on. In case of a tie, assign the average rank to each observation.

Hypothesis:

1) H_o: $\sigma_1^2 = \sigma_2^2$ two-tail test $a=a_1/2$

H_A: $\sigma_1^2 \neq \sigma_2^2$

2) H_o: $\sigma_1^2 \geq \sigma_2^2$ one-tail test $a=a_1$

H_A: $\sigma_1^2 < \sigma_2^2$

Test:

$$T = \sum_{j=1}^{n} R(x_{1j})$$

$$U = n(N+1) - W$$

$$L = W$$

$$W = U_{m,n,a} \qquad \text{(Ref 15:406)}$$

Acceptance Region:

Hypothesis

1	$L \leq T \leq U$
2	$L \leq T$

Go To Reporting Results - pp. A-78

Example:

Determine whether delivery accuracy of iron bombs (MK-82) using fire control computer mode A is better than the accuracy when using mode B, a_1=.1.

Hypothesis:

H_o: The variance around the MPI using mode A is greater or equal to the variance using mode B. $(\sigma_A^2 \geq \sigma_B^2)$

Data:

Mode	A A B A A B A B A B A B B B
Radial miss distance (ft.)	7 8 11 13 14 15 15 22 24 27 28 28 35 39
Rank	1 4 5 8 9 12.5 12.5 14 11 10 6.5 6.5 3 2

$T = (1 + 4 + 8 + 9 + 12.5 + 11 + 6.5)$

$= 52$

$W = U_{m,n,a}$ (Ref 15:406)

$m = 7$

$n = 7$

$a = .1$

$W = 13$

$L = 13$

$L \leq T$ therefore fail to reject H_o at a_1=.1, there is no evidence to support the claim that the variance of mode A is less than the variance of mode B.

Reporting Results

A. One Sample Tests - Fill in parenthesis for the correct combination of H_o and test result.

H_o:

$p=p^*$

$|\mu-\mu^*| \leq E$

$|E(x) - Q^*| \leq E$

$ZEP = Q^*$

Reject H_o: "Reject H_o at $((1-a_1)x100)\%$ confidence level. There is evidence that (MOE) is significantly different from (hypothesized value, e.g. Q^*). The probability that the difference detected is due purely to randomness is equal to (a_1)."

Fail to Reject H_o: "Fail to reject H_o at $((1-a_1)x$ $100)\%$ confidence level. There is insufficient evidence to indicate that (MOE) is significantly different from (hypothesized value)."

H_o:

$$P \geq P^*$$

$$\mu \geq \mu^* - E$$

$$E(x) \geq Q^* - E$$

$$ZEP \geq Q^*$$

Reject H_o: "Reject H_o at $((1-a_1)x100)\%$ confidence level. There is evidence that (MOE) is significantly less than (hypothesized value). The probability that the difference detected is due purely to randomness is equal to (a_1)."

Fail to Reject H_o: "Fail to reject H_o at $((1-a_1)x$ $100)\%$ confidence level. There is insufficient evidence to indicate that (MOE) is significantly less than (hypothesized value)."

$\underline{H_o}$:

$p \leq p^*$

$\mu \leq \mu^* + E$

$E(x) \leq Q^* + E$

$ZEP \leq Q^*$

$\underline{\text{Reject } H_o}$: "Reject H_o at $((1-a_1)x100)\%$ confidence level. There is evidence that (MOE is significantly greater than (hypothesized value). The probability that the difference detected is due purely to randomness is equal to (a_1)."

$\underline{\text{Fail to reject } H_o}$: "Fail to reject H_o at $((1-a_1)x100)\%$ confidence level. There is insufficient evidence that (MOE) is significantly greater than (hypothesized value)."

B. <u>Two Samples</u> - Fill in parenthesis for correct combination of H_o and test results.

$\underline{H_o}$:

$|\mu_1 - \mu_2| \leq E$

$|E(x) - E(y)| \leq E$

$ZEP_1 = ZEP_2$

$p_1 = p_2$

$\underline{\text{Reject } H_o}$: "Reject H_o at $((1-a_1)x100)\%$ confidence level. There is evidence that (MOE_1) is significantly different from (MOE_2). The probability that the difference detected is due purely to randomness is equal to (a_1)."

<u>Fail to reject H_o:</u> "Fail to reject H_o at (($(1-a_1)$x100)% confidence level. There is insufficient evidence to indicate that (MOE_1) is significantly different from (MOE_2)."

<u>H_o:</u>

$$\mu_1 \leq \mu_2 + E$$

$$E(x) \leq E(y) + E$$

<u>Reject H_o:</u> "Reject H_o at (($(1-a_1)$x100)% confidence level. There is evidence that (MOE_1) is greater than (MOE_2). The probability that the difference detected is due purely to randomness is equal to (a_1)."

<u>Fail to reject H_o:</u> "Fail to reject H_o at (($(1-a_1)$x100)% confidence level. There is insufficient evidence to indicate that (MOE_1) is significantly greater than (MOE_2)."

C. <u>More Than Two Samples</u> - Fill in parenthesis.

<u>Reject H_o:</u> "Reject H_o at (($(1-a_1)$x100)% confidence level. There is evidence that at least one of the treatments tends to yield larger values of (MOE) than at least one other treatment. The probability that this result

is due purely to randomness is equal to (a_1)."

<u>Fail to reject H_o</u>: "Fail to reject H_o at ((1-a_1) x100)% confidence level. There is insufficient evidence to indicate that any of the treatments yield larger values of (MOE) than any other treatment."

Using the Guide

An Example: Using the following data determine whether a new avionics component (B) is more reliable than the currently used component (A). The data which has been collected is the time between failure, the selected MOE is MTBF (mean time between failure).

The Data:

Component Failure	1	2	3	4	5	6	7	8	9	10	S
Time of failure											
Component A	2.6	3.7	2.9	2.8	3.3	4.1	3.6	4.2	3.0	3.1	.55
Component B	4.2	3.3	3.8	4.2	3.1	2.9	4.4	3.7	3.9	3.5	.50

STEP I:

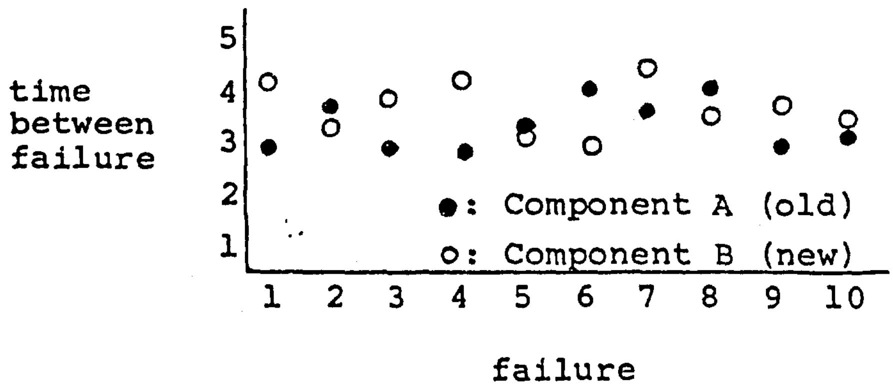


Figure A-2. Time Between Failures vs Failure

No trends are detected.

