STATISTICAL TESTS FOR TRUNCATION IN VALIDATING THE RESULTS OF VARIABLES SAMPLING INSPECTION

By

ROGER WAYNE BERGER

Bachelor of Science
University of Nebraska
Lincoln, Nebraska
1958

Master of Science
Kansas State University
Manhattan, Kansas
1962

Submitted to the faculty of the Graduate
College of the Oklahoma State University
in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY
July, 1968

STATISTICAL TESTS FOR TRUNCATION IN VALIDATING

THE RESULTS OF VARIABLES SAMPLING INSPECTION

Thesis Approved:

M. Palmer Terrell
Thesis Adviser

James E. Shamblin

G. T. Stevens, Jr.

Rollo E. Venn

[illegible]

N. N. Durham
Dean of the Graduate College

PREFACE

Data validation comprises a set of quality assurance methods which have recently shown promise for reducing the cost of quality assurance. There has been considerable development of the data validation technique for the case of attributes-type data; the research reported in this dissertation is part of an effort by the author to extend this important concept into the area of measurement-type data.

The dissertation is organized into seven chapters. In Chapter I the history of previous data validation developments is sketched, and data validation is compared and contrasted with certain other quality assurance methods.

Chapter II describes measurement-type data discrepancies and includes a dichotomization of the data validation problem.

In Chapter III truncation in sample selection is developed as a model of an important type of data discrepancy, and the parameters of the truncated normal distribution are derived.

Then in Chapters IV and V, a set of statistical tests are reviewed and their abilities to detect truncation are analyzed. Chapter IV contains a description and brief history of the development of these tests, which fall into two basic categories: parametric and nonparametric. Equations for the operating characteristic curves of the parametric tests are derived through the application of statistical distribution theory. In Chapter V, the technique of distribution sampling is used to estimate the operating characteristics of the nonparametric tests.

In Chapter VI the different tests are compared on the basis of their relative effectiveness and potential contribution to data validation. Finally, in Chapter VII, the overall scope of the research is summarized and recommendations are made for further research on the validation of variables-type data.

Many individuals have planed an instrumental role in the effort which is culminated in this dissertation. First should be mentioned Dr. Irvin Reis, now Head of Mechanical Engineering at Montana State University, who inspired me by his own example and persuasion first to choose the field of industrial engineering for a profession, then to pursue graduate work, and finally to seek the Ph.D.

The faculty of Oklahoma State University has helped me greatly. Special thanks are due Dr. M. Palmer Terrell, my research advisor, for his advice regarding numerous aspects of the performance of the research and preparation of this thesis. The other members of my dissertation committee were Dr. Paul Torgersen, Dr. James Shamblin, Dr. David Bee, Dr. G. T. Stevens, and Dean Rollo Venn. All have cheerfully provided assistance and guidance and have contributed suggestions incorporated in the final version of the dissertation. Professor Wilson Bentley has given me valuable personal and professional counsel as well as helpful administrative support. Dr. David Weeks stimulated my interest in statistics and provided me with a model of the art of teaching that will remain long in my memory.

Financial support for two years of full time academic work was provided by the National Aeronautics and Space Administration. The Atomic Energy Commission, through its contract with Sandia Corporation, also contributed indirectly to my support as well as to the excellent professional environment at Sandia. The vital aid derived from these two Federal agencies will be recalled each April as I file my income tax return.

Drafts of the dissertation were typed by my wife Kay, whose unflagging loyalty and moral support made the whole idea feasible. The final typing was by Mrs. Rebecca H. Ellis and the figures drawn by Eldon Hardy, both of whom deserve credit for their quality workmanship.

Finally, acknowledgement must be made to Mr. L. E. Snodgrass, who was for five years my supervisor in the Quality Control Systems Division at Sandia Corporation. It was under the supervision of Mr. Snodgrass that my initial interest in data validation developed. In appreciation of his many personal kindnesses, his constant encouragement concerning my graduate work, and for the influence he has had, both as friend and supervisor, on my career, this dissertation is respectfully dedicated to Larry Snodgrass.

TABLE OF CONTENTS

Chapter | Page

LIST OF TABLES

LIST OF FIGURES

Figure		Page

CHAPTER I

INTRODUCTION

One of the major functions of any inspection, quality control, or quality assurance organization is to provide assurance to management and to the customer that the firm's products are of satisfactory quality. Over the years many different approaches have been developed toward meeting this goal.

Historical Background

One approach which has recently been under study is the technique of data validation. The major development to data has been done by procurement agencies of the U. S. Government. The U. S. Department of Defense established (1954) a uniform policy on quality assurance as related to acceptance inspection. This policy statement enunciated the concept of data validation by requiring that data generated by the supplier of a product should be utilized to as great an extent as feasible in determining the acceptability of material submitted by the supplier. A prerequisite to this usage of supplier data is determining that the data is reliable. The various methods of establishing that supplier data are in fact reliable have come to be known as data validation techniques.

To assist the Government representative in executing the policies set forth by the above Instructions, the Department of Defense (1960) published *DOD Handbook H-109*, a handbook of statistical procedures for determining the validity of supplier's attributes inspection. The authors of *DOD Handbook H-109* were Harry Elner, who published (1963) the mathematical background, and Joseph Mandelson, who published (1964) a non-technical discussion of the philosophy, purpose, and procedures, of attribute data validation.

The policy requiring independent sampling for validation was laid out in *H-109*:

> The requirement that the units of the sample be selected at random without regard to their quality cannot be verified by reinspecting the items drawn by the supplier. Only an independent sample selected by the consumer can authenticate the over-all effect of the supplier's inspection in assuring conformance of supplies with technical requirements and evaluate the true quality of the supplies offered to the consumer for acceptance. (Department of Defense, 1960, p. 2.)

Thus, the precedent was established requiring independent samples, though explicitly for attributes data only, which was carried over into the initial thinking on a validation system for measurement results (Elner and Mandelson, 1964; Berger, 1966). Some of the implications of this independent sample policy, as regards consumer protection and sample size, are developed in later chapters of this paper.

Relationship of Data Validation to Other Quality Assurance Systems

The principles of data validation should be compared with some of the other quality assurance approaches. One such approach is product acceptance sampling. A large number of techniques of this type have been developed over the past four decades. The numerous techniques available have been documented by Cowden, (1957, Ch. 30-39) and in other standard quality control references.

It may seem natural to make a direct comparison between data validation sampling and product acceptance sampling, because both consist, basically, of sample inspection by a consumer in order to purchase material submitted by a supplier. However, a fundamental difference exists between the two concepts that would make direct numerical comparisons misleading if not meaningless. In a data validation scheme it is not the product but the supplier's sampling and inspection capability which is being scrutinized. Data validation procedures should be expected only to tell a consumer that his supplier's inspection results are a valid representation of the material he is buying. The risks involved in a data validation program are related to the supplier's sampling and inspection system, rather than to the product itself. The consumer depends, in the framework of a data validation program, upon his supplier to tell him if the product he submits is of satisfactory quality. The validation program should provide an incentive for the supplier to consistently tell the truth about the material he is selling and should provide a means for the consumer to check to see that the supplier is doing so. There are many forms in which this untruth (that is, invalid data) might appear, and different approaches with differing types of protection may be required to cope with these various forms.

Although it is difficult to relate the technique of data validation to acceptance sampling inspection, a rather close relationship does exist between data validation and the general area of supplier quality audit systems. A number of different quality audit systems are now operating in American industry. They have been designed around the philosophy that for certain types of material, the customer has a need to know, not only the quality of the actual product characteristics inspected, but also the basic capability of the supplier to maintain a satisfactory level of quality through adequate procedures, controls, and equipment.

A brief overview of quality auditing systems is provided by Hansen (1963, p. 83), who explains how the customer can ascertain that the various components of the supplier's entire production-quality process are adequate. A typical quality audit system would include checklists and sampling procedures for reviewing such areas as: drawing release, updating, and retrieval; gage control; calibration of all inspection equipment; personnel training; and general quality control procedures. A customer team would visit the supplier's plant armed with such checklists and would proceed to compare the actual existing situation with the ideals represented by the lists.

The role of quality audit systems and data validation techniques was delineated by Whittlemore (1952) who said that while some manufacturers do not have the facilities nor personnel to determine whether their products comply with technical requirements, others control their processes carefully and test the finished product to see that it meets required quality standards. According to Whittlemore:

> Often the purchaser duplicates these tests, more or less closely, for acceptance. There would be great advantages, particularly in the cost of making the tests, if the quality was determined once for both parties.... It should be possible for the customers, or their association to share in the management of the inspection department and in the cost - then they could accept the results with confidence.

Whereas the quality audit system is intended to verify that the supplier has a capability to produce products of the desired quality, the data validation system is intended to verify that the supplier has accurately reported the true quality of his products. Thus the two systems tend to complement each other.

Data validation is also closely related to a set of techniques which have variously been called round-robin, collaborative, cooperative, interlaboratory, or intercomparison testing (Nelson, 1967). Nelson showed that interlaboratory programs might be instituted for a variety of reasons, such as to compare variabilities or precision of methods, to determine the biases between methods and ascertain reasons for their existence, to check operators or laboratories for evaluation of their efficiency, qualification, or certification, and others. His article mentions, as a possible application, cooperative use by supplier and purchaser to determine degree of adherence to purchase specifications. More detailed information about performing interlaboratory evaluation of testing methods can be found in the article by Mandel and Lashoff (1959) which analyzes the practical aspects of application.

The interlaboratory testing technique has been most frequently applied in the chemical industry, but it is equally applicable in any situation where measurement results of two or more organizations must be compared. It is a more general approach than data validation since it provides for various numbers of laboratores, types of testing procedures, and a whole hierarchy of other experimental factors. The development of a data validation program could possibly be strengthened by consideration of the numerous facets of interlaboratory testing techniques, such as those summarized and documented in Nelson's article.

Scope of This Research

The research effort reported upon here is focused upon identification of problems associated with the validation of supplier's variables-type inspection data. The author will show, in the following chapter, that data discrepancies can be divided into two mutually exclusive classes: measurement discrepancies and sampling discrepancies. For the detection of measurement discrepancies it is proposed that the method of paired comparisons be utilized on the basis of existing knowledge about the relative merits of paired and independent samples.

This leaves for more intensive study the problems associated with sampling discrepancies. The remainder of the dissertation is concerned with the development of a specific model of sampling discrepancy: truncation in sample selection; and with the study of a variety of different approaches to the truncation problem. By means of theoretical and empirical analysis, several two-sample statistical tests will be compared on the basis of their power to detect truncation when it exists. With this information available, a better decision can be made in any particular case as to whether the truncation sampling discrepancies should be tested for statistically; and if so, which test offers the best chance of detecting it.

CHAPTER II

DATA DISCREPANCIES

The Data Validation Principle

In a product acceptance program which utilizes data validation, the attention of the customer will be focused on the capability of the supplier's quality inspection system to accurately report the quality of the product submitted for purchase. While the customer is interested most of all in the quality of the product, his validation program will be designed to inform him as to the quality of the data. He may wish, for example, to utilize the quality audit, described in the previous chapter, to establish that his supplier has the capability to accurately and consistently report product quality. He may also see fit to insist on certain procedures, controls, additional training of operators, purchase and calibration of equipment, and other similar measures to be satisfied that the supplier's quality inspection system is adequate.

Once all such problems have been resolved, the prudent customer will still want to make empirical verifications of the quality of his supplier's data, by inspecting material which has been passed as acceptable by the supplier's quality inspection system. At this point the customer must still maintain his focus upon the inspection system rather than the product itself. For instance it is entirely possible that the customer will find defective product in the data validation sample he inspects. In fact, he may find enough defectiveness that he feels compelled to reject the entire lot of material under question, or even to re-evaluate the merit of continuing to purchase material from the particular supplier. However, the quality of the material as revealed by the customer's inspection should be recognized as of secondary importance. The primary criterion will be the degree of agreement or disagreement between the customer's results and the supplier's results.

A difference between supplier results and customer results can be the result of any one (or possibly all) of three different causes. In the first place it is possible that there is a discrepancy in the customer's inspection and validation system. While this is possible, it will behoove the customer to go to considerable pains to minimize this source of difference. He will do this by careful establishment of the calibration standards of equipment, training of operators, and by rechecking his own results whenever a significant difference is noted. For the purposes of this dissertation it will be assumed that the customer's system may be accepted as a standard and is thus not the source of any difference.

A second cause of difference may be pure random variation. It is recognized that all measured variables are subject to a certain amount of random variation. In some instances this variation will be so slight that it can be completely ignored. In others, it will have to be explicitly considered and its effect on the probability of error determined. Besides random measurement variation there is the problem of random product variation. Within the same production lot there will usually be unit-to-unit differences in measurable characteristics, and in many cases the product variation allowable by the purchase specifications is of a greater order of magnitude than the random measurement variation allowable for the inspection and test equipment. This random product variation plays a particularly important role when it is either impossible or impractical for the supplier and customer to both measure identical units of the product in the validation process.

The final cause of difference may be actual discrepancy in the supplier's data. This is, of course, what the quality audit program is designed to prevent, and what the data validation program is designed to detect, when it exists. There are many different ways in which discrepancies in supplier data may exist, either by accident or design. However, it now will be shown that these discrepancies all tend to fall into one of two distinctly different types.

The Dichotomy of Supplier Data Discrepancies

Differences between supplier and customer results can be attributed to one or more of three different causes, as described above. In this connection, one of the greatest concerns is in differentiating pure random variation from discrepancies in the supplier data. Although there are numerous discrepancies which could exist in the supplier's data, it can be seen that all these discrepancies fall into two mutually exclusive classes, namely, (1) discrepancies in the actual measurements recorded, and (2) where supplier sampling is permitted, discrepancies in the sample-to-lot relationship. These two types will be discussed in turn.

Measurement Discrepancies

Perhaps the most common source of measurement discrepancy occurring in practice is bias. Frequently bias is associated with test equipment calibration. Bias occurs when all measurements made have a constant or systematic error component. Thus, bias can be due to excessive passage of time or usage of equipment after a proper calibration or due to an improper calibration. The importance of a well defined and strictly enforced calibration program in preventing this type of discrepancy cannot be overemphasized. Bias may be caused, also, by operator error in consistently misreading an instrument error in the same manner, or by procedural error in applying the wrong voltage to an instrument, missetting a constant type control, or operating the equipment in an unsuitable environment. The important thing to recognize is that each reading recorded will be offset from the true value by a constant amount.

Complementary to bias error is precision error, which is a random component of error, the individual amounts of which cannot be predicted either as to direction or magnitude. As previously mentioned, all measurements are subject, more or less, to random variation, which is synonymous with precision error. Although exact values of this type error cannot be determined, as can bias error, it is entirely feasible to establish, for any given measurement system, the distribution of precision error. This is commonly done by error-of-measurement studies, wherein repeated measurement of calibration standards or identical units of product are made under suitably controlled experimental conditions. Precision error, if within the limits established by such error-of-measurement studies, cannot be considered as data discrepancies. It is simply a source of random variation which must be recognized and lived with by both parties. However, precision error of greater magnitude than that established and agreed upon can be a serious discrepancy. There are several possible causes of excessive precision error. The substitution, inadvertant or deliberate, of an inferior piece of inspection equipment would generally contribute to a loss of precision. Poor adjustment of equipment and poor operator training are other common causes.

Besides bias and precision errors, measurement discrepancies can exist due to the falsification of data. One of the most common versions of data falsification is called "flinching" and is frequently not at all deliberate. However it is a commonly recognized malady among some inspectors. Flinching is the tendency to make a consistent error in one direction when the result of a measurement is very close to a specification limit. For example, an inspector faced with an upper specification limit of 1.500 might consistently reject all units measuring 1.510 or greater, but usually accept marginally defective units measuring in the 1.500 to 1.509 range. More serious than flinching but probably less common is deliberate falsification. An example would be an operator who wished to save a great deal of time by inspecting only a small portion of an inspection batch, and, after determining the range of the data, recorded fake data on the balance of the batch. Falsification could also occur due to carelessness, where an operator simply made a gross error in reading or recording a measurement. The significant distinction between falsification and precision error is that while the latter is an inherent part of a particular inspection system and method, the former is completely avoidable, by greater attention or honesty.

Sampling Discrepancies

In contrast to measurement discrepancy is the problem of a discrepancy in the sample-to-lot relationship. An alternate descriptive term is a biased sample. A familiar example of this type of discrepancy is the basket of apples in which the top layers are composed only of the finest soundest fruit, while soft, rotten, green, or under-sized apples are down at the bottom of the basket. While such an arrangement could occur by coincidence it is more likely a deliberate attempt to deceive those who would inspect the basket only casually. Obviously, such a discrepancy is impossible in those cases where 100% inspection and reporting of a characteristic is required. It would also be impossible if the customer specifies, after the entire lot is completely processed and presented for final inspection, which units are to be inspected.

The general problem of biased samples has been tacitly recognized for a long time in the inspection and quality control profession, as is evidenced by the following quotation from a speech made by C. B. Dudley, a former president of the American Society for Testing Materials:

> ...it is not reasonable or proper or safe to trust the producers in anything by which the validity of the tests might be affected. Not once but many hundred times have we been asked to allow the shippers or producers to send a sample and accept the shipment on its examination. The request was undoubtedly made in good faith and with no other desire than to facilitate the transaction. Perhaps it is needless to say that our belief in the facility with which unintentional mistakes would be made and a sample better than the average of the shipment be sent, has always led us to positively refuse such requests.

One may read a tinge of sarcasm into this statement by Dudley, or take it at pure face value. In either event, his message is clear: do not allow the supplier to perform sample inspection in behalf of the customer under any circumstances. However, with proper safeguards and checks, we may find alternatives to his conservative approach.

Some distinct types of sampling discrepancies can be labeled, "salting," "tampering," and "truncation." The term "salting," meaning to enrich artificially, has its origin in connection with mining operations where a promoter would place a very rich specimen of ore someplace to mislead a prospector into believing he had found a valuable mining site. Thus, a procedure of placing known good material in an inspection sample, even though this product might be completely unrelated to the material which the sample purports to represent, can be labeled as salting the sample.

A distinctly different but similar scheme is to give special attention to the inspection sample in preparing it for inspection. In one version of this deception, which is labeled here as "tampering," a legitimate random sample of the lot could be selected, but the selection made while the material was still in processing. From that point on, the very best operators could be assigned to work on the sample units, and the most careful attention to detail could be given these units, as opposed to the balance of the lot. This would again result in an inspection sample whose characteristics were better than the lot as a whole.

A third type of sampling discrepancy will be called "truncation in sample selection" and defined as the deliberate exclusion from a sample, any units whose measured characteristic(s) exceeds a certain value (or values). A number of different mechanisms exists by which certain units in the lot might be deliberately excluded from the inspection sample, and as a result of which the sample would be an actually valid portion of the lot but yet would not be representative in the random sense. One way to achieve this sort of deception would be to pre-inspect a number of units just prior to the final quality inspection. Whenever a measurement appeared which exceeded a certain limit it would be set aside and another unit selected to take its place. The process has never, to the knowledge of the writer, been explicitly considered in the quality control literature.

The three types of discrepancies in the sample-to-lot relationship described above were referred to as deceptions. It should be pointed out that similar discrepancy could easily exist without any intent to deceive, if the supplier was careless in the method of selecting the inspection sample, and the group which was selected nonrandomly happened to be quite homogeneous within itself but not representative of the entire lot. Much emphasis in the literature of sampling is given to the point that great care is required to achieve a valid random sample. If the effort is not exerted, then probably sampling discrepancies will result.

To recapitulate, it is asserted that discrepancies in the supplier data can be divided into two mutually exclusive and exhaustive categories: measurement and sample-to-lot. Examples of each of these categories have been cited; while other examples can easily be constructed, it is apparent that any discrepancy in which the measured result of a sample is used to represent the characteristic of a lot can be placed in one category or the other.

Detection of Data Discrepancies

Corresponding to the dichotomy of data discrepancies is a similar dichotomy of the methods for detecting these discrepancies. We first consider the validity of the actual recorded measurements. If the data on specific measured units is invalid, this fact can be most directly ascertained by repeating the measurements on the same units and analyzing the paired differences. In contrast to repeat measurements would be the selection, by the customer, of an independent sample and comparing the customer sample results with either the entirety or with a selected sample of the supplier's data. However, the drawback of making an independent sample comparison to validate recorded measurements is that it introduces a source of random variation which must be tolerated and which will tend to obscure true discrepancies of this type. The point is discussed in most statistical texts. For instance, Hald (1952, pp. 401-405) pointed out the applicability and advantages of the paired-difference "Student t" test, and again (pp. 504-507) the advantages of the randomized complete block design in reducing the residual or error variance for testing the significance of some effect of interest. Cochran and Cox (1957, pp. 31-34, 112-114) have quantified the relative efficiency of the paired and independent samples, which are really equivalent to randomized complete block and completely randomized experimental designs, the blocks in this case being specific units of the product.

While the extensive work cited above will not be repeated here, it is worth emphasizing that, contrary to previous published statements on data validation (e.g., Department of Defense, 1960; Elner and Mandelson, 1964) it is not necessarily advantageous for the customer to insist on independent samples in validation. For example, if the customer is concerned _only_ with discrepancies in the recorded data, and if product variation tends to be large relative to the magnitude of discrepancy which is of concern, then clearly paired analysis on repeat measurements would be superior to the use of independent samples. On the other hand, certain factors tend to offset the advantage of paired samples for validating the recorded data. In the first place, the degrees of freedom for error variance is reduced by half. Further, there may be, in some cases, a significant incremental cost in identifying and re-inspecting identical units, above the cost of inspecting an independent random sample of equal size. However, in most practical applications these two factors are of small effect, in comparison with elimination of product variation, on the information obtained at a given cost.

The detection of the biased sample, on the other hand, _cannot_ be made through the means of repeat measurements. If the supplier did choose a sample in a biased fashion, repeating measurements on the same units would shed no light whatsoever on the relationship of these measurements to the balance of a lot or batch of material. Thus it is not only advisable but mandatory that the customer inspect units other than those reported on by the supplier _if_ he wishes to check for a biased sample. A completely independent random sample could be selected, recognizing that there might be some duplication of units into both samples. Or a mutually exclusive sample could be selected, in which no units of the supplier sample were allowed. For the purposes of this dissertation, it is assumed that sampling does not deplete the lot, and in this limiting case, the independent and mutually exclusive samples yield to the same analysis.

The customer may, in any given validation situation, be concerned about either biased sample discrepancy, recorded data discrepancy, or both. Our attention now focuses on the problem of the biased sample, and more specifically, on truncation in sample selection.

CHAPTER III

TRUNCATION IN SAMPLE SELECTION

Truncation in sample selection was defined in the previous chapter as the deliberate exclusion from the sample those units whose measurement(s) exceed a certain value. Consider two points, (a) and (b), on the measurement range. Exclusion of measurements greater than point (b) leads to one-sided truncation on the upper tail; exclusion of measurements less than a given value (a) leads to one-sided truncation on the lower tail. Exclusion of measurements less than (a) and greater than (b) leads to two-sided truncation.

For one-sided truncation, the value (a) or (b) might be any value within the possible range of measurements, depending only on the attitude of the person responsible for the exclusion. For that matter, it could be outside the range of the measurements, but this is a trivial case since it implies either no truncation, or complete truncation and no resultant sample. For two-sided truncation it is requisite that $b > a$ to avoid the trivial case, again, of no sample. The fraction of the lot which lies outside the truncation limits will be referred to as the degree of truncation and denoted by the symbol (γ), for one-sided truncation and by ($\gamma 2$) for two-sided truncation. This term is from Hald (1952) who described the degree of truncation as the proportion of an original population which is excluded from a truncated subpopulation. It is noted that the degree of truncation is a factor relating to the lot and not the sample directly.

Truncation and Biased Samples

Three types of biased-sample mechanisms were defined in Chapter II: salting, tampering, and truncation in sample selection. These are not the only types of biased samples which could be considered, but they do give an indication of the general nature of this type of sampling discrepancy. One distinction between truncation and the other two biasing mechanisms is the relation between the lot being presented for acceptance and the population being sampled. Salted and tampered samples are from populations which may be quite different from the lot. But truncated samples are clearly from the lot itself, albeit only a segment of the lot. This fact leads to a unique feature of truncation, from the point of view of statistical analysis. That is, the truncated population from which the customer's sample is drawn can be defined entirely in terms of the parent population and the point(s) of truncation.

As another distinction, it appears that truncation would be much easier to accomplish than the other two forms of biasing. No alteration of any unit is required, as in the case of tampering; and no separate production or procurement is required, as in the case of salting. All that is required is an inspection and sorting operation; thus, it would be a relatively more tempting form of deception to practice.

Reasons for Truncation in Sample Selection

There are several reasons why a supplier might, either at the policy level or operating level, be motivated to truncate a lot in connection with selection of an acceptance sample. The most obvious reason is that a specification limit exists which would cause any measurements above a certain value to be classified as defective. For instance, the supplier might have the opinion that a certain specification limit is arbitrary rather than functional, and that product which measures in excess of that limit will nevertheless function properly. If such a judgment is correct, then little would be lost by arbitrarily ignoring units which measure greater than an arbitrary limit, as long as this is not discovered by the customer. The supplier's reasoning is that after the product has been sold and put to use, the product, although nonconforming, will provide satisfaction to the customer nevertheless, and will not contribute to loss of customer good will or later rejection. Of course, a short-sighted supplier might choose to truncate his sample at a specification limit without any consideration of the consequences in further usage of the product.