$E = .1$ hours

$$N_{r_A} = \hat{s}_A(\hat{s}_A+\hat{s}_B)\left(\frac{t_{.9,\infty}}{.1}\right)^2$$

$$= .5(1)(1.282/.1)^2 = 84$$

$N_{r_B} = 84$

STEP II:

$a_1 = .1$

$a_2 = .1$

STEP III:

H_o: Sample A is a random sample.

H_A: Sample A is not a random sample.

$\overline{X}_A$: MTBF of sample A.

$\overline{X}_A = 3.33$

$n_1 = 4 \quad n_2 = 6$

Failure:	1	2	3	4	5	6	7	8	9	10
TBF(X_i):	2.6	3.7	2.9	2.8	3.3	4.1	3.6	4.2	3.0	3.1
$X_i>\overline{X}$:	-	+	-	-	-	+	+	+	-	-
runs:	1	2		3			4			5

$T = 5 \quad a_2 = .1$

$L = U_{4,6,.05} = 3$

$U = U_{4,6,.95} = 8$

$L \leq T \leq U$ therefore fail to reject H_o, there is no evidence that sample A is not a random sample.

H_o: Sample B is a random sample.

H_A: Sample F is not a random sample.

$\overline{X}_B = 3.7$

$n_1 = 5 \quad n_2 = 5$

Failure:	1	2	3	4	5	6	7	8	9	10
TBF (X_i):	4.2	3.3	3.8	4.2	3.1	2.9	4.4	3.7	3.4	3.5
$X_i > \overline{X}$:	+	-	+	+	-	-	+	-	+	-
runs:	1	2	3		4		5	6	7	8

$T = 8 \quad a_2 = .1$

$L = U_{5,5,.05} = 3$

$U = U_{5,5,.95} = 8$

$L \leq T \leq U$ therefore fail to reject H_o, there is no evidence that sample B is not a random sample.

Type of MOE? - <u>Average</u>

$N_{r_A} = 34 \qquad n_A \geq N_{r_A}$? - <u>No</u>

$N_{r_B} = 34 \qquad n_B \geq N_{r_B}$? - <u>No</u>

<u>Go To pp. A-45 (3) Nonparametric Tests</u>

One sample? - No

Two samples? - Yes

H_o: $E(B) \leq E(A)$

H_A: $E(B) > E(A)$

$a = .1$ $I = 3$

Paired samples? - No

Go To pp. A-65   Mann-Whitney.

Sample A

X_{Ai}	R_{ij}
2.6	1
3.7	12.5
2.9	3.5
2.8	2
3.3	8.5
4.1	16
3.6	11
4.2	18
3.0	5
3.1	6.5
	84

Sample B

X_{Bi}	R_{ij}
4.2	18
3.3	8.5
3.8	14
4.2	18
3.1	6.5
2.9	3.5
4.4	20
3.7	12.5
3.9	15
3.5	10

$m = n = 10$ $N = 20$

$$T = \sum_{j=1}^{10} R_{Aj} = 84$$

$$L = U_{m,n,.1} = 32$$

$$U = 10(20+1) - 32 = 178$$

$I = 3$

$L \leq T$ therefore fail to reject H_o at $a=.1$.

Go To pp. A-78 Reporting Results

Fail to reject H_o at 90% confidence level. There is insufficient evidence to indicate that the MTBF of the new component is greater than that of the old component.

VIII. REFERENCES

In order to write a truly complete guide for OT&E data analysis, it would be necessary to include large sections of statistical theory and tables of probability distribution functions. This has not been done in the interest of utility, that is, so that this guide will be used as a quick reference, not as a complete treatment of the subject matter. Statistical tables, texts and software guides may also be necessary to perform the statistical analysis suggested in this guide. In order to standardize as much as possible, all of the tables referred to in the guide can be found in the CRC Handbook for Probability and Statistics (Ref 15). The exceptions are the binomial tables which are found in Table II and the Lilliefors test statistic in Table III. All of the texts cited in the bibliography contain statistical tables; however, none are as complete as the CRC (Ref 15).

The textbooks which are referenced in the guide are listed in the bibliography (pp. A-102) by topic in increasing order of mathematical complexity within each topic. Consulting a text prior to using an unfamiliar statistical technique will result in a better understanding by the user. At least two statistical references will be necessary as a minimum, one in the area of general statistics and one in nonparametric statistics.

Two statistical software packages, SPSS (Ref 22) and BMD

(Ref 16) have been cited in this guide because of their availability to the Air Force analyst and the quality of their documentation. SPSS and BMD are available at many bases and the copies can be transferred within the Air Force. Documentation can be ordered from:

BMD

University of California Press
2223 Fulton Street

Berkeley, California 94720

SPSS

SPSS Incorporate Inc.
Suite 3300
44 North Michigan Avenue

Chicago, Illinois 60611

As important to the OT&E analyst as textbooks and software are contacts within the Air Force with OT&E and/or statistical experience. The following is only a partial list of offices where personnel with these backgrounds can be consulted:

Air Force Test and Evaluation Center
AFTEC/OA AV-244-0437

Air Force Institute of Technology
AFIT/Math Department AV-285-3098
Operational Sciences Dept. AV-285-2549

U.S. Air Force Academy
USAFA/Math Department AV-259-4470

Table II Binomial Distribution[a]