The truncation point may not be a specification limit. A certain type of lot sampling plan requires the computation of a variables acceptance limit, e.g., accept only if $\bar{x} + ks \leq USL$ where k and USL are pre-specified constants, and $\bar{x}$ and s are the sample mean and sample standard deviation. A supplier whose product is to be purchased on the basis of such a criterion may be able to markedly improve the chances of lot acceptance by truncation of extremely high readings. This would have the dual effect of reducing both $\bar{x}$ and s. Clark (1957) computed the degree of truncation required to have the desired effect for specified lot distribution parameters and acceptance criteria. It should be noted that Clark was not advocating truncation in sample selection. He proposed screening of the entire lot in order to alter its parameters. The important distinction between lot screening and sample truncation is that in the former case the units which have measurements in excess of a certain value are removed from the lot itself, but in the latter case they are removed only if they appear in the sample.

A third possible reason for truncation is the desire to establish, for the benefit of future contracts or other applications, that the characteristics of a lot are of a higher quality level than they actually are. If process changes were under development which were expected to improve the product distribution, there could be some tendency to attempt to provide evidence of the desired results through selective inspection.

Thus, it is not possible to simply look at a predetermined number such as a specification limit, in deciding whether truncation in sample selection is present. However, a specification limit is probably the most likely truncation point.

The Normal Distribution Model for Truncation Analysis

The truncated normal distribution was chosen as the model for this study. Two versions of truncated normal distributions and their relationships to the parent normal distribution are illustrated in Figure 1 and Figure 2. One-sided truncation is shown by Figure 1. It is apparent that the mean is shifted and the standard deviation reduced if a normal distribution is truncated at one end. Two-sided symmetrical

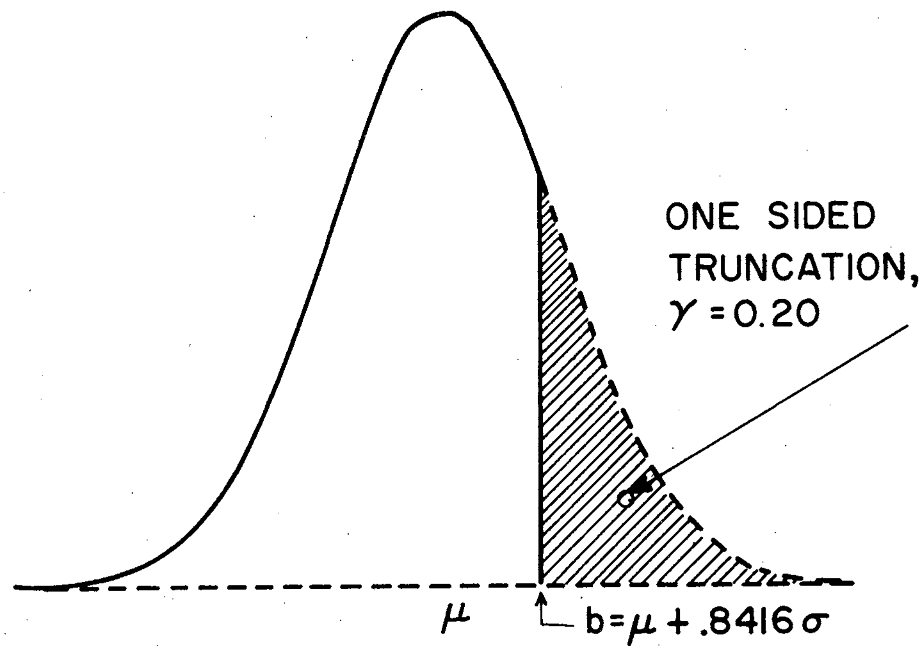

Figure 1. One-sided Truncated Normal Distribution

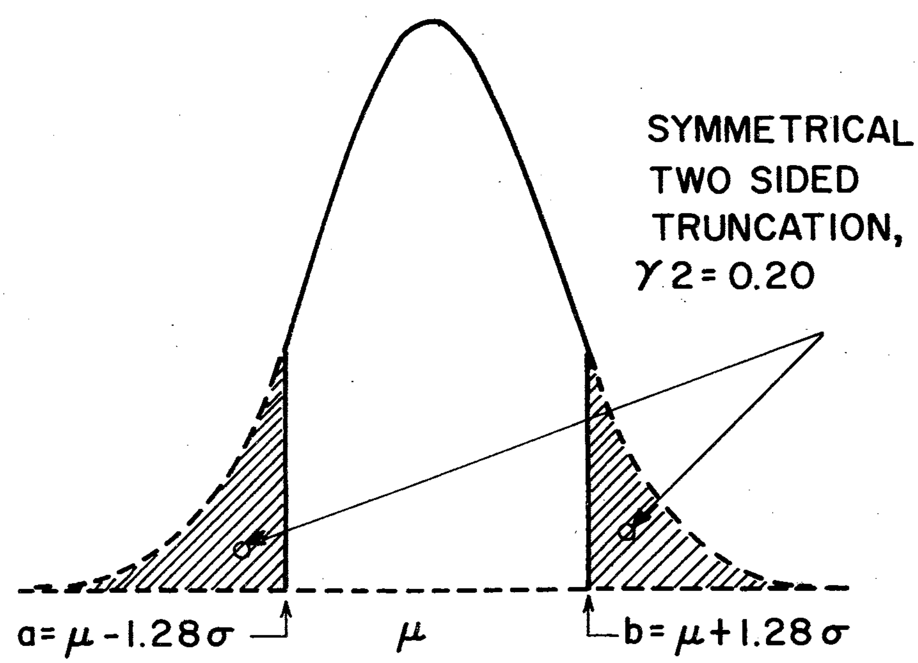

Figure 2. Symmetrical Two-sided Truncated Normal Distribution

truncation is shown in Figure 2. The standard deviation is reduced to a much greater relative extent (compared with one-sided truncation) but the mean is not shifted.

The truncated normal distribution can be defined by four parameters: μ, the parent population mean; σ^2, the parent population standard deviation, and a and b, the two points of truncation. Letting the variable x represent a truncated normal random variable, the density function of x is

$$g(x) = \frac{f(y)}{F} \quad , \quad a < x < b$$

$$= 0 \qquad , \text{ elsewhere,}$$

where

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

and

$$F = \int_a^b f(y)\, dy.$$

Following notation used by Hald (1952) and elsewhere, we define for any number c,

$$\phi(c) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2}$$

and

$$\Phi(c) = \int_{-\infty}^{c} \phi(t)\, dt.$$

Thus $\Phi(c)$ is the normal distribution function of c and $\phi(c)$ is the ordinate of the standard normal density function of c, and

$$F = \Phi(b) - \Phi(a)$$

is the constant factor which makes g(x) a density function.

It will be necessary in the following chapter to know the mean, μ_x, and standard deviation, σ_x, of the truncated normal distribution in terms of the parameters μ, σ, a, and b. These parameters are now derived for the case of general two-sided truncation and then specialized to the cases of symmetrical two-sided truncation and one-sided truncation.

To derive the mean of the truncated normal distribution, we apply the definition

$$\mu_x = E(x) = \int_{-\infty}^{\infty} xf(x)dx$$

which is

$$\mu_x = \int_a^b \{[\ \Phi(b) - \Phi(a)\ ]\ \ \sigma\sqrt{2\pi}\ \}^{-1} xe^{-(x-\mu)^2/2\sigma^2}dx$$

$$= \{\ F\ \sigma\sqrt{2\pi}\ \}^{-1} \int_a^b xe^{-(x-\mu)^2/2\sigma^2}dx.$$

This expression can be separated into two parts,

$$\mu_x = \frac{\sigma\int_a^b (\frac{x-\mu}{\sigma}) e^{-(x-\mu)^2/2\sigma^2}\ dx}{F\ \sigma\sqrt{2\pi}} + \frac{\mu\ \int_a^b\ e^{-(x-\mu)^2/2\sigma^2}dx}{F\ \sigma\sqrt{2\pi}}$$

$$= \frac{\sigma[-e^{-(x-\mu)^2/2\sigma^2}\ \Big|_a^b]}{F\ \sigma\sqrt{2\pi}} + \mu\cdot 1$$

$$= \sigma\ \frac{-\frac{e^{-(b-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} + \frac{e^{-(a-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}}}{F\sigma} + \mu$$

$$= \mu - \sigma\ \frac{\phi(b) - \phi(a)}{\Phi(b) - \Phi(a)} \qquad (3\text{-}1)$$

We now derive $E(x^2)$, and σ_x^2 follows from it.

$$E(x^2) = \frac{1}{F\sqrt{2\pi\gamma}} \int_a^b\ x^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx$$

$$= \frac{1}{F\sqrt{2\pi\sigma}} \int_a^b [(x-\mu)^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} + 2x\mu e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} - \mu^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}]\ dx$$

$$= \frac{1}{F\sqrt{2\pi\sigma}} [\ A \quad + \quad B \quad - \quad C\ ]$$

Where each term denoted by A, B, C, respectively, is to be integrated.

$$A = \frac{\sigma^2}{\sigma\sqrt{2\pi}F} \int_a^b (\frac{x-\mu}{\sigma})^2\ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

$$= \frac{\sigma^2}{\sqrt{2\pi}\ F} \int_a^b (\frac{x-\mu}{\sigma})\ (\frac{x-\mu}{\sigma})\ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\ \frac{dx}{\sigma}$$

Integrating by parts, let

$$u = \frac{x-\mu}{\sigma} \qquad dv = (\frac{x-\mu}{\sigma})\ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} (\frac{dx}{\sigma})$$

$$du = \frac{dx}{\sigma} \qquad v = -e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} (\frac{1}{\sigma})$$

Then

$$A = \frac{\sigma^2}{\sigma\sqrt{2\pi}\ F} \{-[\frac{x-\mu}{\sigma}]\ [\ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\ ]\ \Big|_a^b + \int e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx\}$$

$$= \sigma^2\ [\ (\frac{a-\mu}{\sigma})\ \phi(a) - (\frac{b-\mu}{\sigma})\ \phi(b)\ ]\ /\ F + \sigma^2 \int_a^b f(x)\ dx$$

$$= \sigma^2 [\ 1 + \frac{(\frac{a-\mu}{\sigma})\ \phi(a) - (\frac{b-\mu}{\sigma})\ \phi(b)}{F}\ ]$$

This completes the evaluation of term A.

$$B = \frac{2\mu}{\sigma\sqrt{2\pi}\ F} \int_a^b x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

$$= 2\mu \left[\frac{1}{\sigma\sqrt{2\pi}\ F} \int_a^b x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx\right]$$

$$= 2\mu \left[\mu - \sigma \frac{\phi(b) - \phi(a)}{F}\right]$$

$$= 2\mu^2 - 2\mu\sigma \frac{\phi(b) - \phi(a)}{F}$$

This completes the evaluation of term B.

$$C = \frac{1}{\sigma\sqrt{2\pi}\ F} \int_a^b \mu^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \mu^2 \int_a^b f(x)\ dx = \mu^2$$

This completes the evaluation of term C.

Combining these three terms we have

$$E(x^2) = \sigma^2\left[1 + \frac{\left(\frac{a-\mu}{\sigma}\right)\phi(a) - \left(\frac{b-\mu}{\sigma}\right)\phi(b)}{F}\right] + \mu^2 - 2\mu\sigma \frac{\phi(b) - \phi(a)}{F}$$

Now since $\sigma_x^2 = E(x^2) - \mu_x^2$

and $\mu_x^2 = \mu^2 - 2\mu\sigma \frac{\phi(b) - \phi(a)}{F} + \left[\sigma \frac{\phi(b) - \phi(a)}{F}\right]^2$,

it follows with some algebraic simplification that

$$\sigma_x^2 = \sigma^2 \left\{1 + \frac{\left(\frac{a-\mu}{\sigma}\right)\phi(a) - \left(\frac{b-\mu}{\sigma}\right)\phi(b)}{\Phi(b) - \Phi(a)} - \left[\frac{\phi(b) - \phi(a)}{\Phi(b) - \Phi(a)}\right]^2\right\} \tag{3-2}$$

There are two special cases of the truncated normal distribution considered in more detail herein. In one case, where $a = -\infty$, we have one-sided truncation on the upper tail. In the other case, where a $a = 2\mu - b$, we have symmetrical two-sided truncation.

In the case of one-sided truncation on the upper tail, we have, for $a = -\infty$, $\phi(a) = \Phi(a) = 0$ and thus, from equation (3-1)

$$\mu_x = \mu - \sigma \frac{\phi(b)}{\Phi(b)} \tag{3-3}$$

and likewise

$$\sigma_x^2 = \sigma^2 \{1 - \frac{(\frac{b-\mu}{\sigma})\ \phi(b)}{\Phi(b)} - [\frac{\phi(b)}{\Phi(b)}]^2\} \tag{3-4}$$

In the case of symmetrical two-sided truncation, we have

$$a = \mu - (b-\mu) = 2\mu - b$$

$$a - \mu = \mu - b$$

$$\phi(a) = \phi(b)$$

$$\Phi(a) = 1 - \Phi(b)$$

When these equations are substituted into equation (3-1) we find

$$\mu_x = \mu$$

and from equation (3-2),

$$\sigma_{x_2}^{\ 2} = \sigma^2 \{ 1 - \frac{[(\frac{\mu-b}{\sigma}) - (\frac{b-\mu}{\sigma})]\ [\phi(b)]}{\Phi(b) - 1 + \Phi(b)} - 0\}$$

$$= \sigma^2 \{ 1 - \frac{2(\frac{b-\mu}{\sigma})\ \phi(b)}{2\Phi(b) - 1} \}$$

$$= \sigma^2 \{ 1 - \frac{(\frac{b-\mu}{\sigma})\ \phi(b)}{\Phi(b) - \frac{1}{2}} \} \tag{3-5}$$

where $\sigma_{x_2}^{\ 2}$ is used to emphasize the fact that truncation in this case is symmetrical and two-sided.

The final constant associated with the truncated normal distribution is the degree of truncation, which was defined as the fraction of the parent population excluded by the truncation process. Letting γ represent degree of truncation it is clear that in general

$$\gamma = \Phi(a) + 1 - \Phi(b)$$

while for one-sided truncation on the upper tail,

$$\gamma = 1 - \Phi(b) \qquad (3\text{-}6)$$

and for two-sided symmetrical truncation,

$$\gamma 2 = 1 - \Phi(b) + 1 - \Phi(b) = 2 - 2\Phi(b) \qquad (3\text{-}7)$$

The mean and variance μ_x and σ_x^2 for a truncated normal distribution can be standardized by considering their relationship with the parent population and their parameters μ and σ^2. Thus for the case of one-sided truncation on the upper tail we define

$$k = \frac{b-\mu}{\sigma}$$

$$\lambda = \frac{\mu-\mu_x}{\sigma}$$

and

$$\theta = \frac{\sigma_x^2}{\sigma^2} \quad .$$

Then if $\gamma = 0$, $\mu = \mu_x$, $\sigma^2 = \sigma_x^2$, and the standardized parameters λ and θ equal 0 and 1, respectively. This is depicted in Table I, which shows that as γ approaches 1, or complete truncation, μ_x approaches $-\infty$, and θ approaches 0.

For the case of symmetrical two-sided truncation, we define

$$k2 = \frac{b-\mu}{\sigma}$$

$$\lambda = \frac{\mu-\mu_x}{\sigma}$$

and

$$\theta 2 = \frac{\sigma_{x2}^{\;2}}{\sigma^2} \quad .$$

The relationships for this case are depicted in Table II. It is noted that λ is constant at 0 regardless of the value of γ, but that $\theta 2$ decreases from 1 to 0 as γ increases from 0 to 1.

TABLE I

PARAMETERS, ONE-SIDED TRUNCATED NORMAL DISTRIBUTION

Degree of Truncation	Truncation Point	Mean	Variance	Standard Deviation
γ	k	λ	θ	$\sqrt{\theta}$
0.0010	3.09023	-0.003370	0.989576	0.994774
0.0050	2.57583	-0.014533	0.962354	0.980997
0.0100	2.32635	-0.026922	0.936646	0.967805
0.0200	2.05375	-0.049406	0.896092	0.946621
0.0400	1.75069	-0.089763	0.834795	0.913671
0.0500	1.64485	-0.108564	0.809641	0.899801
0.0600	1.55477	-0.126728	0.786907	0.887078
0.0800	1.40507	-0.161596	0.746834	0.864195
0.1000	1.28155	-0.195000	0.712073	0.843844
0.1500	1.03643	-0.274305	0.640459	0.800287
0.2000	0.84162	-0.349949	0.583011	0.763551
0.2500	0.67449	-0.423700	0.534697	0.731230
0.3000	0.52440	-0.496706	0.492810	0.702004
0.3500	0.38532	-0.569851	0.455695	0.675052
0.4000	0.25335	-0.643904	0.422254	0.649811
0.4500	0.12566	-0.719636	0.391694	0.625855
0.5000	0.00000	-0.797884	0.363381	0.602811
0.5500	-0.12566	-0.879579	0.336869	0.580404
0.6000	-0.25335	-0.965851	0.311831	0.558418
0.6500	-0.38532	-1.058255	0.287863	0.536529
1.0000	$-\infty$	$-\infty$	0	0

TABLE II

PARAMETERS, TWO-SIDED TRUNCATED NORMAL DISTRIBUTION

Degree of Truncation	Truncation Point	Mean	Variance	Standard Deviation
γ_2	k_2	λ	θ^2	$\sqrt{\theta^2}$
0.00100	3.29050	0	0.98830	0.99413
0.00500	2.80700	0	0.95621	0.97786
0.01000	2.57580	0	0.92475	0.96164
0.02000	2.32630	0	0.87345	0.93459
0.04000	2.05370	0	0.79282	0.89041
0.04000	1.96000	0	0.75886	0.87113
0.06000	1.88070	0	0.72768	0.85304
0.08000	1.75100	0	0.67218	0.81986
0.10000	1.64480	0	0.62299	0.78930
0.15000	1.44050	0	0.52102	0.72182
0.20000	1.28160	0	0.43774	0.66162
0.25000	1.15050	0	0.36860	0.60713
0.30000	1.03640	0	0.30955	0.55638
0.35000	0.93460	0	0.25875	0.50867
0.40000	0.84160	0	0.21460	0.46325
0.45000	0.75530	0	0.17612	0.41967
0.50000	0.67450	0	0.14267	0.37771
0.55000	0.59790	0	0.11358	0.33701
0.60000	0.52440	0	0.08833	0.29720
0.65000	0.45390	0	0.06677	0.25840
1.00000	10.00000	0	0	0

It should be noted that the degree of truncation, γ, is the fraction of a lot which is excluded from the chance of being selected as part of a random sample. Thus the sample which is inspected is actually a random sample from the truncation distribution. This is the basis for the expression, truncation in sample selection, to indicate the type of sampling discrepancy being considered. In the context of this dissertation the degree of truncation is identically equal to the fraction of a lot excluded from the chance of being part of a sample by truncation in sample selection.

There are two main reasons for the choice of the normal distribution to study the effect of truncation. First, the normal distribution approximates many of the actual distributions of measured characteristics in manufacturing. Also, in a number of cases, appropriate transformations can be utilized to achieve normality in analyzing non-normal product distributions. The second reason for its choice is the great extent to which statistical test power curves have been developed in terms of the normal distribution. Thus the use of the normal model provides a bench mark or frame of reference for comparing tests to a common criteria, even if it is known that given product distributions are not truly normal. Similarly, it provides for a common basis for comparing tests, such as the F and U tests, which are derived on the assumption of normality, with others, such as the exceedance, maximum difference, and rank-sum tests (all described in the following chapter) which are distribution-free.

Statistical Inference on the Truncated Normal Distribution

An early study of the properties of the truncated normal distribution was by Stevens (1937) who treated the case where the points of truncation are known and the number of individuals of a sample in the truncated portion is also known. (This is sometimes referred to as a censored distribution inasmuch as the number of observations but not the values in the truncated portion is known.) Stevens derived the likelihood function, the maximum likelihood estimates of μ and σ^2, and the variance-covariance matrix of these estimators. Cohen (1950) extended these results to cover other special cases of parent population parameter estimation, all with truncation points known. (Cohen also cited the work of several other researchers not mentioned here.) Hald (1952) discussed the problems associated with estimating the three parameters μ, σ^2, and b, of the one-sided truncated normal distribution, observing that "estimation of these parameters...is very laborious." He pointed out (p. 146) that the point of truncation of an observed distribution could be graphically estimated by plotting the cumulative distribution on normal probability paper and estimating the point on the abscissa to which the estimated line is asymptotic. He stated "in order to determine a truncation less than 10-20%, the number of observations must be very large, as otherwise, for small values of x, the deviation of the fractiles from the straight line will fall within the permissable limits of random variation." Hald was discussing one-sample inference. The results of the research reported herein corroborates his conclusion for the two-sample case.

A procedure for estimation of a truncation point where the parent distribution is completely specified was developed by Robson and Whitlock (1964), who derived approximate point and interval estimates for the one-sample case. They used the largest and second largest order statistics to estimate the point of truncation on the upper end of the distribution. Their results would hold equally for the case of truncation on the lower end of the distribution, using the smallest order statistics. The approximate $100(1 - \alpha)\%$ upper confidence limit on the truncation point is

$$X_{(n)} + \left(\frac{1-\alpha}{\alpha}\right) (X_{(n)} - X_{(n-1)})$$

where the subscript (n) is used to represent the largest order statistic, and (n-1) the next largest.

This estimate is an exact limit only for the case of the uniform distribution, but is claimed to be "approximately" valid in general. The procedure is applicable only to the one-sample case.

The primary limitation on the method of Robson and Whitlock is that the parameters μ and σ^2 of the parent distribution must be known in order to make an inference about the truncation point. But in the case of truncation in sample selection it is generally not known even that truncation exists and although there may be a good guess as to the point of truncation this too must frequently be considered as unknown.

Of more importance however, than estimating the truncation point is testing the hypothesis that truncation exists, which is discussed in the following chapter.

CHAPTER IV

STATISTICAL TESTS FOR TRUNCATION

It has been shown that one form of data discrepancy in supplier inspection data is sample biasing, and that truncation in sample selection is a reasonably logical and simple way in which sample biasing could be effected. Now it must be recognized that several alternatives are open to a customer in protecting himself against such a discrepancy. One form of protection would be to simply not allow supplier sampling at all, either by requiring 100% supplier inspection or by the customer performing the sampling inspection himself. Another approach to protection would be to deal only with suppliers whose reputation for consistent high quality and integrity were unquestioned, and thus could be always trusted to select samples in a scrupulously random manner. Or, as previously suggested, the customer could assume responsibility for selection of the supplier sample.

In any of the above situations the customer's data validation program can be restricted solely to the analysis of measurement discrepancies. However, there may well be situations where a customer is presented with the results of a supplier's sample and must make a determination of whether the sample is random or not. In such situations a comparison of the results of the supplier sample and an independent sample drawn by the customer can be performed.

Several possible statistical tests which could be used to make such a comparison will be discussed in this and the following chapter. These are not all the tests which might be used for such a purpose. From the large number available several were selected on the basis of their degree of acceptance as tests in similar situations. Others were chosen because it appeared on the surface that they might exhibit desirable properties.

For each of the tests discussed, a brief historical summary is presented, followed by the method of computing the test statistic, and then the derivation or a discussion of the operating characteristic function and significance probability of the test.

The significance probability, that is, the probability of rejecting a null hypothesis when it is true, is denoted by the symbol α. It is also called the probability of Type I error. For all tests discussed herein, the source of derivation of the significance probability is cited and selected values are presented in the charts. Where it is appropriate, a derivation of the significance probability is presented.

Of all the aspects of the tests which might be considered, the one which is of the greatest importance in this analysis is the operating characteristic function of the test against the truncation alternative. The operating characteristic function has been defined (e.g., Cowden, 1957) as the "mathematical expression which states the probability of accepting a lot as a function of the fraction defective in the lot." In applications described herein it is the probability of deciding there is no truncation in sample selection by the supplier as a function of the actual degree of truncation. In general it states the probability of accepting a null hypothesis as a function of a specific alternative hypothesis. Commonly called the "O.C. curve" it is the complement of the power function, which states the probability of rejecting a null hypothesis. Thus the O.C. curve is the "power curve turned upside down." The symbol β represents the probability of accepting a null hypothesis and the expression β( ) means the probability of accepting a null hypothesis given the condition expressed inside the parentheses. The use of both O.C. curves and power curves is common; the former are used in this study to be consistent with prevalent quality control practice, where characteristics of sampling plans are generally described in part by displaying their O.C. curves.

For each O.C. curve the degree of truncation is the independent variable and the probability of accepting "no truncation" is the dependent variable, while the significance probability and the size of both supplier sample and customer sample are constants.

The U Test for One-Sided Truncation

One of the most common statistical hypothesis tests is the U test, which is the optimum test for the difference between two means when the distributions are normal and the variances are known. This test is also referred to by some as the normal test. The assumption that the variances be known is somewhat restrictive if small samples are involved. But if sample size is of the order of 30 or above it has been stated (e.g., Hald, p. 389) that there is little practical difference between the U test and the comparable Student's test, which is the optimum test if the distributions have equal but unknown variances.

The U statistic for the two-sample case is the difference between the two sample means, divided by the standard error of the difference. That is,

$$U = \frac{\bar{y} - \bar{x}}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}}$$

where

$\bar{y}$ is the mean of the supplier sample

$\bar{x}$ is the mean of the customer sample

σ^2 is the population variance

m is the number of observations in the supplier sample

n is the number of observations in the customer sample.