n	y	p=.05	.10	.15	.20	.25	.30	.35	.40	.45
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.5400	.6000	.5500
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025
	1	.9975	.9900	.9775	.9600	.9375	.9100	.8775	.8400	.7975
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664
	1	.9928	.9720	.9392	.8960	.8438	.7840	.7182	.6480	.5748
	2	.9999	.9990	.9966	.9920	.9844	.9730	.9571	.9360	.9089
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915
	1	.9860	.9477	.8905	.8192	.7383	.6517	.5630	.4752	.3910
	2	.9995	.9963	.9880	.9728	.9492	.9163	.8735	.8208	.7585
	3	1.0000	.9999	.9995	.9984	.9961	.9919	.9850	.9743	.9590
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503
	1	.9774	.9185	.8352	.7373	.6328	.5282	.4284	.3370	.2562
	2	.9988	.9914	.9734	.9421	.8965	.8369	.7648	.6826	.5931
	3	1.0000	.9995	.9978	.9933	.9844	.9692	.9460	.9130	.8688
	4	1.0000	1.0000	.9999	.9997	.9990	.9976	.9947	.9898	.9815
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0567	.0277
	1	.9672	.8857	.7765	.6554	.5339	.4202	.3191	.2333	.1636
	2	.9978	.9842	.9527	.9011	.8306	.7443	.6471	.5443	.4415
	3	.9999	.9987	.9941	.9830	.9624	.9295	.8826	.8208	.7447
	4	1.0000	.9999	.9996	.9984	.9954	.9891	.9777	.9590	.9308
	5	1.0000	1.0000	1.0000	.9999	.9998	.9993	.9982	.9959	.9917
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152
	1	.9556	.8503	.7166	.5767	.4449	.3294	.2338	.1586	.1024
	2	.9962	.9743	.9262	.8520	.7564	.6471	.5323	.4199	.3164
	3	.9998	.9973	.9879	.9667	.9294	.8740	.8002	.7102	.6083
	4	1.0000	.9998	.9988	.9953	.9871	.9712	.9444	.9037	.8471
	5	1.0000	1.0000	.9999	.9996	.9987	.9962	.9910	.9812	.9643
	6	1.0000	1.0000	1.0000	1.0000	.9999	.9998	.9994	.9984	.9963
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

[a] Y has the binomial distribution with parameters n and p. The entries are the values of $P(Y \le y) = \sum_{i=0}^{y} \binom{n}{i} p^{i} (1-p)^{n-i}$, for p ranging from .05 to .95.

Table II (Continued)