In the derivation of the O.C. curve for the U test of truncation, we make use of two theorems, stated here without proof, which are both standard results in statistics. The first is the Central Limit Theorem, a version of which states that if $x_1, x_2, \ldots, x_n$ are stochastically independent random variables with mean μ_x and variance σ_x^2 then $w = \Sigma_{i=1}^{n} x_i/n$ tends to become normally distributed with mean μ_x and variance σ_x^2/n as n approaches ∞. The second is the addition theorem for normally distributed variables which states that if $v_1, v_2, \ldots, v_n$ are stochastically independent normally distributed random variables, with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, the function

$$w = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n$$

will also be normally distributed with mean

$$\mu_w = a_1 \mu_1 + a_2 \mu_2 + \ldots + a_n \mu_n$$

and variance

$$\sigma_w^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + a_n^2 \sigma_n^2 .$$

An approximate distribution of $\bar{x}$, the mean of the customer sample, is now derived. It is specified that the supplier selects a random sample of size n from a truncated normal distribution, with parameters μ, σ^2, and truncation point b on the upper tail. Then from the results in the previous chapter on the truncated normal distribution, it is seen that the mean and variance of the population being sampled by the supplier are

$$\mu_x = \mu - \sigma \frac{\phi(k)}{\Phi(k)}$$

and

$$\sigma_x{}^2 = \sigma^2[\ 1 - k\,\frac{\phi(k)}{\Phi(k)} - (\frac{\phi(k)}{\Phi(k)})^2]$$

where

$$k = \frac{b-\mu}{\sigma}$$

From the Central Limit Theorem cited above it is seen that $\bar{x}$ is approximately normally distributed with mean μ_x and variance $\sigma_x{}^2/n$. The error of the approximation involved here is an increasing function of the degree of truncation and a decreasing function of sample size. (The magnitude of this error has been examined by means of an empirical distribution sampling study, the results of which are presented in Appendix B.)

The customer selects his sample of size m from the untruncated parent population and the observations in the sample are designated $y_1, y_2, \ldots, y_m$. By the addition theorem for the normal distribution, letting $a_1 = \frac{1}{m}$, $a_2 = \frac{1}{m}$, ... , and $v_1 = y_1$, $v_2 = y_2$, ... , we find that $w = \frac{1}{m} \sum_{i=1}^{m} y_i = y$. Each y observation is randomly selected from a normal distribution with parameters mean μ and variance σ^2. Thus, $\bar{y}$ has a normal distribution with mean

$$\mu_{\bar{y}} = \frac{1}{m}\mu + \frac{1}{m}\mu + \ldots + \frac{1}{m}\mu = \mu$$

and variance

$$\sigma_{\bar{y}}^2 = \frac{1}{m^2}\sigma^2 + \frac{1}{m^2}\sigma^2 + \ldots + \frac{1}{m^2}\sigma^2 = \frac{m}{m^2}\sigma^2 = \frac{\sigma^2}{m}$$

These results are exact, as opposed to the approximate results for the supplier sample mean.

Calling upon the addition theorem again, we see that the difference between sample means,

$$w = \bar{y} - \bar{x} \qquad (4\text{-}1)$$

has a normal distribution with mean

$$\mu_w = (1) \cdot \mu_{\bar{y}} + (-1)\mu_{\bar{x}} = \mu - \mu_x \qquad (4\text{-}2)$$

and variance

$$\sigma_w^2 = (1)^2\sigma_{\bar{y}}^2 + (-1)^2\sigma_{\bar{x}}^2 = \frac{\sigma^2}{m} + \frac{\sigma_x^2}{n} \qquad (4\text{-}3)$$

Given this normally distributed random variable w, a new random variable can be derived by subtracting μ_w from w and dividing this difference by σ_w. Let this random variable be Z. Then

$$Z = \frac{w-\mu_w}{\sigma_w} \qquad (4\text{-}4)$$

and by the addition theorem we find

$$\mu_Z = E(\frac{w}{\sigma_w}) - E(\frac{\mu_w}{\sigma_w}) = \frac{\mu_w}{\sigma_w} - \frac{\mu_w}{\sigma_w} = 0$$

$$\sigma_Z^2 = VAR(\frac{w-\mu_w}{\sigma_w}) = VAR(\frac{w}{\sigma_w}) + VAR(\frac{\mu_w}{\sigma_w})$$

$$= (\frac{1}{\sigma_w})^2\sigma_w^2 + (\frac{1}{\sigma_w})^2(0) = 1$$

Z is thus normally distributed with mean 0 and variance 1. Substituting into equation (4-4) the results of (4-1), (4-2), and (4-3) we have

$$Z = \frac{(\bar{y}-\bar{x}) - (\mu-\mu_x)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma_x{}^2}{n}}} \qquad (4\text{-}5)$$

If the null hypothesis of no truncation is true, i.e. $\gamma=0$, then $\mu=\mu_x$, $\sigma^2=\sigma_x{}^2$, and the above equation reduces to

$$Z = \frac{\bar{y} - \bar{x}}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} = \frac{\bar{y} - \bar{x}}{\sigma\sqrt{\frac{m+n}{mn}}} \qquad (4\text{-}6)$$

It is noted that all elements on the right hand side of this expression are computed from the samples or are known constants. Thus we always can compute this value, whether or not the null hypothesis is true. Therefore, we redesignate it as U. Let

$$U = \frac{\bar{y} - \bar{x}}{\sigma\sqrt{\frac{m+n}{mn}}} \qquad (4\text{-}7)$$

and note that U is a sample statistic, whose distribution is that of Z if H_o is true, and whose distribution remains to be found if H_o is not true. We must now find this distribution.

Whereas the above derivation of the significance probability is presented in various forms in the literature, the following derivation of the O.C. curve is believed to be original with the writer. Let Z_α be a point on the unit normal distribution such that $Pr(U > Z_\alpha) = \alpha$. Then the value α is the significance probability of the test and the region on the real line which satisfies the inequality $[U > Z_\alpha]$ is called the critical region.

The variable U is distributed normal, zero mean, unit variance, only if $\gamma = 0$. If $\gamma > 0$ the expected value of $v = \bar{y} - \bar{x}$ will be greater than zero since $\mu > \mu_x$, and its variance will be less than one, since $\sigma_x^2 < \sigma^2$. Thus the effect of truncation on the upper tail is to shift the distribution of the test statistic U to the right and to narrow slightly the spread of the distribution. To derive the O.C. curve for the test, we need only to find the probability that the computed value U will fall outside the critical region, for any specified value γ_o, of

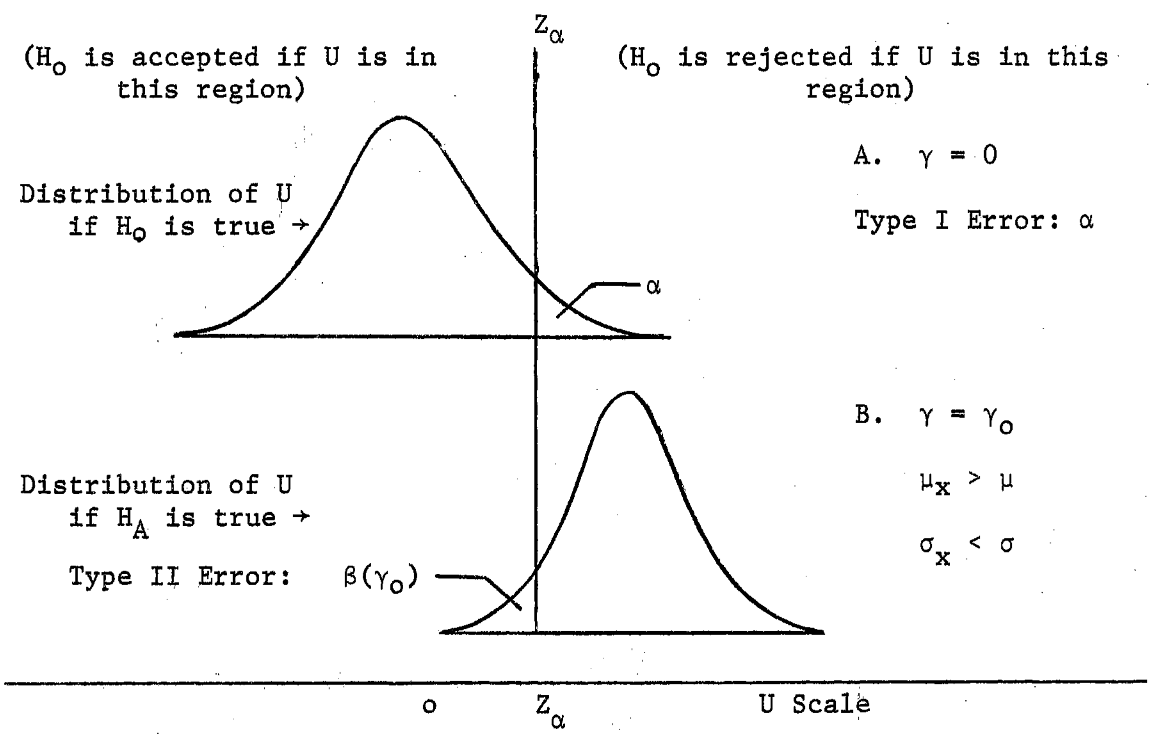

Figure 3. Type I and Type II Errors of U Test

γ greater than zero. These relationships among α, Z_α, γ, and $\beta(\gamma_o)$ are depicted in Figure 3. In part A the condition which leads to Type I error with probability α is depicted: γ is equal to zero, and the distribution of U is centered at zero. In part B, γ takes on some positive value γ_o, which causes the distribution of U to shift to the right and to become somewhat more narrowly dispersed. Type II error has probability β of occurring.

The O.C. function of the U test is

$$\beta(\gamma_o) = \Pr(\, U \le Z_\alpha \mid \gamma = \gamma_o)$$

since the null hypothesis is accepted, regardless of the actual value of γ, whenever $U \le Z_\alpha$. To find $\beta(\gamma_o)$ it is necessary to find the distribution of U given γ_o. Although the truncation will actually reduce the spread of the distribution of U this is not taken into account by the computation of U in which only the values $\bar{x}$ and $\bar{y}$ are computed from the sample. The variance is still assumed to be σ^2. Thus the expected value of the computed U statistic, is again by the addition theorem

$$\mu_u = E\left(\frac{\bar{y} - \bar{x}}{\sigma\sqrt{\frac{m+n}{mn}}}\right) = \frac{\mu-\mu_x}{\sigma\sqrt{\frac{m+n}{mn}}} = \frac{(\mu-\mu_x)/\sigma}{\sqrt{\frac{m+n}{mn}}} = \frac{\lambda}{\sqrt{\frac{m+n}{mn}}} \qquad (4\text{-}8)$$

and the variance is

$$\sigma_u = VAR\left(\frac{\bar{y} - \bar{x}}{\sigma\sqrt{\frac{m+n}{mn}}}\right) = VAR\left[\frac{1}{\sigma\sqrt{\frac{m+n}{mn}}}(\bar{y}) - \frac{1}{\sigma\sqrt{\frac{m+n}{mn}}}(\bar{x})\right]$$

$$= \frac{1}{\sigma^2\left(\frac{m+n}{mn}\right)} VAR(\bar{y}) + \frac{1}{\sigma^2\left(\frac{m+n}{mn}\right)} VAR(\bar{x})$$

$$= \frac{1}{\sigma^2\left(\frac{m+n}{mn}\right)} \cdot \frac{\sigma^2}{m} \quad \frac{1}{\sigma^2\left(\frac{m+n}{mn}\right)} \cdot \frac{\sigma_x{}^2}{n}$$

$$= \frac{\frac{\sigma^2}{m} + \frac{\sigma_x^2}{n}}{\sigma^2 (\frac{m+n}{mn})} = \frac{\frac{1}{m} + \frac{\sigma_x^2}{\sigma^2 n}}{\frac{m+n}{mn}} = \frac{n + (\sigma_x^2/\sigma^2)m}{m + n} = \frac{n + \theta m}{n + m} \qquad (4\text{-}9)$$

Therefore the random variable U is distributed normally with mean and variance as indicated in equations (4-8) and (4-9).

To convert this result into the desired probability statement we recall that for a given value of γ,

$$\beta = \Pr(U \leq Z_\alpha).$$

The inequality is unaffected by subtracting and dividing by constants. So

$$\beta = \Pr\left(\frac{U - \mu_u}{\sigma_u} \leq \frac{Z_\alpha - \mu_u}{\sigma_u}\right)$$

The left hand side of this inequality now meets the requirements of the Z random variable, normal with zero mean and unit variance. This gives

$$\beta = \Pr\left(Z \leq \frac{Z_\alpha - \mu_u}{\sigma_u}\right)$$

Now from the definition of the cumulative normal distribution function of a fixed value v,

$$\Pr(Z \leq v) = \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(v)$$

So it follows that

$$\beta = \Phi(v) \qquad (4\text{-}10)$$

where

$$v = \frac{Z_\alpha - \mu_u}{\sigma_u}$$

The expressions (4-8) and (4-9) developed for μ_u and σ_u can now be substituted into the equation for v.

$$v = \left(Z_\alpha - \frac{\lambda}{\sqrt{\frac{m+n}{mn}}}\right) \Big/ \sqrt{\frac{n + \theta m}{m+n}} \qquad (4\text{-}11)$$

In the case where m = n equation (4-11) can be simplified further to

$$v = \left(Z_\alpha - \frac{\lambda}{\sqrt{\frac{2m}{m^2}}}\right) / \sqrt{\frac{m(1+\theta)}{2m}}$$

$$= \left(Z_\alpha - \frac{\lambda}{\sqrt{2/m}}\right) \cdot \sqrt{\frac{2}{1+\theta}} \qquad (4\text{-}12)$$

To compute the O.C. curve for the U test, either expression (4-11) or (4-12) for v can be solved by specifying α, m, n, and γ. Then the cumulative unit normal distribution function $\Phi(v)$ can be found by the use of available tables (or series approximations, in the case of digital computer analysis). This result, by equation (4-10), is the desired probability.

To illustrate the computation of one point on the O.C. curve of the U test, consider a case where both the supplier and customer inspect samples of size 49. The customer's sample is drawn at random from a population with normal distribution, $\mu = 100$, $\sigma = 20$. The supplier's sample is drawn from a truncated portion of the same distribution, with exclusion of any units measuring above the truncation point b = 135. Then

$$k = \frac{b-\mu}{\sigma} = \frac{135-100}{20} = 0.675$$

and from Table I it is seen that $\gamma = 0.25$, $\lambda = 0.4237$, and $\theta = 0.5347$. Let the test be performed at $\alpha = 0.05$. Applying equation (4-12) gives

$$v = \left(1.645 - 0.4237 / \sqrt{\frac{2}{49}}\right) \sqrt{\frac{2}{1+0.5347}} = -0.509$$

and from a standard table of the cumulative normal distribution,

$$\beta = \Phi(-0.509) = 0.277$$

In Figure 4 a family of O.C. curves for the U test for one-sided truncation is presented, with the degree of truncation as the independent variable and the probability of acceptance as the dependent variable. This set of curves holds for the case of $\alpha = 0.05$ and equal sample sizes varying between 9 and 100.

To put these results in perspective, it is useful to compare this O.C. curve of the U test for truncation with a somewhat similar U test for the difference between two means. Assume a situation in which measurements from a production process are normally distributed with mean of 100, standard deviation of 10, and upper specification limit exactly 3σ above the mean, at 130. A controlled process operating in this manner will produce 0.00135 fraction defective.

Designate the test for a difference between two means as case (a) and the test for truncation as case (b). In case (a) assume that the population mean shifts upward, increasing the fraction defective produced, but that a bias in a supplier's measurement system offsets exactly this shift in the mean. In case (b) assume that a shift occurs but that truncation in sample selection results in supplier samples which indicate no defectiveness. In either case, a customer sample will probably contain defective units if the shift is at all large, though it is not the purpose of the customer's inspection in this analysis to find defectives, per se. Rather, we assume in case (a) the customer is interested in testing for measurement bias by inspecting an independent random sample; while in case (b) he is interested in testing for truncation in sample selection by inspecting an independent random sample. Assume equal samples of size 49.

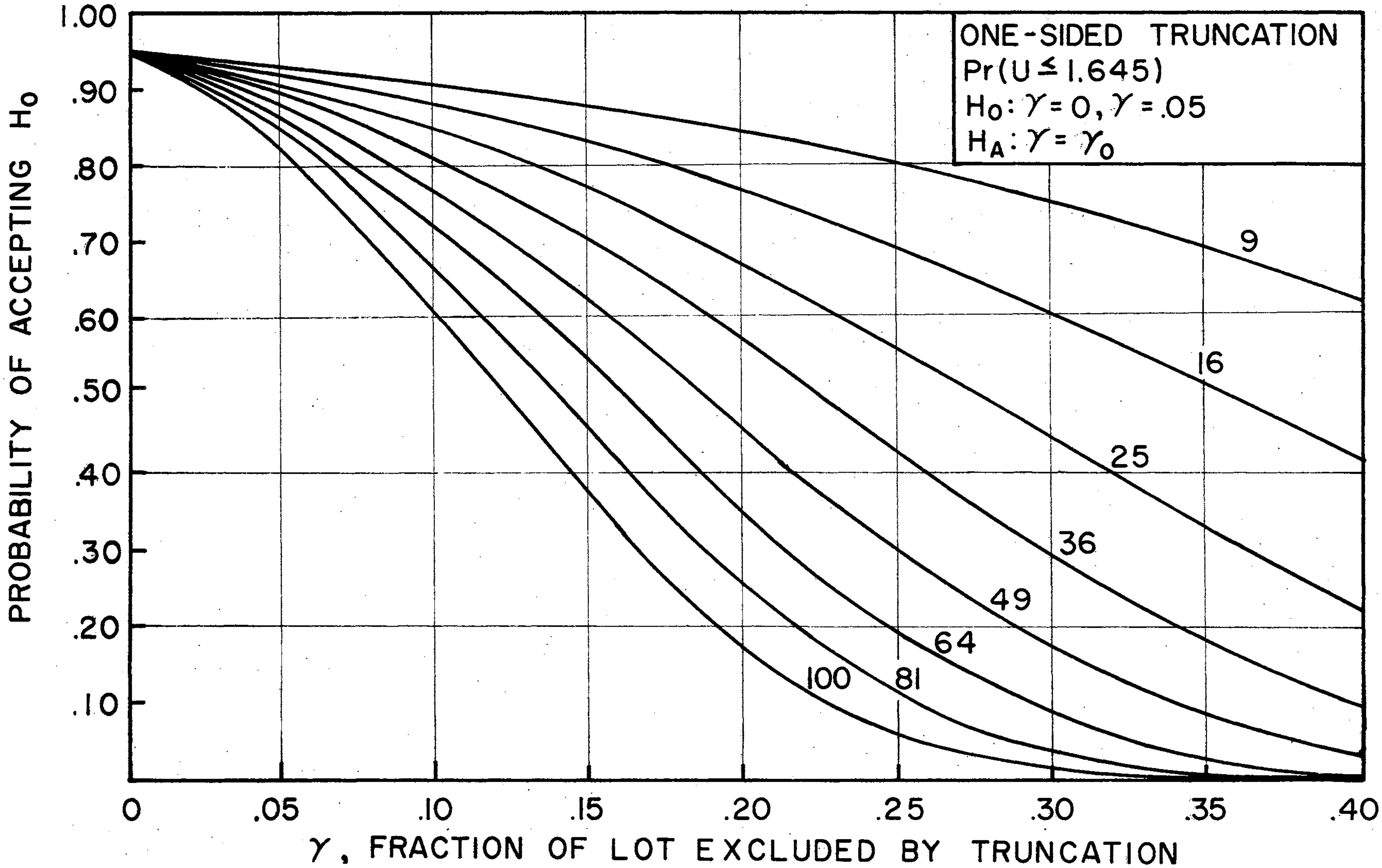

Figure 4. O.C. Curves of U Test for One-Sided Truncation

Consider case (a), with points on the O.C. curve which would result in 0.005, .01, and .05 fraction defective, denoted by p. An increase in the fraction defective from p = 0.00135 to p = 0.005 is equivalent to a decrease in the distance between the mean and the upper specification limit from 3σ to 2.576σ, or a shift upwards by 0.4241σ. Similarly an increase to p = 0.01 is equivalent to a shift of (3-2.326) = .614σ and an increase to p = 0.05 is equivalent to a shift of (3 - 1.645) = 1.355σ. The equation for the O.C. curve of a U test for a shift in the mean is

$$\beta(\delta) = \Pr\left(Z \leq Z_\alpha - \frac{\delta}{\sqrt{\frac{m+n}{mn}}} \right)$$

$$= \Phi\left(Z_\alpha - \frac{\delta}{\sqrt{\frac{m+n}{mn}}}\right)$$

where

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

The equation is derived, e.g. by Bowker and Lieberman (1959, p. 164-165). Applying this equation to the above shifts of δ = .424, .614 and 1.355, with α = .05, Z_α = 1.645, and m = n = 49, gives the following:

For p = .005

$$\beta\ (\delta = .424) = \Phi\left[1.645 - \frac{.424}{\sqrt{\frac{98}{2401}}}\right] = \Phi(-0.4536) = 0.324$$

For p = .01

$$\beta\ (\delta = .614) = \Phi[1.645 - (.614)\ (\frac{7}{\sqrt{2}}) = \Phi(-1.3941) = 0.081$$

For p = .05

$$\beta(1.355) = \Phi[1.645 - (1.355)\ (\frac{7}{\sqrt{2}})] = \Phi(5.0618) = 0.00$$

Now consider case (b) in which the U test is used to detect truncation. With $\alpha = .05$, $n = m = 49$, the equation for the O.C. curve is

$$\beta(\gamma = \gamma_o) = \Phi\{[Z_\alpha - (\lambda_o)\ (\frac{7}{\sqrt{2}})]\sqrt{\frac{2}{1 + \theta_o}}\}$$

which was solved in the earlier part of this chapter for $\gamma = 0.25$.

The O.C. curves of these two tests are plotted to the same scale in Figure 5, which clearly indicates the greater power of a U test for bias, when both are plotted as a function of the fraction of the lot being sampled which is outside the specification limit.

There is a great difference between the two tests when compared on this basis. Two explanations are offered. First consider the physical differences between the two discrepancies in the data. The bias effects all observations in the supplier sample by the amount equal to the bias. But the truncation in sample selection affects only those observations which actually fall above the specification limit. For example, if $\gamma = 0.10$ this amounts to only 10% of the observations.

Second, consider the actual average shift in the mean of the supplier sample. In the case of a bias which results in 0.05 fraction defective, the shift in the mean is 1.355σ. But when truncation in sample selection results in 0.05 fraction defective in the lot yet no defective in the sample, the difference in the true mean of the population being sampled and the population submitted to the customer is only $.1086\sigma$. In summary, for comparable differences between means, the two tests actually perform very comparably, but a much larger fraction

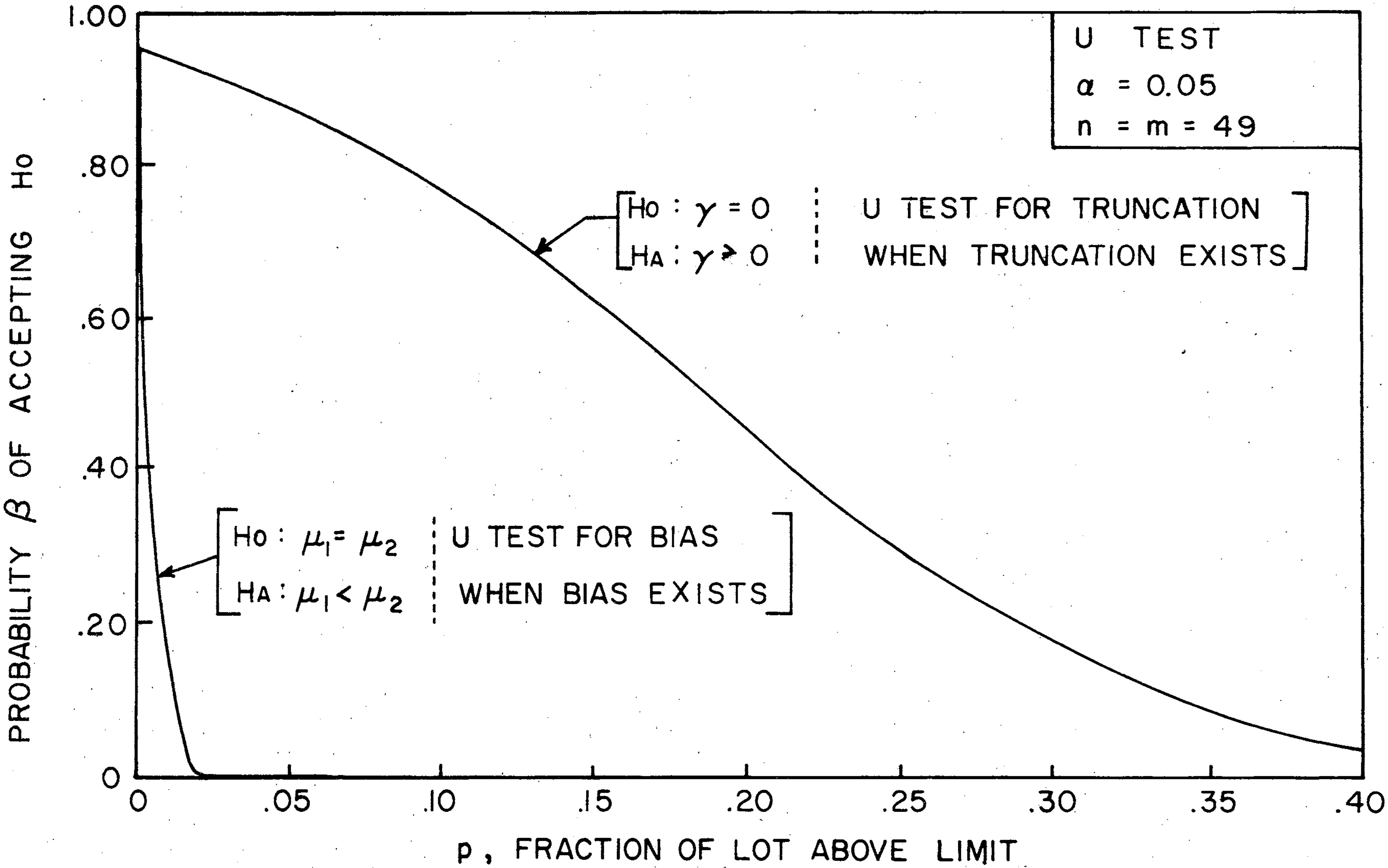

Figure 5. Comparison of OC Curve for U Test of Truncation and U Test of Shift

defective can be concealed with resulting difference in the means by truncation than by measurement bias.

The results of the above analysis should not be over-generalized. It does show the difficulty of detecting truncation but does not indicate the differences among the different possible tests for truncation, which is a more important concern here, and will be dealt with in Chapter VI.

The F Test for Two-sided Truncation

The optimum test for the equality of two variances when the populations are normally distributed is the F test. It is not necessary that the means be either known or equal.

The F statistic is computed from the samples by taking the ratio of the two variances,

$$F = \frac{S_y^2}{S_x^2}$$

where

$$S_y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y})^2$$

and

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

It can be shown (e.g. Bowker and Lieberman, 1959, p. 87) that the variable v^2 formed by the ratio

$$\frac{S_y^2/\sigma^2}{S_x^2 \; \sigma_{x2}^2} = v^2 \qquad (4\text{-}12)$$

is distributed as an F random variable with degrees of freedom m-1, n-1. The degrees of freedom constitute the parameters of the F distribution.

This statement assumes that the distribution of x is normal with variance $\sigma_{x2}{}^2$. While it is true that the variance of the truncated population is $\sigma_{x2}{}^2$, x is not normally distributed. Thus the O.C. curve derived are only approximate.

The decision rule in applying the F test to this problem is to reject the null hypothesis of no truncation if $F > F_{\alpha,\ m-1,\ n-1}$ where $F_{\alpha,\ n-1,\ m-1}$ is the point on the F distribution with m-1, n-1 degrees of freedom such that $Pr(F_{\alpha,\ n-1,\ m-1} < F) = \alpha$. (In subsequent development the appropriate degrees of freedom will be implied rather than explicitly written.)

The null hypothesis will be accepted if

$$\frac{S_y{}^2}{S_x{}^2} \leq F_\alpha .$$

Therefore the probability of accepting H_0: $\gamma 2 = 0$ as a function of the true value $\gamma 2_0$ is

$$\beta(\gamma 2_0) = Pr\ \left(\frac{S_y{}^2}{S_x{}^2} \leq F_\alpha \mid \gamma 2 = \gamma 2_0\right) \qquad (4\text{-}13)$$

In equation (4-12) it was stated that the ratio

$$\frac{S_y{}^2/\sigma^2}{S_x{}^2/\sigma_{x2}{}^2}$$

has an F distribution. This holds true whether or not $\sigma_{x2}{}^2 = \sigma^2$. Therefore the next step is to multiply both sides of the inequality in equation (4-13) by $\frac{1/\sigma^2}{1/\sigma_{x2}{}^2}$ giving

$$\beta = Pr\left(\frac{S_y^2/\sigma^2}{S_x^2/\sigma_{x2}^2} \leq F_\alpha \frac{\sigma_{x2}^2}{\sigma^2} \right) \qquad (4\text{-}14)$$

and the left hand side of the inequality is an F random variable. In Chapter III the ratio $\theta 2$ was defined to be

$$\theta 2 = \frac{\sigma^2}{\sigma_{x2}^2}$$

and a table of $\theta 2$ as a function of $\gamma 2$ was presented as Table II.

Substituting these relationships on both sides of the inequality of equation (4-14) gives

$$\beta(\gamma 2) = Pr(F \leq F_\alpha \cdot \theta 2)$$

which can be solved for a given value of α, $\gamma 2$, n, and m by reference to appropriate tables of the F distribution. A family of O.C. curves so computed is presented in Figure 6. An example of the computation for $n = m = 49$, $\alpha = 0.05$, and $\gamma 2 = 0.25$ is

$$\beta(.25) = Pr\ (F \leq 1.63 \cdot 0.3686)$$

$$= Pr\ (F \leq 0.601)$$

$$= 0.044.$$

So far as is known by the writer, the above derivation and O.C. curves for the F test for two-sided truncation have not been previously developed. However, it should be pointed out that O.C. curves for this test can be obtained by reference to the article by Ferris, et al (1946). In their paper, β was graphed as a function of $\sqrt{\theta 2}$, where $\sqrt{\theta 2}$ denotes the ratio of the larger standard deviation over the smaller one. The parameter $\sqrt{\theta 2}$ can be related to the degree of truncation $\gamma 2$ since both are functions of k:

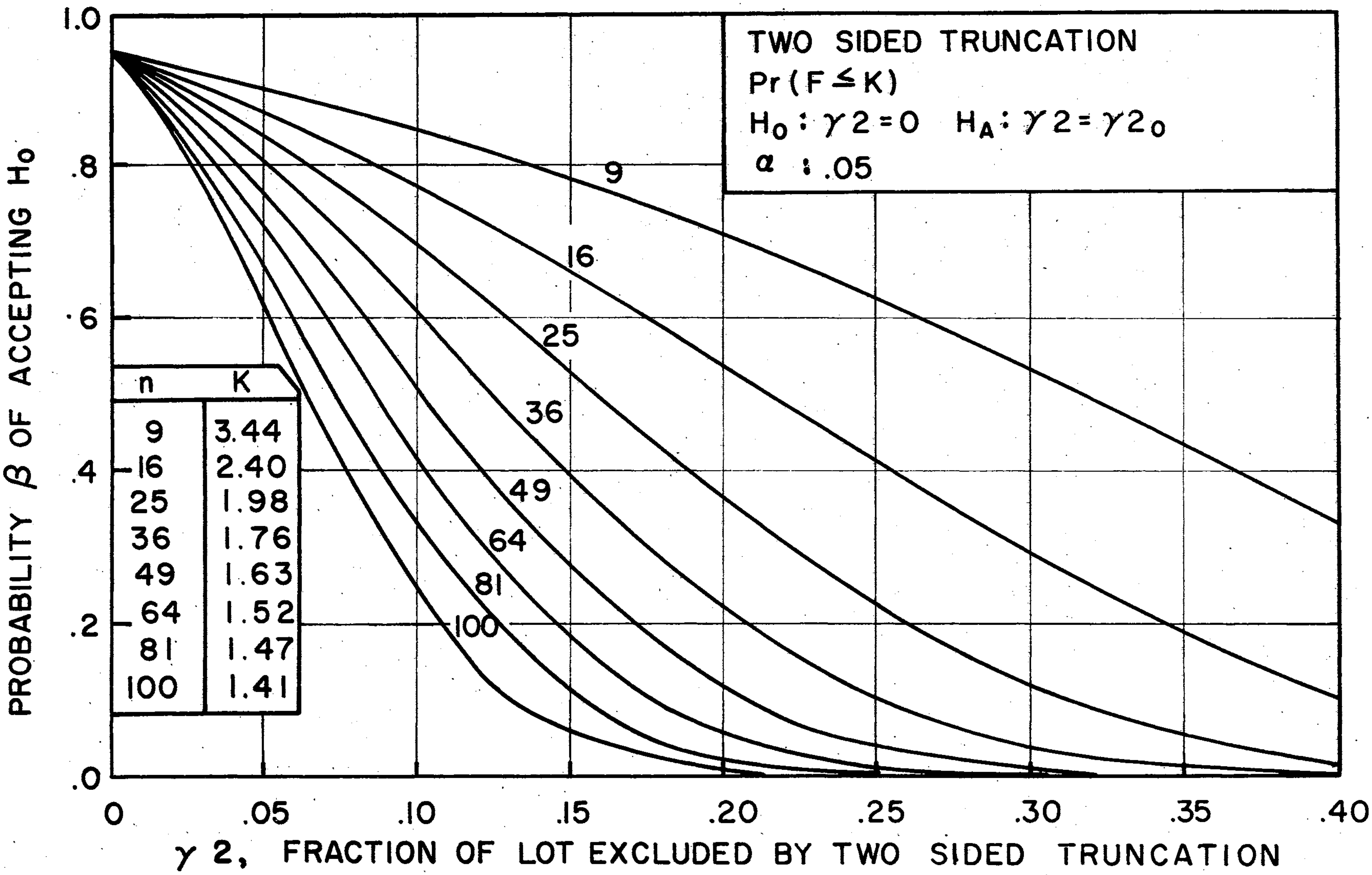

Figure 6. O.C. Curves of F Test for Two-Sided Truncation

$$\sqrt{\theta 2} = \frac{\sigma}{\sigma_x} = \frac{1}{\sqrt{1-\frac{k\phi(k)}{\Phi(k)-1/2}}}$$

and

$$\gamma 2 = 1-\int_{-k}^{k} \phi\ (u)\ du = 2 - 2\Phi(k)$$

The Ferris O.C. curves were presented for $\alpha = .05$ only, and include 28 sample size combinations ranging from $n = 4$, $m = 4$ to $n = 100$, $m = 100$, with equal sample sizes, 2:1 size ratios and 3:1 size ratios. Also, Bowker and Lieberman (1959, p. 191) have presented a comparable chart for $\alpha = .01$.

The χ^2 Test for Two-sided Truncation

The χ^2 test would be applicable to the detection of two-sided truncation if the population variance were known. Inasmuch as the variance is assumed known in applying the U test, it would not seem unreasonable to make the same assumption in the case of two-sided truncation. But, in the two-sided case, if variance is known, then there is no need to draw a customer sample in order to test for supplier sample truncation. For this reason, the χ^2 is not explicitly treated here since attention is focused on two-sample tests. The development of the O.C. curves would be along lines very similar to those of the F test, and it would be, of course, a more powerful test as well as a more economical one.

Nonparametric Tests

Three nonparametric tests are studied in this investigation. They are selected from a large number of possible nonparametric tests on the basis of published evidence regarding their relative power characteristics for the standard alternate hypotheses, and on the basis of the appropriateness of their method of computation. For example, two common nonparametric tests are the median test and the runs test. Both are omitted from this investigation because several authors (e.g. Siegel, 1956) have indicated their power to be lower than competing tests described here. Some relatively powerful tests, such as Walsh's test and Fisher's randomization test are excluded because the complexity of computing these tests would rule them out as tests easy to learn and apply in field operations.

The three tests studied are the Kolmogorov-Smirnov test, also referred to in this thesis as the maximum difference test since it is based on the maximum difference in the empirical distribution functions of two samples; the Mann-Whitney test, referred to as the rank-sum test; and the Wilks-Rosenbaum or exceedance test. Of the three tests mentioned, the first two are well described and tabled in most standard references on nonparametric statistics. Because the exceedance test is less known, it is described in more detail herein.

The Maximum Difference Test

The maximum difference two-sample test was first proposed by Smirnov, (1939), who extended a one-sample test previously developed by Kolmogorov, thus giving rise to the more common name, the Kolmogorov-Smirnov test. The formula for the significance probabilities developed by Smirnov was actually an asymptotic or limit formula. Massey, (1951) developed a computation formula for obtaining the small-sample exact significance probabilities. Hodges, (1957), published findings on the degree of error of the Smirnov formula for small samples, as well as several alternative approaches, both exact and approximate, for calculating the significance probability. Each of these sources contains tables of significance probabilities, as does Siegel's book (1956).

The term maximum difference relates to the method of performing the test. For each sample, the empirical distribution function is formed, and then the range of the variable is searched for the maximum difference between the two functions. This is the value of the test statistic.

To apply the maximum difference test to the hypothesis $F(x) = G(y)$, define the empirical distribution function of one sample, of size m, to be the step function

$$G_m(v) = \frac{k}{m}$$

where k is the number of observations not greater than v, and v is any particular value of y. Likewise for the other sample, of size n, the empirical distribution function is $F_n(v)$. Now define

$$D = \max_{\text{all } v} \left| F_n(v) - G_m(v) \right|$$

for a two-sided test, the alternate hypothesis being H_A: $F(x) \neq G(y)$. Define

$$D^+ = \max_{\text{all } v} (F_n(v) - G_m(v))$$

for a one-sided test, the alternate hypothesis being H_A: $F(x) > G(y)$. Thus the test is actually a test on the equality of two population cumulative distribution functions, and it has the property that it can, in theory, be used to detect any type difference whatsoever between the two distributions, e.g., central tendency, dispersion, or shape. As a practical matter, of course, its ability to detect differences is dependent on sample sizes.

The one-sided version of the maximum difference test is appropriate for detection of one-sided truncation. For detecting two-sided symmetrical truncation the two-sided maximum difference would seem to be the logical choice, but in actuality either of the two alternative one-sided tests would give about the same risks. In Chapter IV the O.C. curves of the maximum difference test against the truncation alternative are estimated by means of a distribution sampling experiment. In the experiment, the two-sided test is used against two-sided truncation and the one-sided test against one-sided truncation.

The Rank-Sum Test

The rank-sum test was originally developed by Wilcoxon, (1954), for the case of equal sample sizes. It was studied in greater detail by Mann and Whitney, (1949), who extended it to the case of unequal sample sizes. It is now frequently referred to as the Mann-Whitney test. Siegel, (1956), called it "one of the most powerful of the nonparametric tests for location." Thus it is a reasonable candidate test for the detection of one-sided truncation, which does affect location. Siegel's book contains tables of significance probabilities and also refers to other, more extensive tables. Mann and Whitney in their 1949 paper show the mathematical basis of computing the test statistic and derive its probability function and associated critical values.

There are several alternate methods of computing the rank-sum statistic for two independent samples. A straightforward method which illustrates the name rank sum, follows: Combine the observations from both samples into a single group, maintaining sample identity within the group. Assign ranks to all observations in the combined group. Rank the lowest valued reading as "1", the next lowest "2", and so on. Then sum the ranks assigned to the customer sample. (The choice of the customer sample is arbitrary, but the computational and test procedure would have to be altered if the test statistic were based upon the supplier sample instead.) Call this sum of all the ranks of the customer sample ΣR_c. Next, compute the product of the two sample sizes, nm, and the product $\frac{n}{2}(n+1)$. Finally, letting RS represent the rank sum statistic,

$$RS = nm + \frac{n}{2}(n+1) - \sum_{i=1}^{m} R_{ci}$$

If the suspected truncation is on the upper tail this method will result in an RS value which can be compared with the critical value of Table J or K of Siegel's book to determine its significance. If truncation were suspected on the lower tail of the customer sample the same result could be obtained by computing ΣR_c in the opposite direction, i.e., assigning the rank "1" to the largest observation in the combined sample, the rank "2" to the next largest observation, until the entire sample was accounted for.

The Exceedance Test

The exceedance test is one which is not discussed much, if at all, in texts on statistical methods. However, it has a fairly long history of development, dating back at least to 1942. Although there exist many versions of exceedance-type tests they all can be related to the simple idea of counting the number of observations in one sample which exceed, (in either a positive or negative direction) a specified ranked observation in the other sample. In this paper we consider only exceedance based upon the largest (or smallest) order statistic.

To define the exceedance statistic, we specify that from a continuous distribution f(x), a first sample, of size n is drawn and the observations $x_1, \ldots, x_n$ are ranked in their order of magnitude. Let $x_{(1)}$ denote the smallest observation and $x_{(n)}$ the largest observation in the sample of size n. In a second sample, $y_1 \ldots, y_m$, of size m, the observations y_i drawn from the same distribution, there will be A observations smaller than $x_{(1)}$, and B observations greater than $x_{(n)}$. Then there will be $E_2 = (A+B)$ observations outside the extreme order statistics of the first sample, where E_2 may be 0, 1, 2, ..., or m. The random variables A and B, are defined as the one-sided exceedances of the second sample.

For the purposes of data validation it is assumed that the direction of sample truncation is known in advance where a one-sided test is appropriate. Therefore the random variable of interest is E_1, where

$$E_1 = A$$

if truncation at the lower end of the sample is to be tested, and

$$E_1 = B$$

if truncation at the upper end of the sample is to be tested.

The exceedance statistic is actually a specialization of both the rank sum and maximum difference statistics. Whereas the maximum difference statistic is found by examining the difference between empirical distribution functions across the entire range of the two variables and picking the maximum, the exceedance statistic is a simple function of the difference between the two distribution functions at a specified end-point. Thus it can be considered a specialization of the maximum difference statistic. Whereas the rank sum statistic is computed by counting, e.g., the number of y's greater than each x, a comparable exceedance statistic is computed by counting the number of y's greater than the largest x observation. Thus the one-sided exceedance is also a specialization of the rank sum statistic.

The mathematical basis of the exceedance test was laid by Wilks (1942) in a classic paper on statistical tolerance limits. Wilks derived general probability formulas for the one-sided and two-sided tests; however, he presented his results as a means for estimating population quantiles and not as a method for testing hypotheses. Gumbel and von Schelling (1950) extended Wilks' work by computing the moments and the cumulative probability function of the number of exceedances on the largest order statistics. They also derived large-sample approximation formulas.

Rosenbaum proposed (1953) that the two-sided exceedance be utilized as a nonparametric test for dispersion, and (1954) that the one-sided exceedance be used for a nonparametric test of location. He presented tables of critical values at the nominal 5% and 1% level for both tests for sample sizes through 50.

Epstein (1954) presented limited tables of the distribution of the one-sided exceedance; he exploited the symmetrical properties of the distribution in order to achieve great brevity in the tables. His tables are for equal sample sizes only, extending through size 20, and provide the probabilities of exceedances for all order statistics.

There are a number of adaptations of the basic test. For example, Tukey proposed (1959) a modification of the exceedance test to increase its power and simplicity of use as a quick-test on location only. His procedure requires that one sample contain the highest value, the other the lowest value. If this requirement is met then an exceedance is computed from each sample, and the two are summed to obtain the test statistic. He found that the numbers 7, 10, and 13 provided roughly 5%, 1%, and 0.1% significance levels for two-sided location tests <u>regardless of sample size</u> so long as they were within a ratio of 4:3 in relative size. Other adaptations and developments of the exceedance test were reported by Rosenbaum (1965) who also reported the results of random sampling experiments which compared the power of the various versions with each other and with other types of tests. The results of these experiments are cited in Chapter V.

The probability function of the statistic which Wilks (1942) derived using the multinomial distribution law, is as follows: Let c be a specific value of the random variable E and m and n be as previously defined. Then for the one-sided exceedance, E_1,

$$Pr(E_1 = c) = \frac{(n)(m!)(m - c + n - 1)!}{(m - c)!\ (m + n)!}$$

For the two-sided exceedance, E_2,

$$Pr(E_2 = c) = \frac{(n)(n - 1)(c + 1)n!(n + m - c - 2)!}{(m - c)!\ (m + n)!}$$

The numerical evaluation of these formulas is somewhat tedious, requiring the use of tables of factorial logarithms, or analogous functions if a computer program is utilized. But the function can also be computed recursively, as will now be shown.

Consider first the one-sided exceedance. As stated previously, it is assumed that if a one-sided exceedance test is to be applied, the choice of upper or lower tail can be specified in advance. Under the null hypothesis, the probabilities are identical regardless of the choice of tail. Choosing the upper tail arbitrarily, rank the (m + n) observations into a single group, preserving only the sample source designations x and y. Based upon the hypothesis that both samples are from the same distribution, all $\binom{n+m}{n}$ possible arrangements are equally likely. The probability that the exceedance equals zero is identical to the probability that the largest observation is an x, which must be equal to the ratio of x's to the combined sample size. Let L represent the largest observation in the combined sample. Then,

$$Pr(E_1 = 0) = Pr(L = x) = \frac{n}{m+n} \tag{4-15}$$

The probability that the exceedance equals one is identical to the probability that the largest observation is y and the next largest is x. Each conditional probability is simply the relative size of the specified sample to the combined sample given the prior condition.

$$Pr(E_1 = 1) = Pr(L = y,\ L\text{-}1 = x) = Pr(L = y)\cdot Pr(L\text{-}1 = x \mid L = y)$$

$$= \frac{m}{n+m}\cdot\frac{n}{n+m-1}\cdot\frac{n}{n+m}\cdot\frac{m}{n+m-1} = Pr(L = x)\cdot\frac{m}{n+m-1}$$

Since $Pr(L = x)$ equals $Pr(E_1 = 0)$ this reduces to $Pr(E_1 = 0)\cdot\frac{m}{n+m-1}$.
The probability that the exceedance equals two is identical to the probability that the largest and next largest observations are y, and the third largest is x.

$$Pr(E_1 = 2) = Pr(L = y,\ L\text{-}1 = y,\ L\text{-}2 = x)$$

$$= \frac{m}{m+n} \cdot \frac{m-1}{m+n-1} \cdot \frac{n}{m+n-2}$$

$$= \frac{n}{m+n} \cdot \frac{m}{m+n-1} \cdot \frac{m-1}{m+n-2}$$

$$= Pr(E_1 = 1) \cdot \frac{m-1}{m+n-2} \tag{4-16}$$

By mathematical induction we extend this recursive relationship to the general term. By means identical to that for $P(E_1 = 1)$ and $Pr(E_1 = 2)$ we see that

$$Pr(E_1 = k) = Pr(L = y,\ L-1 = y,\ \ldots,\ L-(k-1) = y,\ L-k = x)$$

$$= \frac{(m)(m-1)\ \ldots\ (m-(k-1))(n)}{(m+n)(m+n-1)\ \ldots\ (m+n-(k-1))(m+n-k)} \tag{4-17}$$

and

$$Pr(E_1 = k+1) = Pr(L=y,\ L-1 = y,\ \ldots\ ,\ L-(k-1) = y,\ L-k=y,\ L-(k+1) = x)$$

$$= \frac{m(m-1)\ \ldots(m-(k-1))(m-k)(n)}{(m+n)(m+n-1)\ldots(m+n-(k-1))(m+n-k)(m+n-(k+1))} \tag{4-18}$$

Comparing factors in equations (4-17) and (4-18) we see that

$$Pr(E_1 = k+1) = Pr(E_1 = k) \cdot \frac{m-k}{m+n-(k+1)} \tag{4-19}$$

Equation (4-19) gives a recursive equation for the general term. The inductive proof is completed by showing that equation (4-19) holds for the case $k = 1$. Substituting $k = 1$ into equation (4-19) gives

$$Pr(E_1 = 2) = Pr(E_1 = 1) \cdot \frac{m-1}{m+n-2}$$

which was the result derived above as equation (4-16). Now equations (4-15) and (4-19) provide a straightforward method of computing the probability function of the one-sided exceedance.

A similar recursive formula for the two-sided exceedance can be derived by applying the same logic. In this case it is necessary to count the number of y's which are larger than $x_{(1)}$ to obtain E_2. Let S represent the smallest observation in the combined sample. Then,

$$\begin{aligned} \Pr(E_2 = 0) &= \Pr(S = x,\ L = x) \\ &= \Pr(S = x)\ \Pr(L = x \mid S = x) \\ &= \frac{n(n-1)}{(m+n)(m+n-1)} \end{aligned} \tag{4-20}$$

The event $E_2 = 1$ can occur in two equally likely ways, i.e.,

$$S = y,\ S+1 = x,\ \ldots,\ L = x$$

and

$$S = x,\ \ldots,\ L - 1 = x,\ L = y$$

Thus,

$$\begin{aligned} \Pr(E_2 = 1) &= 2 \cdot \Pr(S = y,\ S+1 = x,\ L = x) \\ &= 2 \cdot \frac{mn(n-1)}{(m+n)(m+n-1)(m+n-2)} \\ &= 2 \cdot \Pr(E_2 = 0) \end{aligned}$$

The event $E_2 = 2$ can occur in 3 equally likely ways, i.e.,

$$S = x,\ \ldots,\ L-2 = x,\ L-1 = y,\ L = y$$

$$S = y,\ S+1 = x,\ \ldots,\ L-1 = x,\ L = y$$

$$S = y,\ S+1 = y,\ S+2 = x,\ \ldots,\ L = x$$

Since the probabilities of all of these ways are additive,

$$\Pr(E_2 = 2) = 3 \cdot \frac{mn(m-1)(n-1)}{(m+n)(m+n-1)(m+n-2)(m+n-3)}$$

$$= \Pr(E_2 = 1) \cdot \frac{3}{2} \cdot \frac{m-1}{m+n-3} \qquad (4\text{-}21)$$

In general, there are (k+1) equally likely events for which $E_2 = k$. One of these k configurations is now stated, and the probability factors appear in the same order as the configuration.

$$S = x,\ L = y,\ L-1 = y,\ \ldots,\ L-(k-1) = y,\ L-k = x$$

$$\frac{n}{n+m}\ \frac{m}{n+m-1}\ \frac{m-1}{n-m-2} \cdots \frac{m-(k-1)}{n+m-k} \cdot \frac{n-1}{n+m-(k+1)} \qquad (4\text{-}22)$$

Inasmuch as all k+1 events are equally likely, the probability is

$$\Pr(E_2 = k) = (k+1)\ \frac{n \cdot m \cdot m-1 \cdots m-(k-1) \cdot n}{n+m \cdot n+m-1 \cdot n+m-2 \ldots n+m-k \cdot n+m-(k+1)} \qquad (4\text{-}23)$$

There are (k+2) equally likely events yielding $E_2 = k+1$, a representative one being

$$S = k,\ L = y,\ L-1 = y,\ \ldots,\ L-(k-1) = y,\ L-k = y,\ L-(k+1) = x$$

and it thus follows that

$$\Pr(E_2 = k+1)$$

$$= (k+2)\ \frac{n \cdot m \cdot (m-1) \cdots (m-(k-1))(m-1) \cdot n}{(n+m)(n+m-1)(n+m-2) \ldots (n+m-k)(n+m-(k+1))(n+m-(k+2))} \qquad (4\text{-}24)$$

Now by comparing the factors on the right hand sides of equations (4-23) and (4-24) we see that

$$Pr(E_2 = k+1) = Pr(E_2 = k) \cdot \frac{k+2}{k+1} \cdot \frac{m-k}{n+m-(k+2)} \tag{4-25}$$

The inductive proof is completed by showing that (4-25) holds for the case k = 1.

$$Pr(E_2 = 1 + 1) = \frac{1+2}{1+1} \cdot \frac{m-k}{m+n-(k+2)}$$

$$= \frac{3}{2} \cdot \frac{m-1}{m+n-3}$$

which agrees with equation 4-21. Thus the equation is shown to hold in general.