n	y	p=.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
1	0	.5000	.4500	.4000	.3500	.3000	.2500	.2000	.1500	.1000	.0500
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0	.2500	.2025	.1600	.1225	.0900	.0625	.0400	.0225	.0100	.0025
	1	.7500	.6975	.6400	.5775	.5100	.4375	.3600	.2775	.1900	.0975
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	.1250	.0911	.0640	.0429	.0270	.0156	.0080	.0034	.0010	.0001
	1	.5000	.4252	.3520	.2818	.2160	.1562	.1040	.0608	.0280	.0072
	2	.8750	.8336	.7840	.7254	.6570	.5781	.4880	.3859	.2710	.1426
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	.0625	.0410	.0256	.0150	.0081	.0039	.0016	.0005	.0001	.0000
	1	.3125	.2415	.1792	.1265	.0837	.0508	.0272	.0120	.0037	.0005
	2	.6875	.6090	.5248	.4370	.3484	.2617	.1808	.1095	.0523	.0140
	3	.9375	.9085	.8704	.8215	.7599	.6836	.5904	.4780	.3439	.1855
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	.0312	.0185	.0102	.0053	.0024	.0010	.0003	.0001	.0000	.0000
	1	.1875	.1312	.0870	.0540	.0308	.0156	.0067	.0022	.0005	.0000
	2	.5000	.4069	.3174	.2352	.1631	.1035	.0579	.0266	.0086	.0012
	3	.8125	.7438	.6630	.5716	.4718	.3672	.2627	.1648	.0815	.0226
	4	.9688	.9497	.9222	.8840	.8319	.7627	.6723	.5563	.4095	.2262
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	.0156	.0083	.0041	.0018	.0007	.0002	.0001	.0000	.0000	.0000
	1	.1094	.0692	.0410	.0223	.0109	.0046	.0016	.0004	.0001	.0000
	2	.3438	.2553	.1792	.1174	.0705	.0376	.0170	.0059	.0013	.0001
	3	.6562	.5585	.4557	.3529	.2557	.1694	.0989	.0473	.0158	.0022
	4	.8906	.8364	.7667	.6809	.5798	.4661	.3446	.2235	.1143	.0328
	5	.9844	.9723	.9533	.9246	.8824	.8220	.7379	.6229	.4686	.2649
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	.0078	.0037	.0016	.0006	.0002	.0001	.0000	.0000	.0000	.0000
	1	.0625	.0357	.0188	.0090	.0038	.0013	.0004	.0001	.0000	.0000
	2	.2266	.1529	.0963	.0556	.0288	.0129	.0047	.0012	.0002	.0000
	3	.5000	.3917	.2898	.1998	.1260	.0706	.0333	.0121	.0027	.0002
	4	.7734	.6836	.5801	.4677	.3529	.2436	.1480	.0738	.0257	.0038
	5	.9375	.8976	.8414	.7662	.6706	.5551	.4233	.2834	.1497	.0444
	6	.9322	.9848	.9720	.9510	.9176	.8665	.7903	.6794	.5217	.3017
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.05	.10	.15	.20	.25	.30	.35	.40	.45
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084
	1	.9428	.8131	.6572	.5033	.3671	.2553	.1691	.1064	.0632
	2	.9942	.9619	.8948	.7969	.6785	.5518	.4278	.3154	.2201
	3	.9996	.9950	.9786	.9437	.8862	.8059	.7064	.5941	.4770
	4	1.0000	.9996	.9971	.9896	.9727	.9420	.8939	.8263	.7396
	5	1.0000	1.0000	.9998	.9988	.9958	.9887	.9747	.9502	.9115
	6	1.0000	1.0000	1.0000	.9999	.9996	.9987	.9964	.9915	.9819
	7	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9988	.9993	.9983
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046
	1	.9288	.7748	.5995	.4362	.3003	.1960	.1211	.0705	.0385
	2	.9916	.9470	.8591	.7382	.6007	.4628	.3373	.2318	.1495
	3	.9994	.9917	.9661	.9144	.8343	.7297	.6089	.4826	.3614
	4	1.0000	.9991	.9944	.9804	.9511	.9012	.8283	.7334	.6214
	5	1.0000	.9999	.9994	.9969	.9900	.9747	.9464	.9006	.8342
	6	1.0000	1.0000	1.0000	.9997	.9987	.9957	.9888	.9750	.9502
	7	1.0000	1.0000	1.0000	1.0000	.9999	.9996	.9986	.9962	.9909
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9992
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025
	1	.9139	.7361	.5443	.3758	.2440	.1493	.0860	.0464	.0233
	2	.9885	.9298	.8202	.6778	.5256	.3828	.2616	.1673	.0996
	3	.9990	.9872	.9500	.8791	.7759	.6496	.5138	.3823	.2660
	4	.9999	.9984	.9901	.9672	.9219	.8497	.7515	.6331	.5044
	5	1.0000	.9999	.9986	.9936	.9803	.9527	.9051	.8338	.7384
	6	1.0000	1.0000	.9999	.9991	.9965	.9894	.9740	.9452	.8980
	7	1.0000	1.0000	1.0000	.9999	.9996	.9984	.9952	.9877	.9726
	8	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9983	.9955
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014
	1	.8981	.6974	.4922	.3221	.1971	.1130	.0606	.0302	.0139
	2	.9848	.9104	.7788	.6174	.4552	.3127	.2001	.1189	.0652
	3	.9984	.9815	.9306	.8389	.7133	.5696	.4256	.2963	.1911
	4	.9999	.9972	.9841	.9496	.8854	.7897	.6683	.5328	.3971
	5	1.0000	.9997	.9973	.9883	.9657	.9218	.8513	.7535	.6331
	6	1.0000	1.0000	.9997	.9980	.9924	.9784	.9499	.9006	.8262
	7	1.0000	1.0000	1.0000	.9998	.9988	.9957	.9878	.9707	.9390
	8	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9980	.9941	.9852
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9993	.9978
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
8	0	.0039	.0017	.0007	.0002	.0001	.0000	.0000	.0000	.0000	.0000
	1	.0352	.0181	.0085	.0036	.0013	.0004	.0001	.0000	.0000	.0000
	2	.1445	.0885	.0498	.0253	.0113	.0042	.0012	.0002	.0000	.0000
	3	.3633	.2604	.1737	.1061	.0580	.0273	.0104	.0029	.0004	.0000
	4	.6367	.5230	.4059	.2936	.1941	.1138	.0563	.0214	.0050	.0004
	5	.8555	.7799	.6846	.5722	.4482	.3215	.2031	.1052	.0381	.0058
	6	.9648	.9368	.8936	.8309	.7447	.6329	.4967	.3428	.1869	.0572
	7	.9961	.9916	.9832	.9681	.9424	.8999	.8322	.7275	.5695	.3366
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	.0020	.0008	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0195	.0091	.0038	.0014	.0004	.0001	.0000	.0000	.0000	.0000
	2	.0898	.0498	.0250	.0112	.0043	.0013	.0003	.0000	.0000	.0000
	3	.2539	.1658	.0994	.0536	.0253	.0100	.0031	.0006	.0001	.0000
	4	.5000	.3786	.2666	.1717	.0988	.0489	.0196	.0056	.0009	.0000
	5	.7461	.6386	.5174	.3911	.2703	.1657	.0856	.0339	.0083	.0006
	6	.9102	.8505	.7682	.6627	.5372	.3993	.2618	.1409	.0530	.0084
	7	.9805	.9615	.9295	.8789	.8040	.6997	.5638	.4005	.2252	.0712
	8	.9980	.9954	.9899	.9793	.9596	.9249	.8658	.7684	.6126	.3698
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	.0010	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0107	.0045	.0017	.0005	.0001	.0000	.0000	.0000	.0000	.0000
	2	.0547	.0274	.0123	.0048	.0016	.0004	.0001	.0000	.0000	.0000
	3	.1719	.1020	.0548	.0260	.0106	.0035	.0009	.0001	.0000	.0000