Inasmuch as no tables of the exact significance probabilities of the E_1 test are available for the case where $n \geq 20$, a table is presented in Appendix D which has both E_1 and E_2 tabulated for selected equal sample sizes, from n = 2 to n = 100. All those values of which have significance probabilities between 0.10 and 0.01 are tabulated.

Analysis of the Nonparametric Tests

Of those tests described in this chapter, O.C. curves were derived for only two: the U test and the F test. However, the O.C. curves for the RS, MD, and E tests are of considerable interest in their own right and for comparison with the above derived results. The empirical estimation of these O.C. curves is the subject of the next chapter.

CHAPTER V

OPERATING CHARACTERISTIC CURVES FOR THE NONPARAMETRIC TESTS

Of those tests presented in the previous chapter, three are nonparametric: the E test, the MD test, and the RS test. Nonparametric tests as a class have certain features in common. The most important, from which the name is derived, is that the hypotheses under test are not on specific parameters of a distribution but rather on the distribution itself. Another label often applied to this type of test is "distribution-free". That is, no assumptions need be made about the form of the distribution being sampled in order to compute the significance probability of the test. In applications such as the one being studied here, this distribution-free property has a definite appeal, because it is likely that the normal assumption will not hold. In some cases, nonparametric tests also have an advantage in simplicity of computation. This is particularly true of the exceedance tests. The particular aspect of nonparametric tests which concerns us here is their respective O.C. curve analysis.

Relative Merits of Sampling Experiments

The literature of studies of O.C. curves (or equivalently, power) of nonparametric tests indicates two basically different approaches in developing results. One approach is characterized by the work of Lehman (1953) who formulated alternative hypotheses which were "tailored" to the structure of the test, and made simplifying assumptions so that general analytic expressions could be derived. This procedure was followed in the previous chapter for the U and F tests. Lehman's work has been extended by Gibbons (1964) who also provided a good summary of related developments.

The other approach is to construct a descriptive model of an actual situation of concern for which no mathematical method of analysis is available, and then to make empirical estimates of the desired function, by tabulating the results of repeated trials of sampling experiments.

The analytical methods have traditionally been favored for their rigor and generality of application. But this approach also has its limitations, as pointed out by Hammersly and Handscomb (1964), who emphasized this point: that one of the main strengths of theoretical mathematics is its concern with abstraction and generality, in that one can write symbolic expressions or formal equations which abstract the essence of a problem and reveal its underlying structure. However, this same strength carries with it an inherent weakness: the more general and formal its language, the less is mathematical theory ready to provide a numerical solution in a specific application.

The limitations of mathematical analysis have lead some researchers to exploit the empirical technique of distribution sampling. A number of such studies have been reported. Dixon and Teichroew (1953) determined for various sample sizes, such as $n = m = 5, 10, 20$; $m = 5, n = 10$; $m = 10, n = 20$; and others, the powers of the Wilcoxon ranksum test and the Komogorov-Smirnov maximum difference tests against normal shift alternatives. A normal shift alternative is of the form

$H_A: \mu_1 \neq \mu_2$

where μ_1 and μ_2 are the means of two normally distributed populations with common variance σ^2. The significance levels considered were $\alpha = 0.01$, 0.05, and 0.10. They also made experiments with alternative hypotheses for which variances as well as means were unequal. The number of trials, i.e., pairs of samples, on which their estimates were based was 150, and in some cases 100.

Epstein (1955) compared the power of the runs test, the exceedance test, a truncated variation of the maximum difference test, and the rank sum test against the normal shift alternative. The significance level was maintained at 0.05. The results were based on 200 trials, for sample size 10 only.

An empirical analysis of the probabilities of all rankings of the order statistics of two-sample tests was reported by Teichroew (1955) for very small sample sizes up to and including $n = 3$, $m = 4$. The normal shift alternative was considered. Since all possible rankings were presented, no specific tests or significance levels were considered for the reader could choose any rank-order test and significance level he wished and analyze it for the sample sizes given. The results were based upon 1,000 trials in some cases and 2,000 trials in others.

Van der Laan and Oosterhoff (1965) compared the Wilcoxon rank-sum test with two specialized rank order tests for samples of size six on normal shift alternatives. The significance level was $\alpha = 0.01$ and results were based upon 2,000 trials for each of twelve specific shift alternatives. Also in 1965, Rosenbaum compared several different versions of the one-sided and two-sided exceedance test for normal shift alternatives, sample sizes $n = m = 10$, and nominal significance levels of 0.05 based upon 100 trials.

For this dissertation, a distribution sampling experiment is conducted to estimate the power of the rank-sum, exceedance, and maximum difference tests, as a function of the alternatives of one-sided and symmetrical two-sided truncation, and is discussed in greater detail following the general description of the distribution sampling technique.

General Description of the Distribution Sampling Technique

The sampling method for evaluating the distribution of a two-sample test statistic can be described in general terms as follows:

(a) Perform a large number, (NT), of trials, each trial consisting of two samples (of size NS each), one from the density function f(y) and the other from the density function g(x), where f(y) and g(x) define the two populations under comparison. The two populations are simulated by the use of random number generators.

(b) For each pair of samples, compute the test statistic $S(x_{i1},\ldots, x_{iNS}, y_{i1},\ldots, y_{iNS})$ for i = 1, ... NT

(c) Rank the NT values of the statistic. The number of distinct values will in the case of some statistics be limited by the sample size NS (e.g., MD, E) and in the case of others by the number of trials NT, (e.g., RS for large samples). The number of trials will be the limit whenever the sample space is larger than NT.

(d) Let the set of numbers C_K, K = 1, 2...NS or K = 1, 2, ...NT, represent the ranked set of distinct values of S(x,y), and for any number C_K, compute

$$\hat{F}(C_K) = \frac{NL}{NT}$$

where NL is the number of trials for which the statistic S(x,y) is less than or equal to C_K.

When $\hat{F}(C)$ is an empirical distribution function of the test statistic S(x,y) and NT becomes large, this empirical function will tend to approximately equal the true distribution function. This is a consequence of the law of large numbers, which can be used to show that

$$\lim_{NT \to \infty} [\hat{F}(C)] = F(C)$$

where F(C) is the true distribution function of C.

The distribution sampling technique is applied to the development of approximate O.C. curves of a test for truncation by making one run for each desired value γ_o of the truncation parameter γ, each run consisting of NT trials. Each trial is a single computation of the test statistic for the specified alternative, each run yields an approximate distribution of the test statistic for that alternative, and each set of runs for a given sample size gives a family of sampling distributions from which a family of O.C. curves is derived. Each curve is developed by varying γ and holding C fixed at a certain value. To make one <u>trial</u> requires 2 x NS random numbers to be chosen, stored, and analyzed. Each run requires a sum of NT trials for identical value of all parameters. For each O.C. curve a sequence of approximately six to eight runs is required, varying the truncation parameter γ from zero up to a value such that $\hat{\beta}(\gamma) < .10$. Thus a typical O.C. curve is constructed from 500 x 8 = 4,000 trials, which for samples of 25, represent 200,000 individual random numbers, not including those in the truncated portion of the simulated supplier sample, which are rejected.

The probability of acceptance, β, for $\gamma = 0$ is simply the complement of the significance level α and for all the statistics under study, was computed analytically. By selecting a particular sample size and value of α, the critical point C_α on the distribution of the test statistic is determined. That is C_α is determined by the relationship

$$\alpha = Pr(S > C_\alpha \mid \gamma = 0)$$

in the case of an upper one-sided rejection region and

$$\alpha = Pr(S > |C_\alpha| \mid \gamma = 0)$$

for a two-sided rejection region. For nonparametric tests, C_α can take on only certain discrete values, due to the finite sample space of the test statistic. A conventional practice is to select critical values which closely correspond to nominal α such as 0.10, 0.05, and 0.01.

With the appropriate value of C_α determined analytically, β is estimated empirically by the equation

$$\hat{\beta}(\gamma_o) = \hat{F}(C_\alpha \mid \gamma = \gamma_o)$$

The Sampling Experiment Details

The computer program for this experiment was written in the FORTRAN language. This program was developed to integrate the above factors, generate the required pseudo-random numbers, and tabulate and print out the empirical distribution function and variables identification for each run.

The variables for each run which were specified on a control card were the number of trials, the sample size, and the truncation point. The choice between one-sided and two-sided truncation was made by changing a FORTRAN statement in the NORM subroutine. Figure 7 illustrates the flow chart of the program and a complete listing appears as Appendix A.

The populations from which the sample data were obtained were the normally distributed population with $\mu = 0$, $\sigma = 1$ for f(y), and a truncated derivative of this population for g(x). The truncated function was derived from the parent by simply discarding each deviate which exceeded the point of truncation and drawing an additional deviate, repeating this process until the required sample size was attained.

The random number generator used in this program was written in the COMPASS language by S. Bell, formerly a Sandia Corporation statistician, for the Sandia CDC 3600 computer system, under the code name ANRV, which stands for "A Normal Random Variate." Bell used a "mixed congruential technique" of generating a series of pseudo-random normal deviates. This method was reported by Marsaglia, et. al, (1964) who indicated it had been found to be somewhat faster than other normal deviate generators of comparable quality. It has been programmed into a number of different computer systems.

The statistical characteristics of the pseudo-random member series generated by this "ANRV" subroutine were studied extensively by Bell and Holdridge (1967). They applied four different statistical tests to a large number of different sequences, varying in length from 50 to

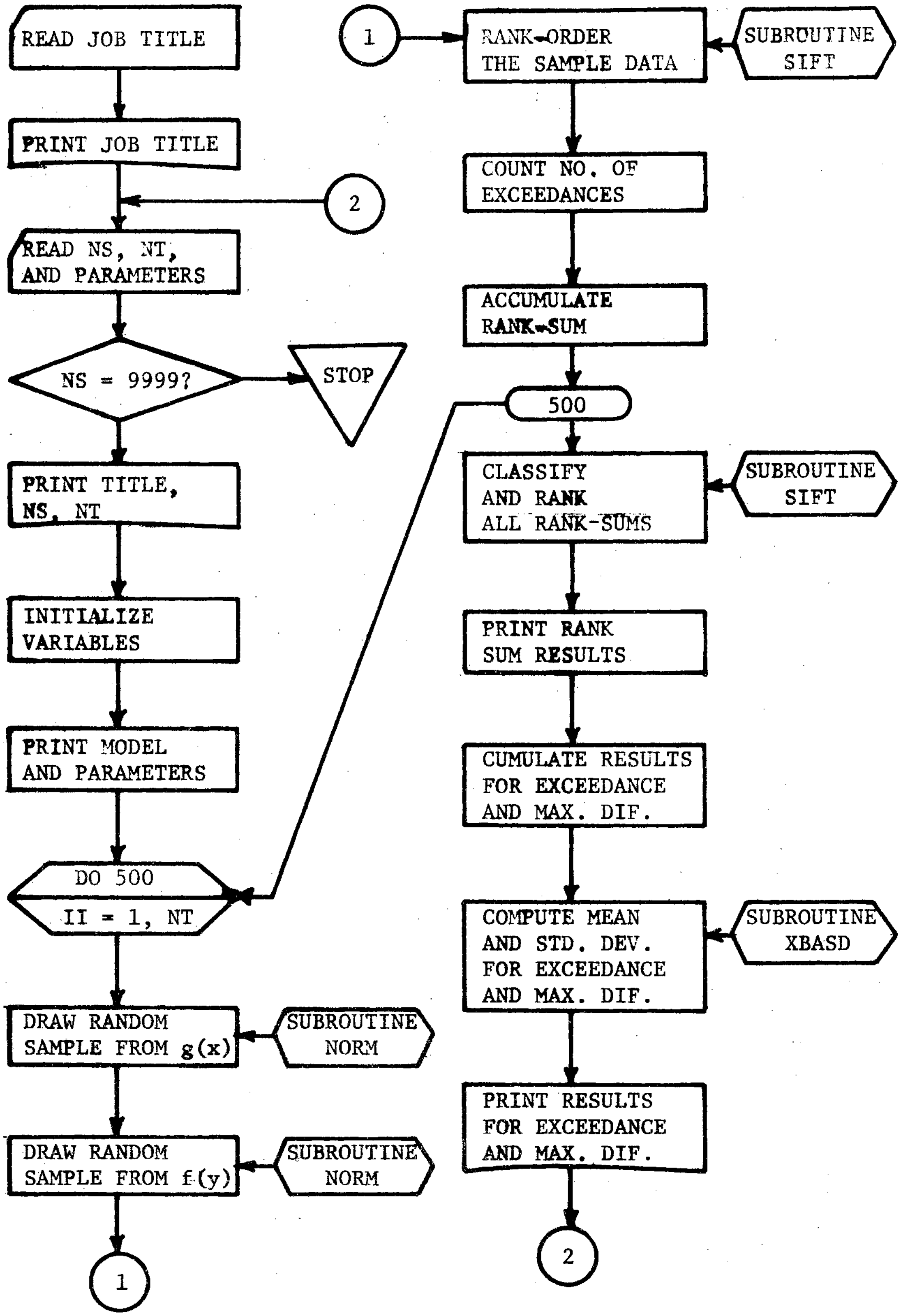

Figure 7. Flow Chart, Distribution Sampling Program

100,000. The tests were the frequency, serial, runs above and below the mean, and runs up and down. The report by Bell and Holdridge describes the nature and outcome of these statistical tests of randomness in considerable detail. The key conclusion of the study was that no evidence of any significant departure from randomness was revealed. The fact that the extensive analysis they performed failed to indicate non-randomness gives a measure of confidence in using the generator in a sampling experiment.

A value of each of the following variables was specified for each run: sample size NS, number of trials NT, truncation point k, one- or two-sided truncation. In then deriving O.C. curves from sampling distributions, appropriate critical values C_{α} were chosen. Each of these variables is now discussed.

Sample Sizes

For this study, equal samples of size 9, 16, ..., 100 were used. The use of equal sample sizes was based on the fact that in most cases equal sample sizes provide the most efficient use of the data, and also to minimize computer storage problems. The choice of squares was to provide a fairly uniform increase in power as a function of sample size increase. It was found that the O.C. curves would be more nearly equally spaced on the sample size scale, since the standard error of estimate of most statistics is a linear function of the square root of sample size. It is not advocated that only squares should be considered when selecting sample sizes. On the contrary, interpolation in a graph of a family of O.C. curves should be made when one wishes to analyze the O.C. curve for a sample size other than those presented. Such an interpolation is a straightforward procedure.

Number of Trials

An optimum choice of the number of trials, i.e., run length, involves a trade-off of increased precision of results on the one hand and increased computer time on the other hand. Reported sampling experiments show wide variance on this point. Van der Laan and Oosterhoff (1965) ran 2,000 trials for each combination of variables. But Rosenbaum (1965) used only 100 trials per combination. Probably each of these extremes could be justified on the basis of different points of emphasis in the respective experiments.

In this experiment the number of trials for each combination of the independent variables was set at 500 for sample sizes 49 and smaller, and at 200 for sample sizes of 64 or larger. There were two considerations in reducing the run length from 500 to 200 trials for the relatively large (64, 81, 100) sample sizes. First, the computer time increased markedly as sample size increased, due mainly to the great number of sorting steps required to rank a large number of numbers. Thus, the run lengths were reduced to help equalize the relative costs of investigation for different sample sizes. The other reason, is that as sample size increases, the O.C. curve becomes steeper, and thus a less precise estimate of the value of β as a function of γ is required in order to have a satisfactory indication of the relationship. This point is discussed in greater detail in Appendix C.

Truncation Points

The alternative programmed into each simulation run was that a specified fraction of the parent distribution was excluded by truncation, from appearing in one of the samples. Both truncation on the upper tail and symmetrical truncation on both tails were studied. The actual truncation points programmed varied from one situation to another: the criteria used were that the points should be close enough to each other to provide a fairly clear idea of inflection points in the resulting O.C. curves, and that the degree of truncation should extend over a great enough range to bring the probability of acceptance down to approximately 0.10 for a specified critical value. The further limitation of a 60% degree of truncation was imposed. After some initial experimentation, increments of .05 were used for successive runs, with additional runs in individual cases where a finer increment was required to adequately estimate a given O.C. curve. Only degrees of truncation through 40% are presented in the O.C. curves, Figures 8 through 15, since it is doubtful that higher degrees would have practical relevance.

Significance Probabilities

The method of distribution sampling estimates the entire distribution function of the test statistic for each set of conditions; thus it is possible to consider any significance probability within the limitation of the discrete probability range of the tests. In the presentation of results, however, the significance probability range was restricted approximately to the interval between 0.03 and 0.15. In the case of actual application of the analyzed test to a data validation problem it is felt that practical considerations would force the user to choose a significance probability somewhere in the 0.05 to 0.10

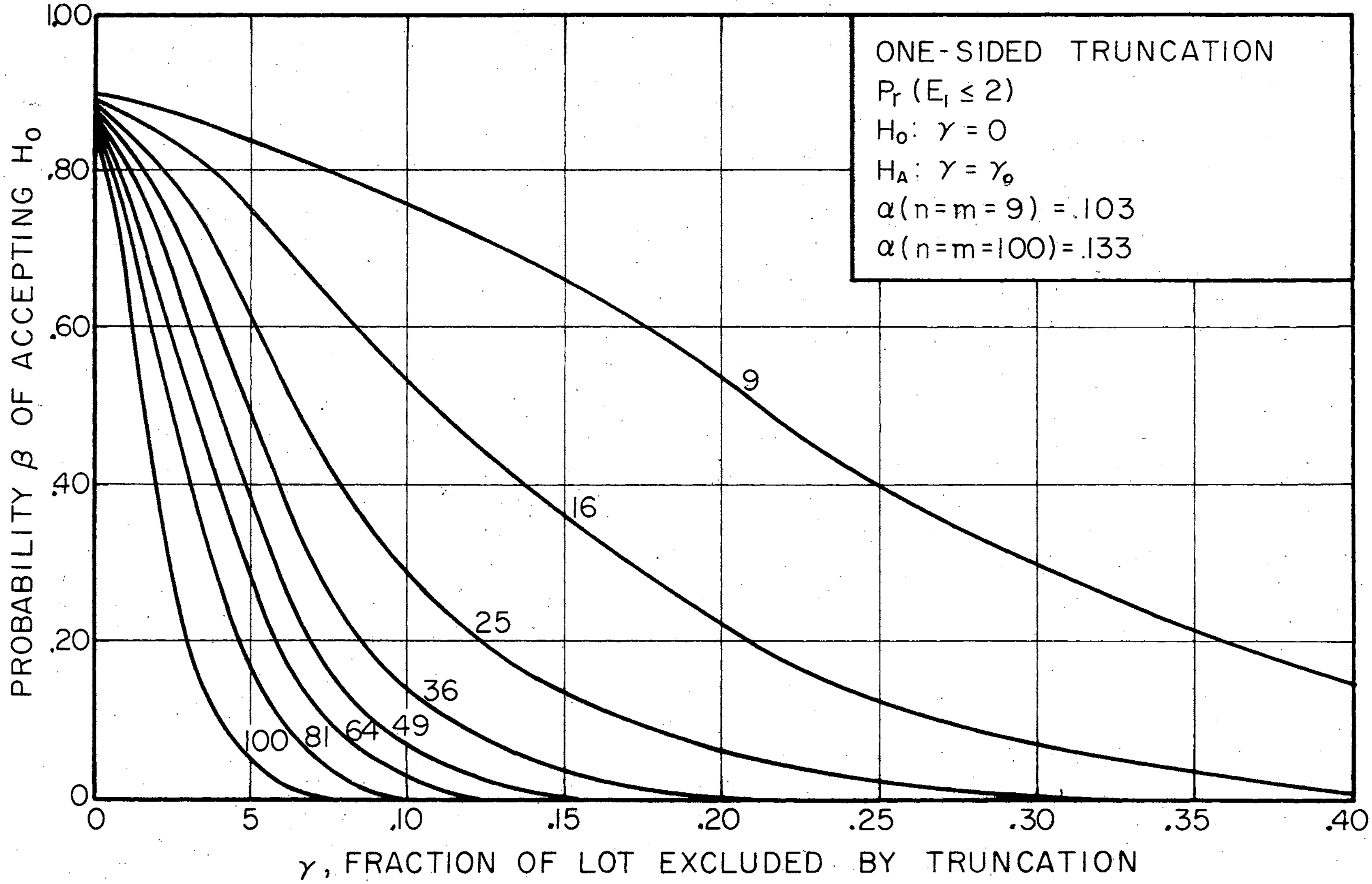

Figure 8. O.C. Curves of Exceedance Test, $E_1 \leq 2$, for One-Sided Truncation

range. Because of the discreteness problem it is not generally possible to choose an exact predetermined significance level for any given sample size. For the exceedance test, which appears to be most applicable to the detection of truncation, several possible acceptance numbers are analyzed for each sample size herein; for the other tests, only a representative set of acceptance numbers, one per each sample size, is presented.

O.C. Curves for Tests of One-Sided Truncation

The operating characteristic curves portrayed in Figure 8 are for the E_1 test with an acceptance number of two; that is, one would accept the null hypothesis of no truncation if two or fewer observations in the customer sample exceed the largest observation in the supplier sample.

The exceedance test significance probability is surprisingly insensitive to sample size. It is noted that all the curves in Figure 8 are for the same acceptance number; although sample size varies from 9 to 100 the significance probability varies only from 0.103 at $n = m = 9$ to 0.133 at $n = m = 100$. This property of the exceedance test makes it feasible to hold the acceptance number constant for a given chart. In examining Figure 8 it is seen that in order to have approximately a 10% chance both Type I and Type II error, detection of truncation of degree .10 requires samples of about size 40, and detection of truncation of degree 0.05 requires samples of about size 90.

When the acceptance number is increased from two to three, the average significance probability is reduced from .118 to .051, and the corresponding O.C. curves all have higher Type II error probability.

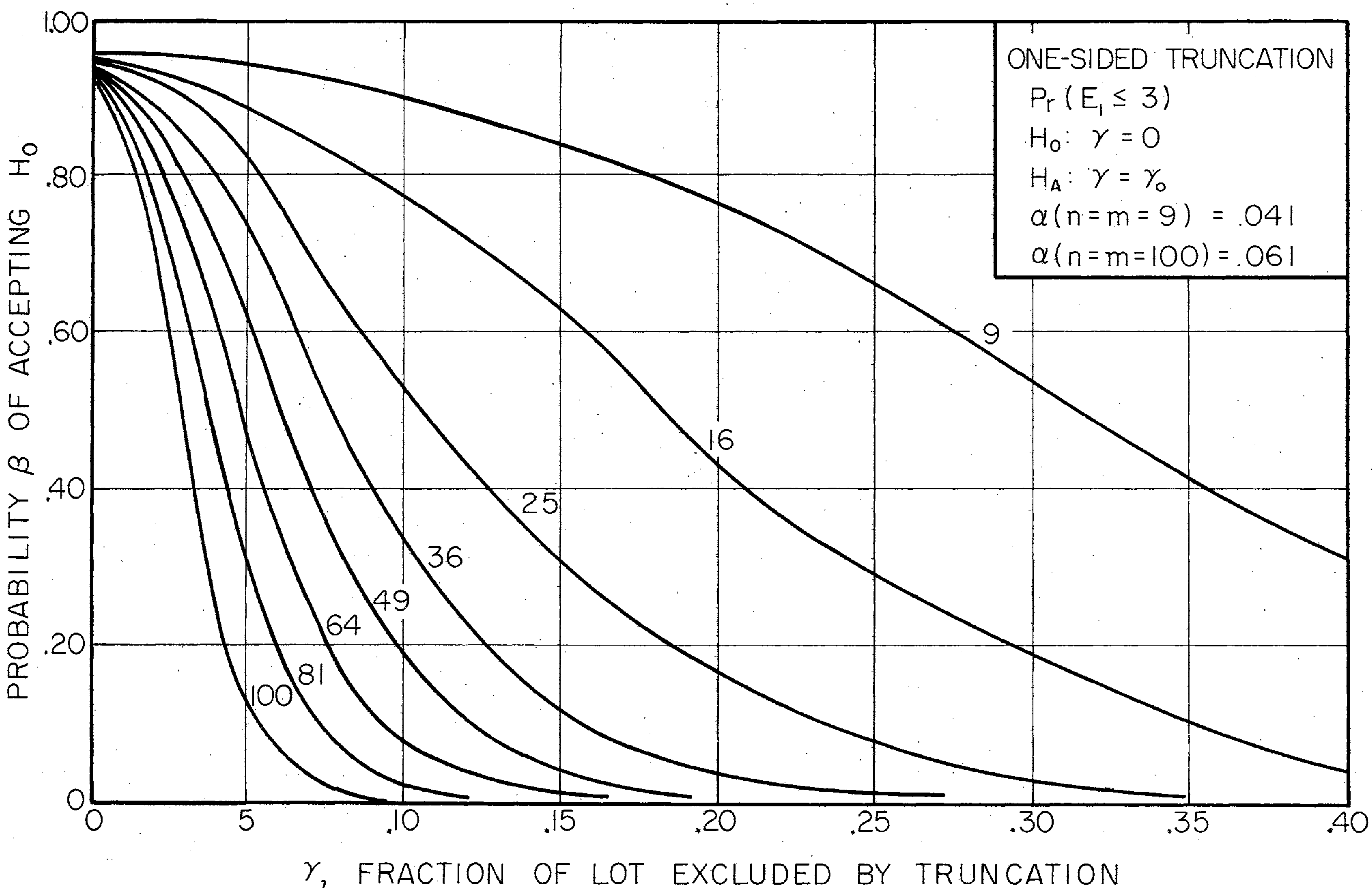

Figure 9. O.C. Curves of Exceedance Test, $E_1 \le 3$, for One-Sided Truncation

This is portrayed in Figure 9. Now, in this case it requires samples of about 60 to detect 10% truncation with error probability of .10, and over 100 to detect 5% truncation.