	4	.3770	.2616	.1662	.0949	.0473	.0197	.0064	.0014	.0001	.0000
	5	.6230	.4956	.3669	.2485	.1503	.0781	.0328	.0099	.0016	.0001
	6	.8281	.7340	.6177	.4862	.3504	.2241	.1209	.0500	.0128	.0010
	7	.9453	.9004	.8327	.7384	.6172	.4744	.3222	.1798	.0702	.0115
	8	.9893	.9767	.9536	.9140	.8507	.7560	.6242	.4557	.2639	.0861
	9	.9990	.9975	.9940	.9865	.9718	.9437	.8926	.8031	.6513	.4013
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	.0005	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0059	.0022	.0007	.0002	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0327	.0148	.0059	.0020	.0006	.0001	.0000	.0000	.0000	.0000
	3	.1133	.0610	.0293	.0122	.0043	.0012	.0002	.0000	.0000	.0000
	4	.2744	.1738	.0994	.0501	.0216	.0076	.0020	.0003	.0000	.0000
	5	.5000	.3669	.2465	.1487	.0782	.0343	.0117	.0027	.0003	.0000
	6	.7256	.6029	.4672	.3317	.2103	.1146	.0504	.0159	.0028	.0001
	7	.8867	.8089	.7037	.5744	.4304	.2867	.1611	.0694	.0185	.0016
	8	.9673	.9348	.8811	.7999	.6873	.5448	.3826	.2212	.0896	.0152
	9	.9941	.9861	.9698	.9394	.8870	.8029	.6779	.5078	.3026	.1019
	10	.9995	.9986	.9964	.9912	.9802	.9578	.9141	.8327	.6862	.4312
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	P=.05	.10	.15	.20	.25	.30	.35	.40	.45
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008
	1	.8816	.6590	.4435	.2749	.1584	.0850	.0424	.0196	.0083
	2	.9804	.8891	.7358	.5583	.3907	.2528	.1513	.0834	.0421
	3	.9978	.9744	.9078	.7946	.6488	.4925	.3467	.2253	.1345
	4	.9998	.9957	.9761	.9274	.8424	.7237	.5833	.4382	.3044
	5	1.0000	.9995	.9954	.9806	.9456	.8822	.7873	.6652	.5269
	6	1.0000	.9999	.9993	.9961	.9857	.9614	.9154	.8418	.7393
	7	1.0000	1.0000	.9999	.9994	.9972	.9905	.9745	.9427	.8883
	8	1.0000	1.0000	1.0000	.9999	.9996	.9983	.9944	.9847	.9644
	9	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9992	.9972	.9921
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9989
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004
	1	.8646	.6213	.3983	.2336	.1267	.0637	.0296	.0126	.0049
	2	.9755	.8661	.7296	.5017	.3326	.2025	.1132	.0579	.0269
	3	.9969	.9658	.9033	.7473	.5843	.4206	.2783	.1686	.0929
	4	.9997	.9935	.9740	.9009	.7940	.6543	.5005	.3530	.2279
	5	1.0000	.9991	.9947	.9700	.9198	.8346	.7159	.5744	.4268
	6	1.0000	.9999	.9987	.9930	.9757	.9376	.8705	.7712	.6437
	7	1.0000	1.0000	.9998	.9988	.9944	.9818	.9538	.9023	.8212
	8	1.0000	1.0000	1.0000	.9998	.9990	.9960	.9874	.9679	.9302
	9	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9975	.9922	.9797
	10	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9987	.9959
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002
	1	.8470	.5846	.3567	.1979	.1010	.0475	.0205	.0081	.0029
	2	.9699	.8416	.6479	.4481	.2811	.1608	.0839	.0398	.0170
	3	.9958	.9559	.8535	.6982	.5213	.3552	.2205	.1243	.0632
	4	.9996	.9908	.9533	.8702	.7415	.5842	.4227	.2793	.1672
	5	1.0000	.9985	.9885	.9561	.8883	.7805	.6405	.4859	.3373
	6	1.0000	.9998	.9978	.9884	.9617	.9067	.8164	.6925	.5461
	7	1.0000	1.0000	.9997	.9976	.9897	.9685	.9247	.8499	.7414
	8	1.0000	1.0000	1.0000	.9996	.9978	.9917	.9757	.9417	.8811
	9	1.0000	1.0000	1.0000	1.0000	.9997	.9983	.9940	.9825	.9574
	10	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9989	.9961	.9886
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9978
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
12	0	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0032	.0011	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0193	.0079	.0028	.0008	.0002	.0000	.0000	.0000	.0000	.0000
	3	.0730	.0356	.0153	.0056	.0017	.0004	.0001	.0000	.0000	.0000
	4	.1938	.1117	.0573	.0255	.0095	.0028	.0006	.0001	.0000	.0000
	5	.3872	.2607	.1582	.0846	.0386	.0143	.0039	.0007	.0001	.0000
	6	.6128	.4731	.3348	.2127	.1178	.0544	.0194	.0046	.0005	.0000
	7	.8062	.6956	.5618	.4167	.2763	.1576	.0726	.0239	.0043	.0002
	8	.9270	.8655	.7747	.6533	.5075	.3512	.2054	.0922	.0256	.0022
	9	.9807	.9579	.9166	.8487	.7472	.6093	.4417	.2642	.1109	.0196
	10	.9968	.9917	.9804	.9576	.9150	.8416	.7251	.5565	.3410	.1184
	11	.9998	.9992	.9978	.9943	.9862	.9683	.9313	.8578	.7176	.4596
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	0	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0017	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0112	.0041	.0013	.0003	.0001	.0000	.0000	.0000	.0000	.0000
	3	.0461	.0203	.0078	.0025	.0007	.0001	.0000	.0000	.0000	.0000
	4	.1334	.0698	.0321	.0126	.0040	.0010	.0002	.0000	.0000	.0000
	5	.2905	.1788	.0977	.0462	.0182	.0056	.0012	.0002	.0000	.0000
	6	.5000	.3563	.2288	.1295	.0624	.0243	.0070	.0013	.0001	.0000
	7	.7095	.5732	.4256	.2841	.1654	.0802	.0300	.0053	.0009	.0000
	8	.8666	.7721	.6470	.4995	.3457	.2060	.0991	.0260	.0065	.0003
	9	.9539	.9071	.8314	.7217	.5794	.4157	.2527	.0967	.0342	.0031
	10	.9888	.9731	.9421	.8868	.7975	.6674	.4983	.2704	.1339	.0245
	11	.9983	.9951	.9874	.9704	.9363	.8733	.7664	.6017	.3787	.1354
	12	.9999	.9996	.9987	.9963	.9903	.9762	.9450	.8791	.7458	.4867
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0009	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0065	.0022	.0006	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0287	.0114	.0039	.0011	.0002	.0000	.0000	.0000	.0000	.0000
	4	.0898	.0462	.0175	.0060	.0017	.0003	.0000	.0000	.0000	.0000
	5	.2120	.1189	.0583	.0243	.0083	.0022	.0004	.0000	.0000	.0000
	6	.3953	.2586	.1501	.0753	.0315	.0103	.0024	.0003	.0000	.0000
	7	.6047	.4539	.3075	.1836	.0933	.0383	.0116	.0022	.0002	.0000
	8	.7880	.6627	.5141	.3595	.2195	.1117	.0439	.0115	.0015	.0000
	9	.9102	.8328	.7207	.5773	.4158	.2585	.1298	.0467	.0092	.0004
	10	.9713	.9368	.8757	.7795	.6448	.4787	.3018	.1465	.0441	.0042
	11	.9935	.9830	.9602	.9161	.8392	.7189	.5519	.3521	.1584	.0301
	12	.9991	.9971	.9919	.9795	.9525	.8990	.8021	.6433	.4154	.1530
	13	.9999	.9998	.9992	.9976	.9932	.9822	.9560	.8972	.7712	.5123
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 11 (Continued)