In Figure 10 a set of O.C. curves of the maximum difference (Kolmogorov-Smirnov) test is displayed. The significance probabilities range from 0.12 for samples of size 9 to 0.04 for samples of size 100.

The acceptance numbers portrayed vary from (K = 3) for (n = 9) to (K = 17) for (n = 100). The acceptance numbers as displayed tend to force the significance probability lower as sample size gets larger. Thus the O.C. curves all tend to cross, as the larger sample sizes have greater discriminating power and also smaller significance probabilities. Upon examining the curves it is seen that, for instance, sample size 100 acceptance number 17 would provide Type I error probability of 0.04 and Type II error probability about 0.10 for degree of truncation 0.20.

The rank-sum (Mann-Whitney) test for one-sided truncation is exhibited in Figure 11. The rank sum statistic has a discrete probability function as do the two previous tests; however, the sample space is much larger for the rank-sum test and therefore a pre-specified significance level can usually be obtained. Thus all the curves in Figure 11 have a significance probability of almost exactly 0.05.

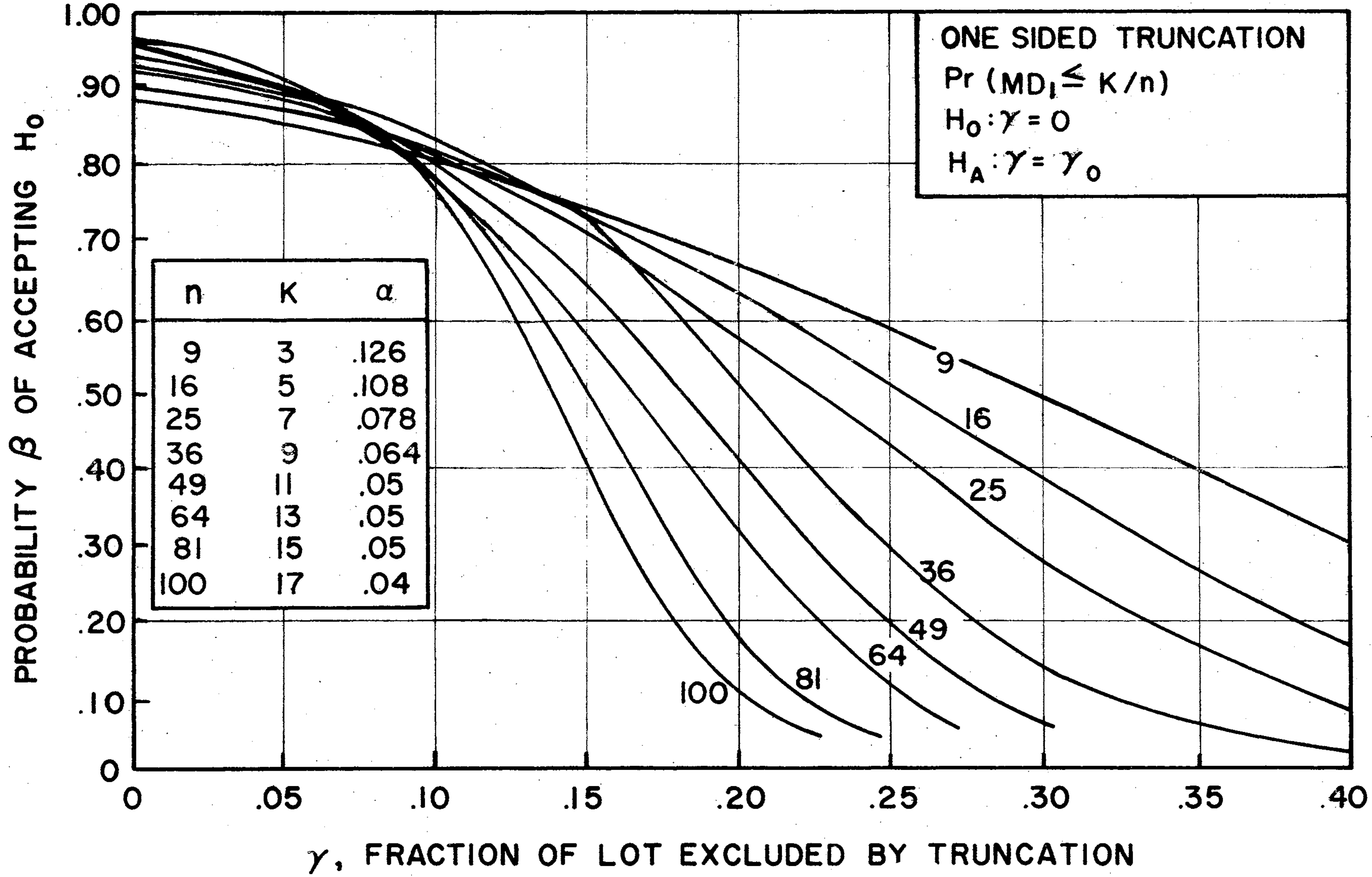

Figure 10. O.C. Curves of Maximum Difference Test for One-Sided Truncation

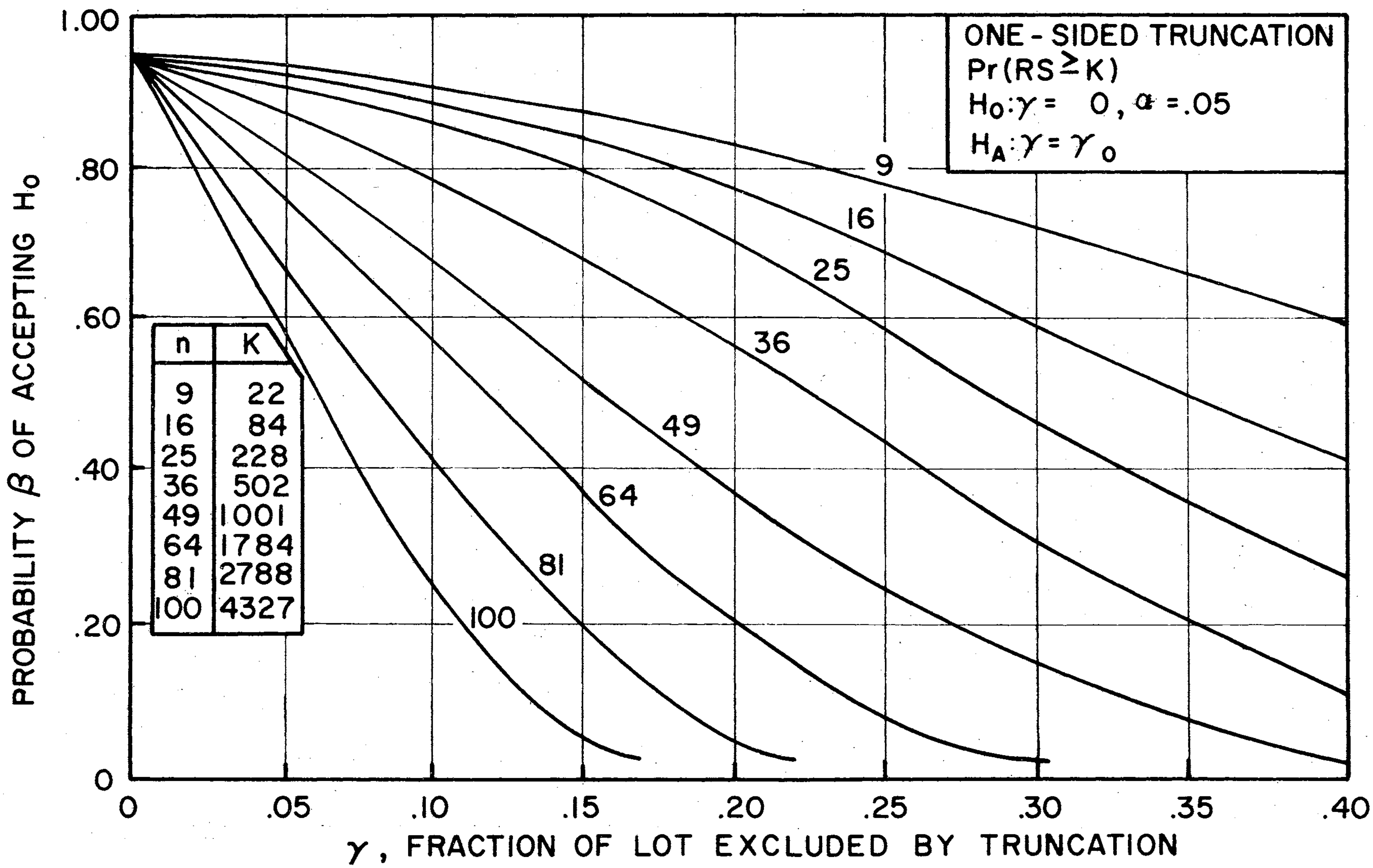

Figure 11. O.C. Curves of Rank-Sum Test for One-Sided Truncation

O.C. Curves for Tests of Two-sided Truncation

The O.C. curves for tests of two-sided truncation are displayed in Figures 12 through 15. The first three figures are for the two-sided exceedance test, acceptance numbers three, four, and five, respectively. These figures are self-explanatory, subject to the same interpretation as was offered Figures 8 and 9 previously. The reason three figures, rather than two, are presented for the exceedance test, is that possible values of the two-sided exceedance statistic are somewhat more densely spaced than in the one-sided case, and it was felt desirable for this particular test to present full results in the range of interest of significance probability. With the three figures we have full coverage of possible significance probabilities between 0.06 and 0.147, and partial coverage from 0.025 to 0.06 and from 0.147 to 0.185.

The two-sided maximum difference test is displayed in Figure 12. It should be noted that the maximum difference test, while theoretically admissible, is exceptionally poor in detecting two-sided truncation. It would exhibit much the same weakness in any case where it was applied to distributions with equal average values and unequal degrees of variation. As for truncation, symmetrical two-sided truncation doubles the degree of truncation without increasing the amount of difference in the cumulative distribution at either end. Thus, such a data discrepancy requires essentially double the degree of truncation in the two-sided case to have the same probability of detection as the one-sided test would give to one-sided truncation. Figure 12 shows that the O.C. curves are quire flat, compared to those for other tests portrayed in this chapter and the previous chapter. For samples of size 9, 16, and

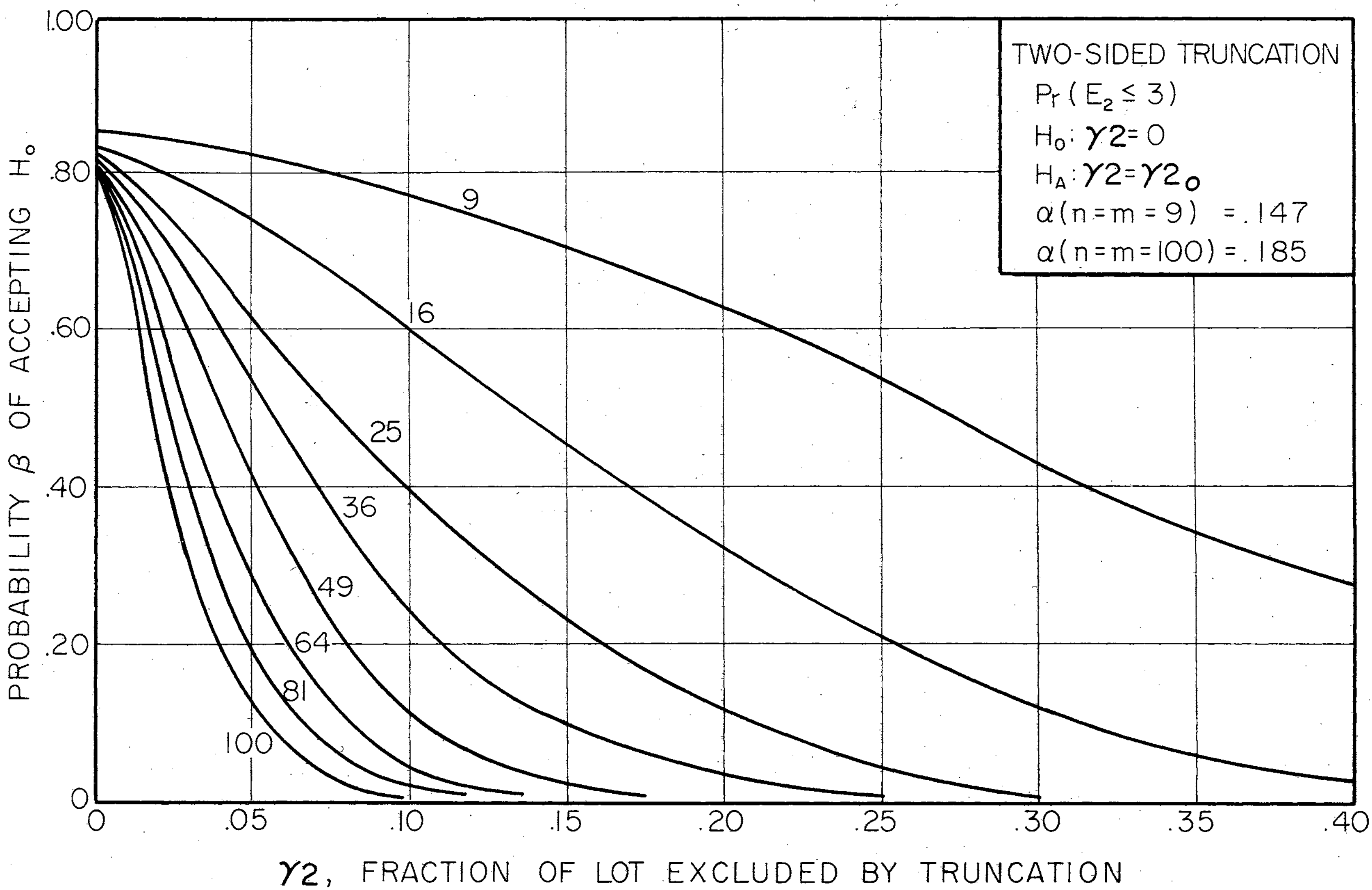

Figure 12. O.C. Curves of Exceedance Test, $E_2 \leq 3$, for Two-Sided Truncation

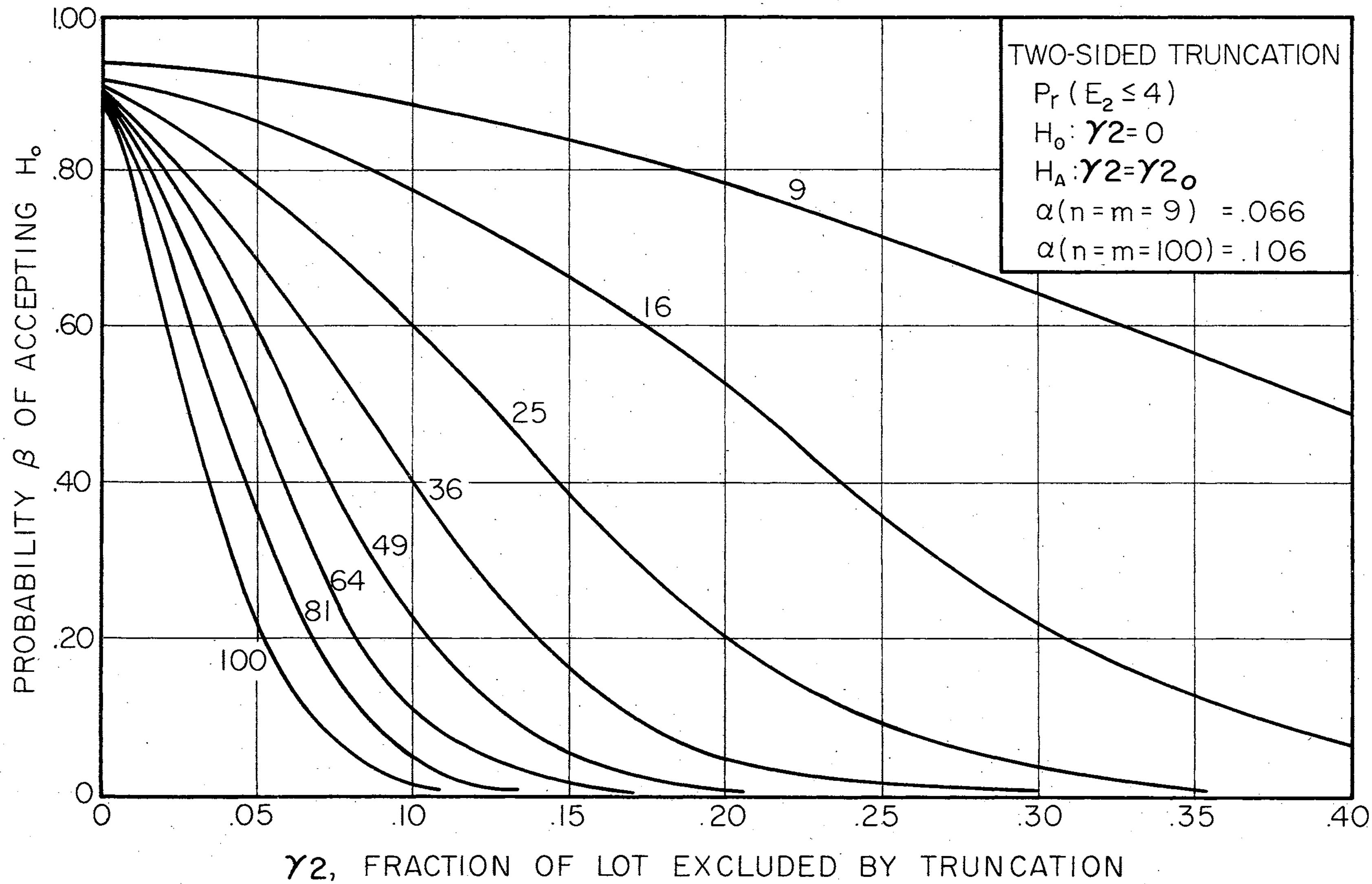

Figure 13. O.C. Curves of Exceedance Test, $E_2 \leq 4$, for Two-Sided Truncation

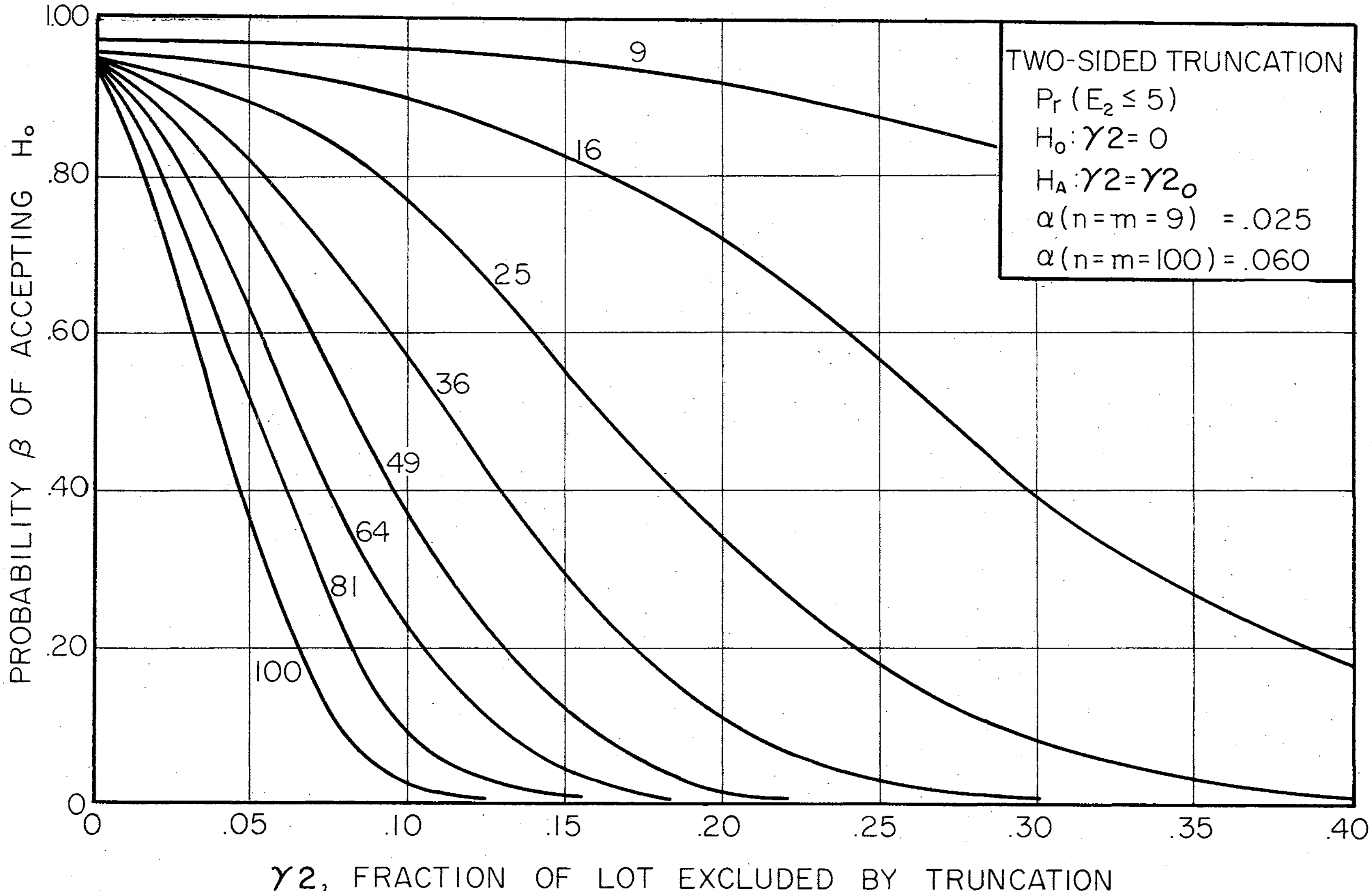

Figure 14. O.C. Curves of Exceedance Test, $E_2 \leq 5$, for Two-Sided Truncation

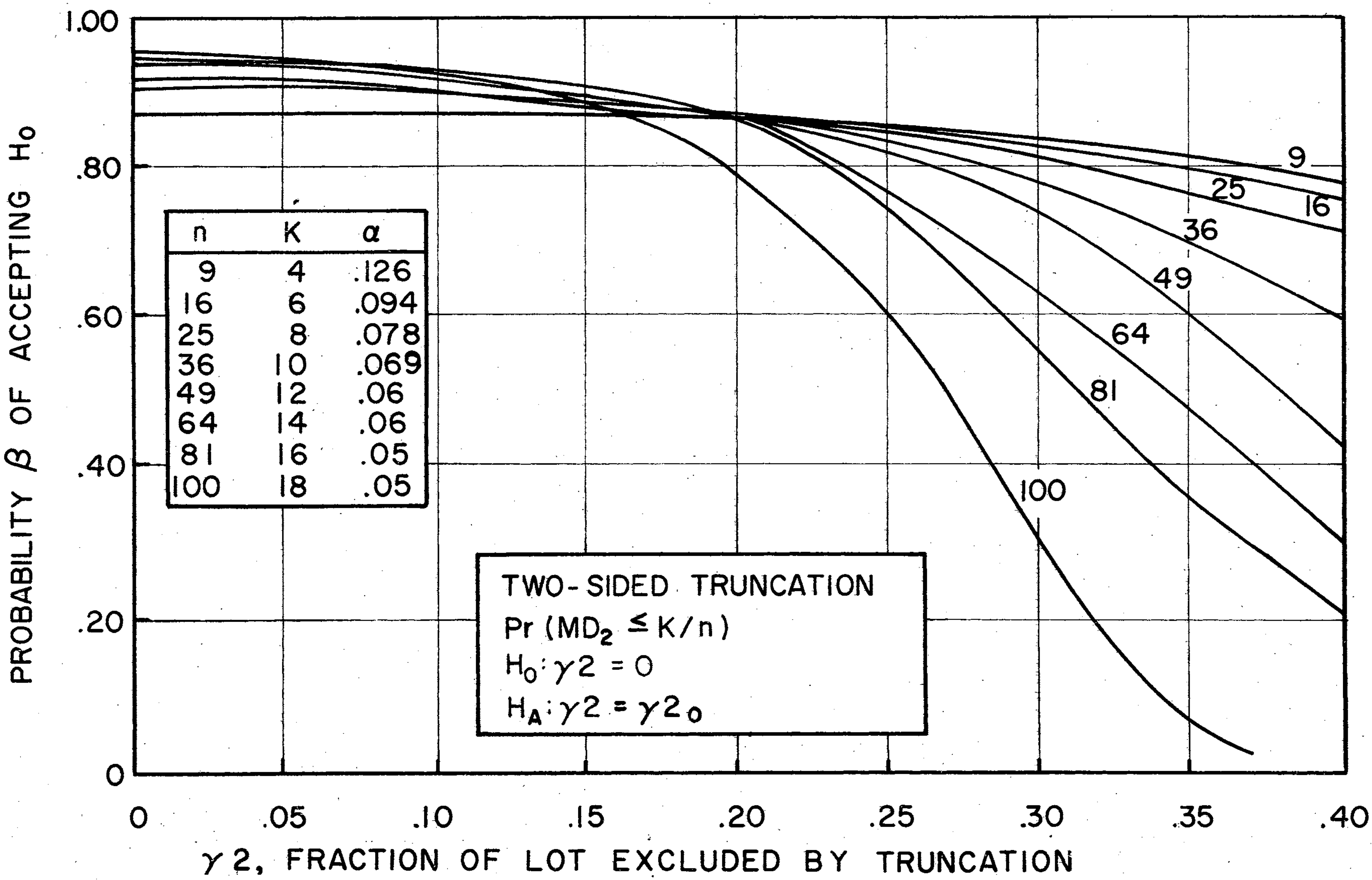

Figure 15. O.C. Curves of Maximum Difference Test, Two-Sided, for Two-Sided Truncation

25, the O.C. curves are almost insensitive to truncation of degree less than 0.40 and only samples of size 100 achieve a probability of Type II error as low as 0.10 for any degree of truncation displayed on the figure.