n	y	p=.05	.10	.15	.20	.25	.30	.35	.40	.45
15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001
	1	.8290	.5490	.3186	.1671	.0802	.0353	.0142	.0052	.0017
	2	.9638	.8159	.6042	.3980	.2361	.1268	.0617	.0271	.0107
	3	.9945	.9444	.8227	.6482	.4613	.2969	.1727	.0905	.0424
	4	.9994	.9873	.9383	.8358	.6865	.5155	.3519	.2173	.1204
	5	.9999	.9978	.9832	.9389	.8516	.7216	.5643	.4032	.2608
	6	1.0000	.9997	.9964	.9819	.9434	.8689	.7548	.6098	.4522
	7	1.0000	1.0000	.9994	.9958	.9827	.9500	.8868	.7869	.6535
	8	1.0000	1.0000	.9999	.9992	.9958	.9848	.9578	.9050	.8182
	9	1.0000	1.0000	1.0000	.9999	.9992	.9963	.9876	.9662	.9231
	10	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9972	.9907	.9745
	11	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9981	.9937
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997	.9989
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001
	1	.8108	.5147	.2839	.1407	.0635	.0261	.0098	.0033	.0010
	2	.9571	.7892	.5614	.3518	.1971	.0994	.0451	.0183	.0066
	3	.9930	.9316	.7899	.5981	.4050	.2459	.1339	.0651	.0281
	4	.9991	.9830	.9209	.7982	.6302	.4499	.2892	.1666	.0853
	5	.9999	.9967	.9765	.9183	.8103	.6598	.4900	.3288	.1976
	6	1.0000	.9995	.9944	.9733	.9204	.8247	.6881	.5272	.3660
	7	1.0000	.9999	.9989	.9930	.9729	.9256	.8406	.7161	.5629
	8	1.0000	1.0000	.9998	.9985	.9925	.9743	.9329	.8577	.7441
	9	1.0000	1.0000	1.0000	.9998	.9984	.9929	.9771	.9417	.8759
	10	1.0000	1.0000	1.0000	1.0000	.9997	.9984	.9938	.9809	.9514
	11	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9987	.9951	.9851
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9991	.9965
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
15	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0037	.0011	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0176	.0063	.0019	.0005	.0001	.0000	.0000	.0000	.0000	.0000
	4	.0592	.0255	.0093	.0028	.0007	.0001	.0000	.0000	.0000	.0000
	5	.1509	.0769	.0338	.0124	.0037	.0008	.0001	.0000	.0000	.0000
	6	.3036	.1818	.0950	.0422	.0152	.0042	.0008	.0001	.0000	.0000
	7	.5000	.3465	.2131	.1132	.0500	.0173	.0042	.0006	.0000	.0000
	8	.6964	.5478	.3902	.2452	.1311	.0566	.0181	.0036	.0003	.0000
	9	.8491	.7392	.5968	.4357	.2784	.1484	.0611	.0168	.0022	.0001
	10	.9408	.8796	.7827	.6481	.4845	.3135	.1642	.0617	.0127	.0006
	11	.9824	.9576	.9095	.8273	.7031	.5387	.3518	.1773	.0556	.0055
	12	.9963	.9893	.9729	.9383	.8732	.7639	.6020	.3958	.1841	.0362
	13	.9995	.9983	.9948	.9858	.9647	.9198	.8329	.6814	.4510	.1710
	14	1.0000	.9999	.9995	.9984	.9953	.9866	.9648	.9126	.7941	.5367
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
16	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0021	.0006	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0106	.0035	.0009	.0002	.0000	.0000	.0000	.0000	.0000	.0000
	4	.0384	.0149	.0049	.0013	.0003	.0000	.0000	.0000	.0000	.0000
	5	.1051	.0486	.0191	.0062	.0016	.0003	.0000	.0000	.0000	.0000
	6	.2272	.1241	.0583	.0229	.0071	.0016	.0002	.0000	.0000	.0000
	7	.4018	.2559	.1423	.0671	.0257	.0075	.0015	.0002	.0000	.0000
	8	.5982	.4371	.2839	.1594	.0744	.0271	.0070	.0011	.0001	.0000
	9	.7728	.6340	.4728	.3119	.1753	.0796	.0267	.0056	.0005	.0000
	10	.8949	.8024	.6712	.5100	.3402	.1897	.0817	.0235	.0033	.0001
	11	.9616	.9147	.8334	.7108	.5501	.3698	.2018	.0791	.0170	.0009
	12	.9894	.9719	.9349	.8661	.7541	.5950	.4019	.2101	.0684	.0070
	13	.9979	.9934	.9817	.9549	.9006	.8729	.6482	.4386	.2108	.0429
	14	.9997	.9990	.9967	.9902	.9739	.9365	.8593	.7161	.4853	.1892
	15	1.0000	.9999	.9997	.9990	.9967	.9900	.9719	.9257	.8147	.5599
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.05	.10	.15	.20	.25	.30	.35	.40	.45
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000
	1	.7922	.4818	.2525	.1182	.0501	.0193	.0067	.0021	.0006
	2	.9497	.7618	.5198	.3096	.1637	.0774	.0327	.0123	.0041
	3	.9912	.9174	.7556	.5489	.3530	.2019	.1028	.0464	.0184
	4	.9988	.9779	.9013	.7582	.5739	.3887	.2348	.1260	.0596
	5	.9999	.9953	.9681	.8943	.7653	.5968	.4197	.2639	.1471
	6	1.0000	.9992	.9917	.9623	.8929	.7752	.6188	.4478	.2902
	7	1.0000	.9999	.9983	.9891	.9598	.8954	.7872	.6405	.4743
	8	1.0000	1.0000	.9997	.9974	.9876	.9597	.9006	.8011	.6626
	9	1.0000	1.0000	1.0000	.9995	.9969	.9873	.9617	.9081	.8166
	10	1.0000	1.0000	1.0000	.9999	.9994	.9968	.9880	.9652	.9174
	11	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9970	.9894	.9699
	12	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9975	.9914
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995	.9981
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000
	1	.7735	.4503	.2241	.0991	.0395	.0142	.0046	.0013	.0003
	2	.9419	.7338	.4797	.2713	.1353	.0600	.0236	.0082	.0025
	3	.9891	.9018	.7202	.5010	.3057	.1646	.0783	.0328	.0120
	4	.9985	.9718	.8794	.7164	.5187	.3327	.1886	.0942	.0411
	5	.9998	.9936	.9581	.8671	.7175	.5344	.3550	.2088	.1077
	6	1.0000	.9988	.9882	.9487	.8610	.7217	.5491	.3743	.2258
	7	1.0000	.9998	.9973	.9837	.9431	.8593	.7283	.5634	.3915
	8	1.0000	1.0000	.9995	.9957	.9807	.9404	.8609	.7368	.5778
	9	1.0000	1.0000	.9999	.9991	.9946	.9790	.9403	.8653	.7473
	10	1.0000	1.0000	1.0000	.9998	.9988	.9939	.9788	.9424	.8720
	11	1.0000	1.0000	1.0000	1.0000	.9998	.9986	.9938	.9797	.9463
	12	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9986	.9942	.9817
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9987	.9951
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9998	.9990
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
17	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0012	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0064	.0019	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	4	.0245	.0086	.0025	.0006	.0001	.0000	.0000	.0000	.0000	.0000
	5	.0717	.0301	.0106	.0030	.0007	.0001	.0000	.0000	.0000	.0000
	6	.1662	.0826	.0348	.0120	.0032	.0006	.0001	.0000	.0000	.0000
	7	.3145	.1834	.0919	.0383	.0127	.0031	.0005	.0000	.0000	.0000
	8	.5000	.3374	.1989	.0994	.0403	.0124	.0026	.0003	.0000	.0000
	9	.6855	.5257	.3595	.2128	.1046	.0402	.0109	.0017	.0001	.0000
	10	.8338	.7098	.5522	.3812	.2248	.1071	.0377	.0083	.0008	.0000
	11	.9283	.8529	.7361	.5803	.4032	.2347	.1057	.0319	.0047	.0001
	12	.9755	.9404	.8740	.7652	.6113	.4261	.2418	.0987	.0221	.0012
	13	.9936	.9816	.9536	.8972	.7981	.6470	.4511	.2444	.0826	.0088
	14	.9988	.9959	.9877	.9673	.9226	.8363	.6904	.4802	.2382	.0503
	15	.9999	.9994	.9979	.9933	.9807	.9499	.8818	.7475	.5182	.2078
	16	1.0000	1.0000	.9998	.9993	.9977	.9925	.9775	.9369	.8332	.5819
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0007	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0038	.0010	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	4	.0154	.0049	.0013	.0003	.0000	.0000	.0000	.0000	.0000	.0000
	5	.0481	.0183	.0058	.0014	.0003	.0000	.0000	.0000	.0000	.0000
	6	.1189	.0537	.0203	.0062	.0014	.0002	.0000	.0000	.0000	.0000
	7	.2403	.1280	.0576	.0212	.0061	.0012	.0002	.0000	.0000	.0000
	8	.4073	.2527	.1347	.0597	.0210	.0054	.0009	.0001	.0000	.0000
	9	.5927	.4222	.2632	.1391	.0596	.0193	.0043	.0005	.0000	.0000
	10	.7597	.6085	.4366	.2717	.1407	.0569	.0163	.0027	.0002	.0000
	11	.8811	.7742	.6257	.4509	.2783	.1390	.0513	.0118	.0012	.0000
	12	.9519	.8923	.7912	.6450	.4656	.2825	.1329	.0419	.0064	.0002
	13	.9846	.9589	.9058	.8114	.6673	.4813	.2836	.1206	.0282	.0015
	14	.9962	.9880	.9672	.9217	.8354	.6943	.4990	.2798	.0982	.0109
	15	.9993	.9975	.9918	.9764	.9400	.8647	.7287	.5203	.2662	.0581
	16	.9999	.9997	.9987	.9954	.9858	.9605	.9009	.7759	.5497	.2265
	17	1.0000	1.0000	.9999	.9996	.9984	.9944	.9820	.9464	.8499	.6028
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.05	.10	.15	.20	.25	.30	.35	.40	.45
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000
	1	.7547	.4203	.1985	.0829	.0310	.0104	.0031	.0008	.0002
	2	.9335	.7054	.4413	.2369	.1113	.0462	.0170	.0055	.0015
	3	.9869	.8850	.6841	.4551	.2631	.1332	.0591	.0230	.0077
	4	.9980	.9648	.8556	.6733	.4654	.2822	.1500	.0696	.0280
	5	.9998	.9914	.9463	.8369	.6678	.4739	.2968	.1629	.0777
	6	1.0000	.9983	.9837	.9324	.8251	.6655	.4812	.3081	.1727
	7	1.0000	.9997	.9959	.9767	.9225	.8180	.6656	.4878	.3169
	8	1.0000	1.0000	.9992	.9933	.9713	.9161	.8145	.6675	.4940
	9	1.0000	1.0000	.9999	.9984	.9911	.9674	.9125	.8139	.6710
	10	1.0000	1.0000	1.0000	.9997	.9977	.9895	.9653	.9115	.8159
	11	1.0000	1.0000	1.0000	1.0000	.9995	.9972	.9886	.9648	.9129
	12	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9969	.9884	.9658
	13	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9993	.9969	.9891
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9994	.9972
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9995
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000
	1	.7358	.3917	.1756	.0692	.0243	.0076	.0021	.0005	.0001
	2	.9245	.6769	.4049	.2061	.0913	.0355	.0121	.0036	.0009
	3	.9841	.8670	.6477	.4114	.2252	.1071	.0444	.0160	.0049
	4	.9974	.9568	.8298	.6296	.4148	.2375	.1182	.0510	.0189
	5	.9997	.9887	.9327	.8042	.6172	.4164	.2454	.1256	.0553
	6	1.0000	.9976	.9781	.9133	.7858	.6080	.4166	.2500	.1299
	7	1.0000	.9996	.9941	.9679	.8982	.7723	.6010	.4159	.2520
	8	1.0000	.9999	.9987	.9900	.9591	.8867	.7624	.5956	.4143
	9	1.0000	1.0000	.9998	.9974	.9861	.9520	.8782	.7553	.5914
	10	1.0000	1.0000	1.0000	.9994	.9961	.9829	.9468	.8725	.7507
	11	1.0000	1.0000	1.0000	.9999	.9991	.9949	.9804	.9435	.8692
	12	1.0000	1.0000	1.0000	1.0000	.9998	.9987	.9940	.9790	.9420
	13	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9985	.9935	.9786
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9984	.9936
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997	.9985
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9997
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table II (Continued)