CHAPTER VI

COMPARISON OF THE TESTS FOR TRUNCATION

The operating characteristic curves have now been presented for all tests of truncation studied in this dissertation. It is appropriate to compare these tests on the basis of the ease of applying them in operational situations, and upon their relative effectiveness in detecting truncation when it exists.

Ease of Application

The relative ease of application of the tests depends in some part on the method of computation utilized. If a validating organization is set up in such a way as to routinely prepare computer input forms for analysis, then there is probably not a great deal of difference among the tests. Given the same sample size, they will all require about the same amount of work in preparing input forms, and the computer time required for all computations will be slight, compared to the input preparation work. However, much of the time validation will be performed by a roving field inspector who visits various supplier plants and performs inspections and analyses manually. In such circumstances the training time required for the inspector, who is generally not statistically sophisticated, and the time required to perform the tests, become important factors.

Of all the tests considered, the simplest to apply manually is certainly the exceedance test. This test requires only that the largest observation in the supplier sample be identified, and then a count made of the number of customer observations exceeding it (for the case of a one-sided test on the upper tail.) The most difficult to apply is the rank-sum test. Ranking a large group of numbers is a tedious procedure done manually. Midway between the exceedance and the rank-sum tests in ease of application are the parametric U and F tests and the maximum difference test. Both these types of tests can be performed by straightforward application of the test statistic formulas, or by a "grouped data" method which will materially reduce the computation time required for manual analysis. Details concerning this method can be found, e.g., in Bowker and Lieberman, (1959, pp. 4-10) for computation of sample mean and standard deviation, and in Berger (1966, pp. 7-8) for computation of the maximum difference. Experimental application of the tests by data reduction clerks and college students, however, indicate that even with the time saved by grouping data, the latter tests are still considerably more time-consuming than the exceedance test.

Relative Power

By reference to the O.C. curves presented in Chapters IV and V a number of comparisons of the power of the different tests can be made. One basis for comparison is to select certain probabilities of Type I and Type II errors, and then find the degree of truncation for which the different tests will detect with the given probabilities. For example, in Figure 16, the four different tests for one-sided truncation are displayed, with sample size 49 and alpha of 0.05. The

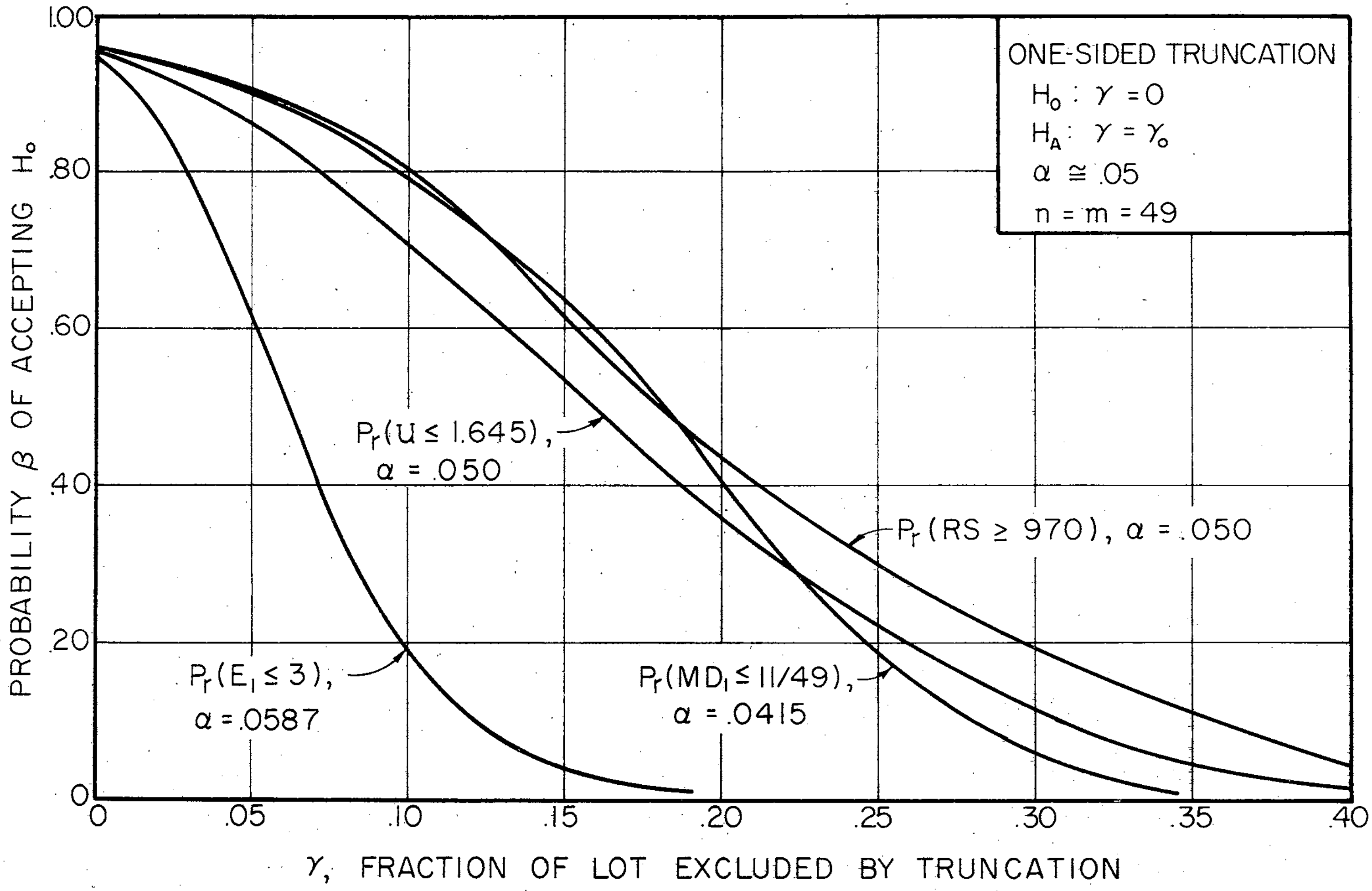

Figure 16. Comparison of Four Tests for One-Sided Truncation, Significance Level and Sample Size Held Constant

superiority of the exceedance test in this case is clear. For $\beta = 0.10$, the exceedane test will detect truncation of degree 0.12, whereas the maximum difference test will detect truncation of degree 0.26, the normal test degree 0.30 and the rank-sum test degree 0.35. Similar comparisons can be made for other degrees of truncation and acceptance probabilities. Viewing Figure 16 as a whole, the conclusion is that there is not a great deal of difference among the latter three tests, but that there is considerable difference between them as a group and the exceedance test.

The three tests for two-sided truncation are displayed in Figure 17, for the sample size 49 and significance probability 0.05. The most striking point of interest in this diagram is the extreme relative weakness of the maximum difference test. Even with truncation of degree 0.40 there is a 0.50 probability of accepting the hypothesis of no truncation. However both the other tests clearly would have zero probability of acceptance in such a case. Thus it can be seen that the maximum difference test, while comparable at least to the normal and rank-sum tests for one-sided truncation, is completely unsuitable for detection of two-sided truncation.

In comparing the F test and the exceedance test, little difference is noted between the two for truncation of less than about 0.08. As truncation increases beyond this level, the relative advantage of the exceedance test increases, and whereas the exceedance test will detect truncation of degree 0.16 with 0.10 probability of Type II error, the F test is indicated to have about 0.22 probability of Type II error, and in order for the Type II error probability to be reduced to 0.10, the degree of truncation would have to be about 0.22.

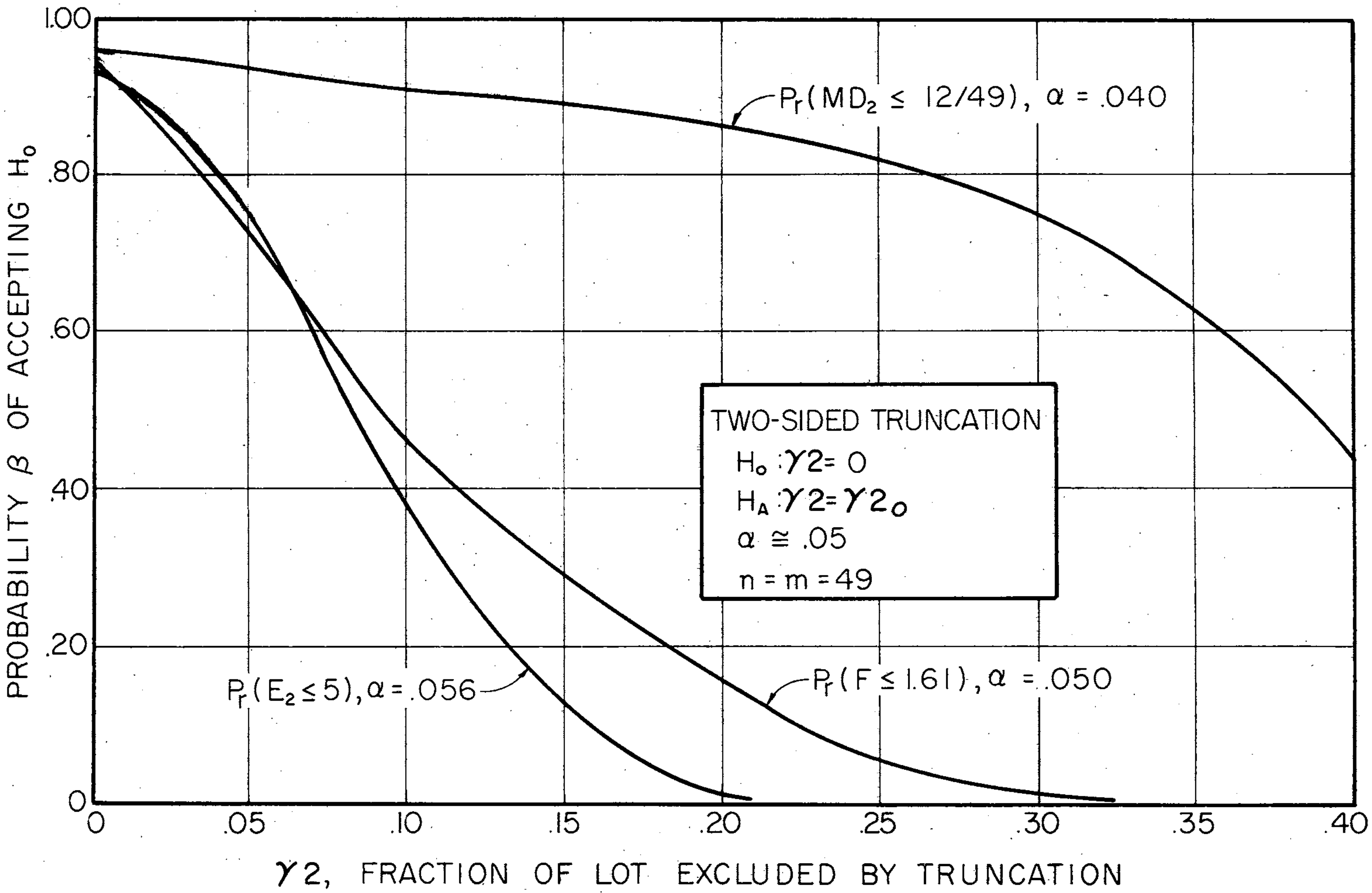

Figure 17. Comparison of Three Tests for Two-Sided Truncation, Significance Level and Sample Size Held Constant

For truncation of degree zero to about 0.06, there is a slight advantage to the F test. However, the relative difference, when compared with inherent sampling errors for the estimation of the curve for the exceedance test, and approximation errors for the derivation of the F test, is so slight as to be completely negligible from a practical standpoint.

Figures 16 and 17 are representative of a whole set of comparisons which can be derived from the figures of Chapters IV and V. All these comparisons will provide evidence of the consistently higher power of the exceedance tests when compared to others studied.

CHAPTER VII

SUMMARY AND CONCLUSIONS

Summary

The problems associated with data validation are directly related to the data discrepancies which can be postulated to occur in supplier data. It is customary for suppliers to do sampling inspection in generating variables-type data. In such situations, the discrepancies which may occur are divided into two mutually exclusive classes, viz: (1) measurement and (2) sampling.

Associated with this dichotomy of data discrepancies is a similar one between basic approaches of data analysis: paired samples and independent samples. For detection of measurement discrepancies, either repeat measurement of units previously inspected by the supplier, followed by paired-sample analysis; or independent sample analysis can be used by the customer. But for detecting sampling discrepancies, only independent samples can be used.

Reference has been made to a portion of the large body of literature on experimental design which indicates the relative merits of independent and paired-sample analysis for measurement discrepancies.

Little has been published about the properties of statistical tests for sampling discrepancies. Therefore the major portion of this research is devoted to (1) the construction of a mathematical model of a certain type of sampling discrepancy, which is referred to as truncation in sample selection; and (2) to the development of the operating characteristic curves of several different statistical tests for truncation. Two of the O.C. curves are derived analytically, the others are estimated by the technique of distribution sampling. Comparisons among the different tests are made.

Conclusions

Several conclusions follow from this research. The first conclusion is that independent samples are not well suited to the overall solution to the data validation problem. This conclusion contradicts previously published doctrine on validation of attributes data which indicates that only by the use of independent samples can the customer make a fair appraisal of the supplier results. This contradiction is explained by the fact that previous research did not explicitly recognize the dichotomy of data discrepancies. Nor was there any actual knowledge of the magnitude of the problem of detecting sampling discrepancies by purely statistical means.

It is concluded that where circumstances dictate that sampling discrepancies are not a concern of the customer (e.g. 100% supplier inspection; customer selection of supplier sample) data validation can be accomplished efficiently, economically, and with a significant reduction in the size of the customer sample as compared to conventional acceptance sampling procedures.

It is concluded that, of those tests for truncation which have been studied, the exceedance test is the best. This was found to be true for both one-sided and two-sided truncation.

A general conclusion can be made about sampling discrepancies. If a supplier should feel inclined to report on the quality of his lot by the use of a biased sample, then truncation in sample selection is a particularly effective way for him to do so, for two reasons. First, it is quite simple to implement, in comparison with some other possible methods. Secondly, it is extremely difficult to detect by purely statistical means. The basis for the latter statement is the set of O.C. curves presented in Chapters IV and V which indicate that for moderate amounts of truncation, in the order of five to ten percent, relatively large samples, in the order of 81 to 100 for both the supplier and customer would be required to have probability of both Type I and Type II errors in the order of 0.05 or less. This should be recognized as larger than most conventional acceptance samples, and thus relatively unattractive.

In view of the difficulty of statistical detection of truncation with small samples, there is only one possible legitimate argument for the use of small, independent samples in a data validation program. This is the psychological effect such sampling may have on suppliers. They can be informed by the customer's quality organization that they are being monitored for sampling discrepancies. Such independent samples could be drawn on random occasions, and could be integrated into a data validation program for measurement discrepancies without either a great deal of cost or of significance attached to the results.

The important conclusion from comparison of the O.C. curves for all tests studied, is that the exceedance test in both its versions is the best for detecting the two versions of truncation considered here. This further suggests that the exceedance-type tests which have been developed may deserve more attention than they have received up to this time from statisticians and quality control personnel.

Recommendations for Further Research

Validation of measurement-type data is a quality control technique about which very little research has been conducted. It appears that fruitful research could be pursued along several lines.

The use of equal sample sizes in this research is a valid means for establishing the relative power of the different tests. However, in actual practice a customer may have a supplier sample of a certain size which he wishes to validate for sampling discrepancies. He needs to know the sample size required to detect a certain amount of truncation. Thus the research can be extended by establishing O.C. functions for various sample size combinations.

Only one form of sampling discrepancy has been considered here. Explicit analysis of other forms of sampling discrepancies such as salting and tampering, and the methods required to detect them, should be made.

This research has established that the exceedance test is more powerful than the normal test for detecting one-sided truncation. However, it is well known that the normal test is optimum for detecting a shift in the mean. Thus there is some combination of shift and truncation for which both tests would have equal power. An investigation of this "indifference point" could yield interesting results.

Finally, this is, to the best of the author's knowledge, the first time that the exclusive use of independent samples for data validation has been questioned. Thus it appears that similar research for the case of attributes data could be performed to determine if these procedures could be improved by careful study of paired sample analysis of attribute-type inspection data.

A SELECTED BIBLIOGRAPHY

Bell, S. and D. B. Holdridge. *The Random Number Generators for the Sandia 3600 and Their Statistical Properties.* Sandia Corporation Research Report, SC-RR-67-65. Albuquerque, New Mexico: Sandia Corporation, 1967.

Berger, Roger. *Use of the Kolmogorov-Smirnov Statistic to Validate Supplier's Inspection Results.* Technical Memorandum, SC-TM-66-613. Albuquerque, New Mexico: Sandia Corporation, 1967.

Bowker, A. H. and G. J. Lieberman. *Engineering Statistics.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1959.

Cohen, A. C. "Estimating the Mean and Variance of Normal Populations From Singly Truncated and Doubly Truncated Samples." *Annals of Mathematical Statistics,* Vol. 21 (1950), 557-569.

Clark, Frank Eugene. "Truncation to Meet Requirements on Means." *Journal of American Statistical Association,* Vol. 52 (1957), 527-536.

Cochran, W. G., and Cox, Gertrude M. *Experimental Designs, 2nd. Ed.* New York: John Wiley and Sons, Inc., 1957.

Cowden, Dudley J. *Statistical Methods in Quality Control.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1957.

Dixon, W. J. "Power Under Normality of Several Non-Parametric Tests." *Annals of Mathematical Statistics,* Vol. 24 (1954), 610-613.

Dixon, W. J. and D. Teichroew. *Some Sampling Results on the Power of Non-Parametric Tests Against Normal Alternatives.* Report, U. S. Department of Commerce. Washington, D. C.: U. S. Government Printing Office, 1954.

Dudley, C. B. "The Enforcement of Specifications." *ASTM Proceedings,* Vol. 7 (1907), 28.

Duncan, Acheson J. *Quality Control and Industrial Statistics, Revised Edition.* Homewood, Illinois: Richard D. Irwin, Inc., 1959.

Elner, Harry and Joseph Mandelson. "Statistical Procedures for Validating the Results of Sampling Inspection by Variables." Draft Copy, Unpublished, 1964.

Epstein, Benjamin. "Comparison of Some Non-Parametric Tests Against Normal Alternatives." *Journal of American Statistical Association*, Vol. 50 (1955), 894-900.

_____. "Tables for the Distribution of the Number of Exceedances." *Annals of Mathematical Statistics*, Vol. 25 (1954), 762-768.

Ferris, C. D., F. E. Grubbs, and C. L. Weaver. "Operating Characteristics for the Common Statistical Tests of Significance." *Annals of Mathematical Statistics*, Vol. 17 (1946), 178-197.

Gibbons, Jean D. "On the Power of Two-Sample Rank Tests on the Equality of Two Distribution Functions." *Journal of the Royal Statistical Society*, Series B, Vol. 26 (1964), 293-304.

Gumbel, E. J., and H. von Schelling. "The Distribution of the Number of Exceedances." *Annals of Mathematical Statistics*, Vol. 21 (1950), 247-262.

Hald, Anders. *Statistical Theory With Engineering Applications*. New York, New York: John Wiley and Sons, Inc., 1952.

Hansen, Bertrand. *Quality Control: Theory and Applications*. Englewood Cliffs, New Jersey: Prentice Hall, 1963.

Hodges, J. L., Jr. "The Significance Probability of the Smirnov Two-Sample Test." *Arkiv For Matematik*, Bond 3 nr 43 (1957), 469-486.

Lehman, E. L. "The Power of Rank Tests." *Annals of Mathematical Statistics*, Vol. 24 (1953), 23-43.

Mandel, John and T. W. Lashof. "The Interlaboratory Evaluation of Testing Methods." *ASTM Bulletin*, Vol. 239 (1959), 53-61.

Mandelson, Joseph. "Product Verification." *Industrial Quality Control*, Vol. 20 (1964), 8-11.

Mann, H. B. and D. R. Whitney. "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other." *Annals of Mathematical Statistics*, Vol. 11 (1949), 367-392.

Marsaglia, G., M. D. MacLaren, and T. A. Bray. "A Fast Procedure for Generating Normal Random Variables." *Communications of the Association for Computing Machinery*, Vol. 7 (1964), 4-10.

Massey, Frank J., Jr. "The Distribution of the Maximum Deviation Between Two Sample Cumulative Step Functions." *Annals of Mathematical Statistics*, Vol. 22 (1951), 125-128.

Nelson, Benjamin N. "Survey and Application of Interlaboratory Testing Techniques." *Industrial Quality Control*, Vol. 23 (1967), 554-559.

Owen, Donald B. *Handbook of Statistical Tables*. Reading, Mass.: Addison-Wesley Publishing Co., 1962.

Robson, D. S. And J. H. Whitlock. "Estimation of a Truncation Point." *Biometrika*, Vol. 51 (1964), 33-39.

Rosenbaum, S. "Some Non-Parametric Two-Sample Tests." *Journal of American Statistical Association*, Vol. 24 (1965), 1118-1127.

_____. "Tables for a Non-Parametric Test of Dispersion." *Annals of Mathematical Statistics*, Vol. 24 (1953), 663-668.

_____. "Tables for a Non-Parametric Test of Location." *Annals of Mathematical Statistics*, Vol. 25 (1954), 146-150.

Siegel, Sidney. *Non-Parametric Statistics for the Behavioral Sciences*. New York, New York: McGraw-Hill Book Co., Inc., 1956.

Smirnov, N. "On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples." *Bulletin Mathematical, University of Moscow*, Vol. 2 (1939).

Stevens, W. L. "The Truncated Normal Distribution." Appendix to "Calculation of the Time-Mortality Curve," by C. I. Bliss. *Annals of Applied Biology*, Vol. 24 (1937), 846-852.

Teichroew, D. "Empirical Power Functions for Non-Parametric Two Sample Tests for Small Samples." *Annals of Mathematical Statistics*, Vol. 26 (1955), 340-344.

Tukey, J. W., "A Quick, Compact, Two-Sample Non-Parametric Test to Duckworth's Specifications." *Technometrics*, Vol. 1 (1959), 31-48.

U. S. Department of Defense. *Department of Defense Procurement Inspection Policies and Procedures for Items Covered by Military and Federal Specifications*. Washington, D. C. : U. S. Government Printing Office, 1967.

_____. *Department of Defense Quality Assurance and Policy*. Instruction 4155.6. Washington, D.C.: U.S. Government Printing Office, 1954.

_____. *Statistical Procedures for Determining Validity of Supplier's Attributes Inspection*. Washington, D.C.:U.S. Government Printing Office, 1960.

_____. *Quality Control and Reliability Handbook (Interim) H109, Statistical Procedures for Determining Validity of Suppliers Attributes Inspection*. Washington, D.C.: U.S. Government Printing Office, 1960.

Van der Laan, P. and T. Oosterhoff. "Monte Carlo Estimation of the Powers of the Distribution-Free Two-Sample Tests of Wilcoxon, van der Waerden and Terry and Comparison of These Powers." *Statistica Neerlandica*, Vol. 19 (1965), 265.