n	y	p=.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
19	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0004	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0022	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	4	.0096	.0028	.0006	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	5	.0318	.0109	.0031	.0007	.0001	.0000	.0000	.0000	.0000	.0000
	6	.0835	.0342	.0116	.0031	.0006	.0001	.0000	.0000	.0000	.0000
	7	.1796	.0871	.0352	.0114	.0028	.0005	.0000	.0000	.0000	.0000
	8	.3238	.1841	.0885	.0347	.0105	.0023	.0003	.0000	.0000	.0000
	9	.5000	.3290	.1861	.0875	.0326	.0089	.0016	.0001	.0000	.0000
	10	.6762	.5060	.3325	.1855	.0839	.0287	.0067	.0008	.0000	.0000
	11	.8204	.6831	.5122	.3344	.1820	.0775	.0233	.0041	.0003	.0000
	12	.9165	.8273	.6919	.5188	.3345	.1749	.0676	.0163	.0017	.0000
	13	.9682	.9223	.8371	.7032	.5261	.3322	.1631	.0537	.0086	.0002
	14	.9904	.9720	.9304	.8500	.7178	.5346	.3267	.1444	.0352	.0020
	15	.9978	.9923	.9770	.9409	.8668	.7369	.5449	.3159	.1150	.0132
	16	.9996	.9985	.9945	.9830	.9538	.8887	.7631	.5587	.2946	.0665
	17	1.0000	.9998	.9992	.9969	.9896	.9690	.9171	.8015	.5797	.2453
	18	1.0000	1.0000	.9999	.9997	.9989	.9958	.9856	.9544	.8649	.6226
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	3	.0013	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	4	.0059	.0015	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	5	.0207	.0064	.0016	.0003	.0000	.0000	.0000	.0000	.0000	.0000
	6	.0577	.0214	.0065	.0015	.0003	.0000	.0000	.0000	.0000	.0000
	7	.1316	.0580	.0210	.0060	.0013	.0002	.0000	.0000	.0000	.0000
	8	.2517	.1308	.0565	.0196	.0051	.0009	.0001	.0000	.0000	.0000
	9	.4119	.2493	.1275	.0532	.0171	.0039	.0006	.0000	.0000	.0000
	10	.5881	.4086	.2447	.1218	.0480	.0139	.0026	.0002	.0000	.0000
	11	.7483	.5857	.4044	.2376	.1133	.0409	.0100	.0013	.0001	.0000
	12	.8684	.7480	.5841	.3990	.2277	.1018	.0321	.0059	.0004	.0000
	13	.9423	.8701	.7500	.5834	.3920	.2142	.0867	.0219	.0024	.0000
	14	.9793	.9447	.8744	.7546	.5836	.3828	.1958	.0673	.0113	.0003
	15	.9941	.9811	.9490	.8818	.7625	.5852	.3704	.1702	.0432	.0026
	16	.9987	.9951	.9840	.9556	.8929	.7748	.5886	.3523	.1330	.0159
	17	.9998	.9991	.9964	.9879	.9645	.9087	.7939	.3951	.3231	.0755
	18	1.0000	.9999	.9995	.9979	.9924	.9757	.9308	.8244	.6083	.2642
	19	1.0000	1.0000	1.0000	.9998	.9992	.9968	.9885	.9612	.8784	.6415
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

For n larger than 20, the rth quantile y_r of a binomial random variable may be approximated using $y_r = np + w_r\sqrt{np(1-p)}$, where w_r is the rth quantile of a standard normal random variable, obtained from Table A1.

Table III Quantiles of the Lilliefors Test Statistic for Normality

	p=.80	.85	.90	.95	.99
Sample size n = 4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.142	.147	.158	.173	.200
30	.131	.136	.144	.161	.187
Over 30	$\frac{.736}{\sqrt{n}}$	$\frac{.768}{\sqrt{n}}$	$\frac{.805}{\sqrt{n}}$	$\frac{.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

SOURCE. Adapted from Table 1 of Lilliefors (1967), with corrections, from Conover (10).

BIBLIOGRAPHY

General Statistics (in order of mathematical complexity)

1. Freund, J.E., Livermore, P.E., Manual for Experimental Statistics, Prentice-Hall, Englewood Cliffs, 1960.

2. Romano, A., Applied Statistics for Science and Industry, Allyn and Bacon, Inc., Boston, 1977.

3. McClave, J.T., Benson, P.G., Statistics for Business and Economics, Second Edition, Dellen Publishing Co., San Francisco, 1982.

4. Miller, I., Freund, J.E., Probability and Statistics for Engineers, Prentice-Hall Inc., Englewood Cliffs, 1965.

5. Snedecor, G.W., Cochran, W.G., Statistical Methods, Seventh Edition, The Iowa State University Press, Ames, 1980.

6. Hines, W.W., Montgomery, D.C., Probability and Statistics in Engineering and Management Science, Second Edition, John Wiley & Sons, New York, 1980.

7. Mendenhall, W., Scheaffer, R.L., Wackerly, D.D., Mathematical Statistics with Applications, Second Edition, Duxbury Press, Boston, 1981.

8. Hogg, R.V., Craig, A.T., Introduction to Mathematical Statistics, Third Edition, Macmillan Publishing Co., Inc., New York, 1970.

Nonparametric Statistics (in order of mathematical complexity)

9. Siegel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Co., New York, 1956.

10. Conover, W.J., Practical Nonparametric Statistics, Second Edition, John Wiley & Sons, New York, 1980.

11. Bradley, J.V., Distribution-Free Statistical Tests, WADD Technical Report 60-661, (DTIC AD-249268), Wright Air Development Division, Wright-Patterson AFB, 1960.

12. Tate, M.W., Celland, R.C., Nonparametric and Shortcut Statistics, Interstate Printers and Publishers, Inc., Danville, 1957.

13. Walsh, J.E., Handbook of Nonparametric Statistics, D. Van Nostrand Co., Inc., Princeton, 1962.

14. Walsh, J.E., Handbook of Nonparametric Statistics, II, D. Van Nostrand Co., Inc., Princeton, 1965.

Special Statistical Topics

15. Beyer, W.H., (editor), CRC Handbook of Tables for Probability and Statistics, Second Edition, The Chemical Rubber Co., Cleveland, 1968.

16. Brown, M.B., (editor), BMDP-77 Biomedical Computer Programs P-Series, University of California Press, Berkeley, 1977.

17. Hald, A., Statistical Theory with Engineering Applications, John Wiley & Sons, New York, 1952.

18. Hicks, C.R., Fundamental Concepts in the Design of Experiments, Second Edition, Holt, Rinehart and Winston, New York, 1974.

19. Johnson, R.A., Wichern, D.W., Applied Multivariate Statistical Analysis, Prentice-Hall, Inc., Englewood Cliffs, 1982.

20. Kleinbaum, D.G., Kupper, L.L., Applied Regression Analysis and Other Multivariable Methods, Duxbury Press, North Scituate, 1978.

21. Neter, J., Wasserman, W., Applied Linear Statistical Models, Richard D. Irwin Inc., Homewood, 1974.

22. Nie, N.H., et al, Statistical Package for the Social Sciences, Second Edition, McGraw-Hill Book Co., New York, 1975.

Operational Testing

23. Analyst Notebook, Commander Operational Test and Evaluation Force Instruction 3960.7A, 16 February 1982.

24. Management of Operational Test and Evaluation, Air Force Manual 55-43, Vol. 1, 2, 13 June, 1979.

25. Report to the President and Secretary of Defense on the Department of Defense by the Blue Ribbon Defense Panel 1 July, 1970.

26. Research and Development Test and Evaluation, Air Force Regulation 80-14, 12 September, 1980.

27. Test and Evaluation, DOD Directive 5000.3, 26 December, 1979.

28. Cook, I.D., Measures of Munitions Delivery Accuracy, Working Paper, OAA File 4-41, AFTEC, Kirtland AFB, 1974

29. Cox, D.P., et al, Study of Operational Testing and Evaluation Experience, (MTR-6084), The MITRE Corp., Washington D.C., 1971.

30. Simon, D.E., et al, Standard Procedures for Air Force Operational Test and Evaluation, RADC Technical Report 74-270, (DTIC AD-B000-365, B000-366), Rome Air Development Center, AFSC, Griffis AFB, 1974.

31. Stevens, R.T., Operational Test and Evaluation: A Systems Engineering Process, John Wiley & Sons Inc., New York, 1979.

32. Waller, S.L., "Operational Test and Evaluation in Support of Executive Decisions," Astronautics & Aeronautics, October, 1977.

VITA

Gregg M. Burgess was born at Mather AFB, Sacremento, California on 14 August 1956. He graduated from the United States Naval Academy in 1978 with a Bachelor of Science degree in Operations Analysis. He was commissioned as a Second Lieutenant upon graduation.

He was assigned to the F-16 Multinational Operational Test and Evaluation as an operations analyst. His following assignment was to the Air Force Institute of Technology in June 1981.

Permanent Address: 216 Little Falls St.
Falls Church, Va. 22406

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/GOR/OS/82S-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A GUIDE FOR THE APPLICATION OF STATISTICAL ANALYSIS METHODS TO OPERATIONAL TEST AND EVALUATION DATA		5. TYPE OF REPORT & PERIOD COVERED MS Thesis 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gregg M. Burgess Captain, USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology Wright-Patterson AFB, Ohio 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE July 1982 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Statistics
Test and Evaluation
User's Guide
Statistical Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This research effort was directed toward developing a practical, easy-to-use guide for statistical analysis of Test and Evaluation data. Ease of use and completeness have both been stressed in writing this guide. The guide is written for users with little mathematical background, however should be useful to all Test and Evaluation analysts and managers. The guide includes a discussion of the basic statistical terminology as well as a step-by-step

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. ABSTRACT

analysis procedure. The analysis procedure is incorporated into a flow chart and includes discussions and examples of individual statistical tests. The statistical tests include parametric and nonparametric tests which can be applied to most test data types.