Whittlemore, H. L. *Ideas on Specifications*. New York, New York: Columbia Graphs, 1952.

Wilcoxon, F. "Individual Comparisons by Ranking Methods." *Biometrics*, Vol. 7 (1945), 80-83.

Wilks, Samuel S. "Statistical Prediction With Special Reference to the Problem of Tolerance Limits." *Annals of Mathematical Statistics*, Vol. 13 (1942), 400-409.

APPENDIX A

DISTRIBUTION SAMPLING PROGRAM

All nonparametric test O.C. curves presented in this dissertation were estimated by means of the distribution sampling program described in Chapter V and listed here. The program consists of a main program, STASIM (an acronym for STAtistical SIMulation), and subroutines SIFT (a catchword for a rank-ordering routine), NORM, (for NORMal distribution generator), and XBASD (an acronym for X-Bar And Standard Deviation).

The main program STASIM provides for all input and output, controls the use of the normal variate generator, and computes the values of all the tests statistics under study. The choice of one-sided or two-sided truncation and one-sided and two-sided tests is controlled by changing appropriate statements in the main program. The determination of run length, test statistic sample size, truncation point, is made through input data cards.

The subroutine SIFT is used at several points to rank a set of observation in numerical order. The normal random variates are generated by the use of a library function, ANRV, described in Chapter V. This function is part of the NORM subroutine which controls the mean, variance, and truncation point of the distributions being sampled. The XBASD subroutine is used to compute the mean and standard deviation of the sampling distribution of the MD and E tests.

```
      PROGRAM STASIM
C
C
      TYPE INTEGER CUM
      DIMENSION W(200),X(100),Y(100),XANDY(200),KTSUMY(200),
     1KDIF(200),PARM(6),ISUM(100),ICT(100),CUM(100),
     2,KUM(100),F(10)
     3KTSUMX(200),JSUM(100),CUMP(100)
      DIMENSION MW(500),IROG(500),ITSUM(500),ICUMF(500),
     1ACUMF(500)
      DIMENSION AMW(500)
C
C  READ JOB TITLE
C
      READ(5,2) F
    2 FORMAT(10A8)
C
C  PRINT JOB TITLE
C
      WRITE(6,1) F
    1 FORMAT(10A8 //)
      PARM(1)=0.
      PARM(2)=0.
      PARM(4)=0.
      PARM(5)=0.
      PARM(6)=10.
C
C  READ NS, NT, AND PARAMETERS
C
C  NS IS SAMPLE SIZE
C  NT IS NUMBER OF TRIALS
    4 READ (5,555) NS, NT, PARM(3)
  555 FORMAT (2I2, F10.0)
C
C  TEST NS = 9999
C
      IF (NS-9999) 6, 3, 6
    3 CALL EXIT
C
C  PRINT TITLE, NS, NT
C
    6 WRITE (6,81)
   81 FORMAT (1H0)
      WRITE (6,8) NT,NS
    8 FORMAT(1H0,10X,I3,2X,7HTRIALS,1X,
     111HSAMPLE SIZE, 1X,I4)
      N = NS
C
C  INITIALIZE VARIABLES
C
      JZERO = 0
      DO 189 J = 1,N
  189 ICT(J) = J
      DO 193 MAXDIF = 1,N
      ISUM(MAXDIF) = 0
  193 CUM(MAXDIF) = 0
      KZERO = 0
      DO 194 KONT = 1, N
      JSUM (KONT) = 0
  194 KUM(KONT) = 0
      MZERO = 0
C
C  PRINT MODEL AND PARAMETERS
C
   11 WRITE(6,16) PARM(1),PARM(2),PARM(3),PARM(4)
   16 FORMAT(10X,13HNORMAL, MU1 =,F6.2,2X,12HTRUNCATED A
     15HSIGMA,2X,5HMU2 =,F6.2,2X,12HTRUNCATED AT,F6.2,2X)
C
C  REPEAT THE BASIC SIMULATION THRU ST. 500 NT TIMES
C
      DO 500 II=1,NT
C
C  DRAW RANDOM SAMPLE FROM F OF X
C
   15 CALL NORM (NS,PARM(1),PARM(2),X)
C
C  DRAW RANDOM SAMPLE FROM G OF Y
C
      CALL NORM (NS,PARM(3),PARM(4),Y)
C
C  RANK-ORDER THE SAMPLE DATA
C
   25 M=NS
      MS=2*NS
      DO 30 I=1,NS
      M=M+1
      W(I)=X(I)
      W(M)=Y(I)
   30 CONTINUE
      CALL SIFT (NS,X)
      CALL SIFT (NS,Y)
      CALL SIFT (MS,W)
      N = NS
      M = NS
```

```
C  X IS THE VECTOR ON WHICH TRUNCATION OCCURS
C      Y(I)   I SMALLEST TO LARGEST I=1,M
C      X(J)   J SMALLEST TO LARGEST J = 1,N
C
C  COUNT NUMBER OF EXCEEDANCES
C
      KONT = 0
      DO 701 I=1,M
      IF(Y(I) - X(N)) 701,701,702
  702 KONT = KONT +1
  701 CONTINUE
C
C  ACCUMULATE RANK-SUM
C
C     MWT IS NON-SUBSCRIPTED EQUIV OF MW
      MWT = 0
      DO 620 I= 1,N
      DO 620 J= 1,M
      IF(Y(J)-X(I)) 610,610,620
  610 MWT = MWT + 1
  620 CONTINUE
      AMW(II)= MWT
      J=1
      D=W(1)
      DO 107 I=2,MS
      IF(D-W(I))105,110,110
  105 XANDY(J)=D
      D=W(I)
      J=J+1
  110 ED=W(I)
  107 CONTINUE
C
C  COMPUTE MAXIMUM DIFFERENCE BETWEEN CUMULATIVES
C
      XANDY(J)=ED
      KCX=1
      KCY=1
      KTX=0
      KTY=0
      DO 142 I=1,J
  112 IF(XANDY(I)-Y(KCY))120,115,120
  115 KCY=KCY+1
      KTY=KTY+1
      GO TO 112
  120 KTSUMY(I)=KTY
  125 IF(XANDY(I)-X(KCX))135,130,135
  130 KCX=KCX+1
      KTX=KTX+1
      GO TO 125
  135 KTSUMX(I)=KTX
      IF (KTSUMX(I)-KTSUMY(I)) GO TO 143
  141 KDIF(I)=0
      GO TO 142
  143 KDIF(I)=KTSUMX(I)-KTSUMY(I)
  142 CONTINUE
      DUM = KDIF(1)
      DO 150 I=2,J
      IF(DUM-KDIF(I)) 145,145,150
  145 DUM=KDIF(I)
  150 CONTINUE
      MAXDIF=DUM
      IF(MAXDIF) GO TO 199
      MZERO = MZERO +1
  199 CONTINUE
      ISUM(MAXDIF) = ISUM(MAXDIF)+1
      IF(KONT)1333,1333,1335
 1333 KZERO = KZERO + 1
 1334 GO TO 500
 1335 JSUM(KONT) = JSUM(KONT) + 1
C
C  END OF MAIN SIMULATION -DO- LOOP
C
  500 CONTINUE
C
C  CLASSIFY AND RANK ALL RANK-SUMS
C
      CALL SIFT(NT,AMW)
      DO 4005 KI=1,NT
 4005 MW(KI)= AMW(KI)
      J=1
      ID=MW(1)
      DO 401 I= 2,NT
      IF(ID-MW(I)) 402,403,403
  402 IROG(J)=ID
      ID= MW(I)
      J=J+1
  403 JD=MW(I)
  401 CONTINUE
C NOW HAVE DISTINCT MW VALUES. BUILD UP FREQUENCY COUNT
      IROG(J) = JD
      ICX=1
      DO 406 I = 1,J
      ITX = 0
  409 IF(IROG(I)-MW(ICX)) 407, 408, 407
  408 ICX = ICX +1
      ITX = ITX +1
```

```
      GO TO 409
  407 ITSUM(I) = ITX
  406 CONTINUE
C NOW HAVE CELLS IROG(I) WITH FREQ. ITSUM(I)
      ICUMF(1)=ITSUM(1)
      DO 410 I=2,J
  410 ICUMF(I)=ICUMF(I-1)+ ITSUM(I)
      ANT = NT
      DO 411 I=1,J
      ACUMF(I) = ICUMF(I)
  411 CUMP(I) = ACUMF(I)/ANT
C
C  PRINT RANK-SUM RESULTS
C
      WRITE (6,4165) J
 4165 FORMAT (1H0,10HTHERE ARE ,I3,   20H DISTINCT RAN
     1K SUMS)
      RJ = J
      R = RJ/20.
      LOG = INT(R)
      RLOG = LOG
      LAG = 1
      LEG = 20
      IF(LOG-0) 416,416,417
  417 DO 418 L= 1,LOG
      WRITE(6,302)(IROG(I), I=LAG,LEG)
      WRITE(6,303)(ITSUM(I), I=LAG,LEG)
      WRITE(6,304)(ICUMF(I), I=LAG,LEG)
      LAG = 1+ L*20
      IF(LOG-L) 419, 419, 420
  420 LEG = 20 * (L+1)
  418 CONTINUE
  419 IF(J-20*LOG) 421,421,416
  416 LEG = J
      WRITE(6,302)(IROG(I), I=LAG,LEG)
      WRITE(6,303)(ITSUM(I), I=LAG,LEG)
      WRITE(6,304)(ICUMF(I), I=LAG,LEG)
  421 CONTINUE
  302 FORMAT(1H0,15HRANKSUM VALUE  , 20I5)
  303 FORMAT(1H ,15HFREQUENCY      , 20I5)
  304 FORMAT(1H ,15HCUM FREQUENCY  , 20I5)
  305 FORMAT(1H ,15HCUM PROB. FN.  , 20(1X,F4.3))
C
C  CUMULATE RESULTS FOR EXCEEDANCE AND MAXIMUM DIFFERENCE
C
      CUM(1) = ISUM(1) +MZERO
      DO 201 MAXDIF = 2, N
  201 CUM (MAXDIF) = CUM(MAXDIF-1)+ISUM(MAXDIF)
      KUM(1) = KZERO + JSUM(1)
      DO 1338 KONT = 2,N
 1338 KUM(KONT) = KUM(KONT-1) +JSUM(KONT)
C
C  COMPUTE MEAN  AND STD DEV FOR EXCEEDANCE AND MAXIMUM DIF
C
         CALL XBASD(N, NT,JSUM,WBAR,RMR)
         CALL XBASD(N, NT,ISUM,XBAR,RMS)
C
C  PRINT RESULTS FOR EXCEEDANCE AND MAXDIF
C
      WRITE (6,291) JZERO,(ICT(J),J=1,N)
      WRITE(6,2915)
      WRITE(6,292) KZERO,(JSUM(KONT),KONT=1,N)
      WRITE(6,296) WBAR, RMR
      WRITE(6,293) KZERO,(KUM(KONT),KONT=1,N)
      WRITE(6,2925)
      WRITE(6,294) MZERO,(ISUM(MAXDIF), MAXDIF=1,N)
      WRITE (6,295) MZERO,(CUM(M),M=1,N)
      WRITE(6,296) XBAR, RMS
  291 FORMAT(1H0,12HARGUMENT    ,26I4)
 2915 FORMAT(117X,16H  MEAN   STDDEV )
  292 FORMAT(1H ,12HEXCEEDANCES ,26I4)
 2925 FORMAT(1H )
  293 FORMAT(1H ,12HCUM EXCEED  ,26I4)
  294 FORMAT(1H ,12HMAXDIFF     ,26I4)
  295 FORMAT(1H ,12HCUM MAXDIF  ,26I4)
  296 FORMAT(1H+,117X,2F7.2)
      GO TO 4
      END
```

```
C
C
C
      SUBROUTINE SIFT (N,X)
      DIMENSION X(40)
      M=N
    1 M=M/2
      IF(M) 3,2,3
    2 RETURN
    3 K=N-M
      J=1
    4 I=J
    5 L=I+M
      IF (X(I)-X(L))7,7,6
    6 A=X(I)
      X(I)=X(L)
      X(L)=A
      I=I-M
      IF(I)7,7,5
    7 J=J+1
      IF(J-K)4,4,1
      END
C
C
C
      SUBROUTINE NORM(N,X,Y,S)
C
C     N IS NUMBER OF VARIATES RETURNED
C     X IS MEAN OF DISTRIBUTION, NORMALLY ZERO
C     Y IS TRUNCATION CONSTANT ON UPPER TAIL
C     S IS THE VECTOR OF RANDOM VARIATES SELECTED
C     STD DEV IS FIXED AT ONE
      DIMENSION S(20)
      DO 10 I=1,N
    9 T = ANRV(0) + X
      IF(T-Y) 10,9,9
   10 S(I) = T
C     ANY VALUE OF T WHICH IS LESS THAN Y SHOULD BE ACCEPTED
      RETURN
      END
```

```
C
C
C
           SUBROUTINE XBASD(N, NT,I,X,R)
C
      DIMENSION I(50)
      S=0.0
      S2=0.0
      DO 5 K=1,N
      D=K
      F=I(K)
      S=S+F*D
    5 S2=S2+F*D*D
         A= NT
      X=S/A
      R=SQRTF(S2/A-X**2)
      RETURN
      END
```

APPENDIX B

VALIDITY OF THE NORMAL APPROXIMATION

In Chapter IV the distribution of $\bar{x}$ was taken to be approximately normal with mean μ_x and variance σ_x^2/n. According to the Central Limit Theorem this result holds exactly as $n \to \infty$. But it is generally felt to hold for relatively small sample sizes, as well. An empirical study was performed to determine how good the approximation really is. For this study, samples of size n = 2, 5, 10, 25, and 50, are considered along with degrees of truncation γ = .05, .10, .20, .40, .60. For each combination of sample size and degree of truncation, 1,000 trials are performed. Then the mean, variance, coefficient of skewness, and coefficient of kurtosis, are computed, according to equations presented below.

Duncan (1959, pp. 496-501) discusses the computation of moments of a frequency distribution and explains that the coefficients of skewness and kurtosis are zero for a normal distribution. Further details concerning computation and interpretation of skewness and kurtosis coefficients are available in Duncan's book. (Duncan uses the symbol γ_1 for CS as defined herein, and γ_2 for CK.)

Let v_j be the jth moment of the frequency distribution of $\bar{x}$'s about zero. In our case, the frequency distribution consists of 1,000 $\bar{x}$'s. So, letting N = 1,000,

$$v_j = \sum_{i=1}^{1000} (x_i)^j/1000$$

The first moment of the frequency distribution about zero is the mean of the frequency distribution, which should be, with 1,000 trials, a good estimate of the population mean. Letting the ^ over a parameter denote a sample estimate of the parameter,

$$\hat{\mu}_{\bar{x}} = \sum_{i=1}^{1000} \bar{x}_i / 1000.$$

The second moment about the mean is the variance,

$$\hat{\sigma}_{\bar{x}}^2 = \sum_{i=1}^{1000} (\bar{x}_i - \hat{\mu}_{\bar{x}})^2/1000 = v_2 - v_1^2$$

The third moment about the mean is a measure of skewness, but a more common measure is the third moment divided by σ^3. Let CS denote the coefficient of skewness. Then

$$\hat{CS}_{\bar{x}} = (v_3 - 3v_2v_1 + 2v_1^2) \; / \; (\hat{\sigma}_{\bar{x}}^2)^{3/2}$$

Finally, let the coefficient of kurtosis be denoted by CK. Then

$$\hat{CK}_{\bar{x}} = [v_4 - 4v_3v_1 + 6v_2v_1^2 - 3v_1^4) \; / \; (\hat{\sigma}_{\bar{x}}^2)^2] - 3.0$$

If the Central Limit Theorem held exactly then the following would be the result:

$$E(\hat{\mu}_{\bar{x}} - \mu) = 0$$
$$E[\hat{\sigma}_{\bar{x}}^2 / (\sigma_x^2/n)] = 1$$
$$E(\hat{CS}_{\bar{x}}) = 0$$
$$E(\hat{CK}_{\bar{x}}) = 0$$

The results of the sampling experiment showed that for moderate amounts of truncation in the order of 0.10 or less, samples of two or more gave a good approximation to the expected normal distribution. As the degree of truncation increased, so also did the required sample size.

The mean and standard deviation were consistently very close to their expected values, within the limits of sampling variation. The results for $\hat{CS}$ and $\hat{CK}$ are presented in Table III and Table IV. Both

TABLE III

SKEWNESS OF EMPIRICAL DISTRIBUTION OF $\bar{x}$

γ ↓ \ n→	2	5	10	25	50
.05	-.2846	-.1617	-.1070	.0454	-.0179
.10	-.3219	-.2372	-.2322	-.1069	-.0839
.20	-.5720	-.3488	-.0939	-.1743	-.1531
.40	-.5953	-.3807	-.2497	-.2905	-.1713
.60	-.6629	-.5430	-.2459	-.2820	-.1987

TABLE IV

KURTOSIS OF EMPIRICAL DISTRIBUTION OF $\bar{x}$

γ ↓ \ n→	2	5	10	25	50
.05	.0851	-.0994	.2318	.1828	-.0925
.10	.0111	-.0031	.0713	-.2105	.0550
.20	.4017	.1008	-.1481	-.0746	.0128
.40	.3549	.0151	-.1430	.2507	.0773
.60	.3968	.1742	.0569	-.1204	.0767

these measures should approach zero as sample size increases. In examining the tables it is seen that this is in fact the case.

It is difficult to say just what sample size is required for the normal approximation to be sufficiently valid since the degree of accuracy required will vary from one application to another. However, from examining the tables, a sample size of 10 appears to be adequate, inasmuch as very little improvement is noted for larger samples.

Another form of analysis was performed by using the Kolmogorov-Smirnov goodness-of-fit test, and comparing the simulated empirical distributions with normal distributions having the mean and variance which are predicted by the Central Limit Theorem.

The mechanics and rationale of the test are described by Siegel (1956, pp. 47-52). The results are portrayed in Table V. This shows that, as sample size increases, the one-sample maximum difference statistic, MD_o, decreases but does not approach zero. Actually, the expected value of MD_o is $1/\sqrt{N}$ if the empirical and theoretical distributions are identical, where N is the number of trials in the experiment. Thus the expected value of MD_o is $E(MD_o) = 1/\sqrt{1000} = .031$.

For samples of ten or better, all the values of MD_o range from 0.015 to 0.038, while for samples of two and five there are several values larger than 0.05.

The two different methods of analysis tend to support the conclusion that for samples of ten or greater the Central Limit Theorem can be safely applied.

TABLE V

MAXIMUM DIFFERENCE BETWEEN EMPIRICAL AND THEORETICAL

CUMULATIVE DISTRIBUTION FUNCTIONS OF $\bar{\bar{x}}$

$n \rightarrow$ / $\gamma \downarrow$	2	5	10	25	50
.05	.020	.026	.031	.017	.038
.10	.015	.070	.026	.016	.018
.20	.056	.038	.020	.026	.021
.40	.051	.039	.036	.034	.033
.60	.055	.052	.029	.015	.021

APPENDIX C

PRECISION OF ESTIMATED O.C. CURVES

In this experiment the primary goal was to generate O.C. curves which would have sufficient precision that they could be used to compare different tests and to choose appropriate sample sizes. The adequacy of precision of the O.C. curves for their intended purpose is established by two different methods.

One method is to compare O.C. curves estimated separately to check on the degree of agreement. Some typical results are summarized in Table VI and plotted in Figure 18. These are the results of three separate sequences on the one-sided exceedance statistic equal samples of size 25. The acceptance number chosen for the comparison is $E_1 = 3$. Of these three O.C. curves compared, two are programmed on equal increments of the truncation point k, while the third is incremented on the degree of truncation parameter γ. This is the reason for the difference in the point locations in the horizontal direction.

In inspecting Figure 18, it is seen that the three curves are in relatively close agreement. The largest discrepancy is about a 0.06 difference in $\hat{\beta}$ of the two extreme curves which occurs at approximately $\gamma = 0.07$ and $\gamma = 0.115$. Then at $\gamma = 0.23$ there is a 0.05 difference. The three curves all converge at the end points ($\gamma = 0$, $\beta = 0.945$) and ($\gamma = 0.40$, $\beta = 0$). By coincidence they are also equal at the point ($\gamma = 0.16$, $\beta = 0.26$).

Since four times as many trials per point, (or NT = 2000) would be required to reduce sampling error by 50%, and since the agreement exhibited in Figure 18 is sufficient for evaluation purposes, it is concluded that NT = 500 is an adequate number of trials.

Another method of evaluating the precision of results is to establish a confidence band about the estimated O.C. curve, by use of a

TABLE VI

COMPARISON OF THREE O.C. CURVE SAMPLING SEQUENCES

$E_1 \leq 3$, NT = 500, NS = 25. α = 0.055

k	γ	NL	$\hat{\beta}$
Run Sequence A			
3.0	.001	471	.942
2.0	.0228	455	.910
1.75	.04	429	.858
1.5	.0668	348	.696
1.25	.1056	255	.510
1.0	.1587	138	.276
0.75	.2266	30	.060
0.5	.3085	10	.002
0.25	.4013	1	.002
Run Sequence B			
3.0	.001	477	.954
2.0	.0228	458	.916
1.75	.04	436	.872
1.50	.0668	378	.756
1.25	.1056	248	.496
1.00	.1587	139	.278
0.75	.2266	53	.106
0.50	.3085	3	.006
0.0	.50	0	.000
Run Sequence C			
2.326	.01	471	.942
1.645	.05	421	.842
1.282	.10	295	.590
1.036	.15	157	.314
0.8416	.20	75	.150
0.6745	.25	29	.058
0.5244	.30	11	.022
0.2530	.40	1	.002

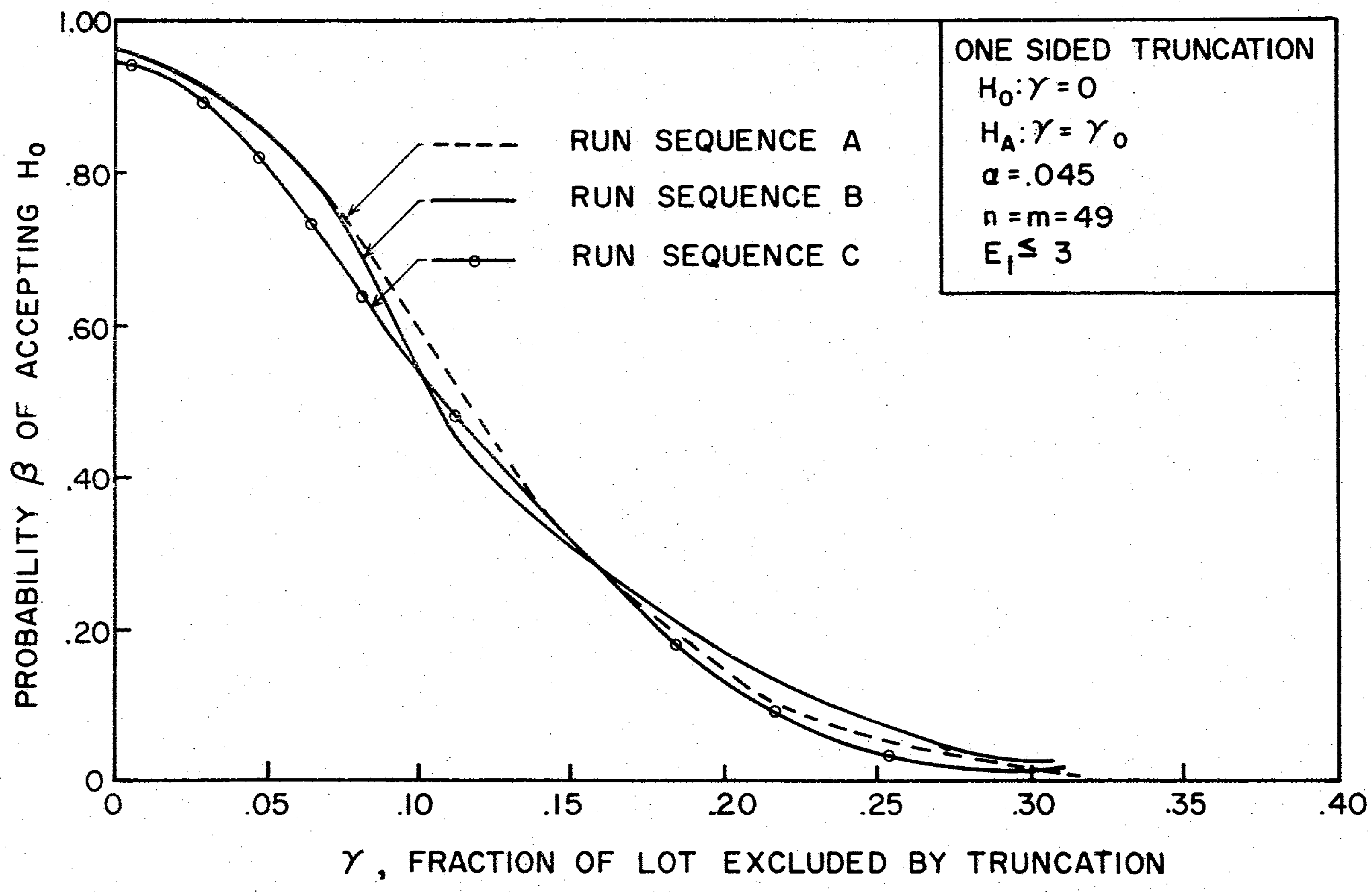

Figure 18. Comparison of Three O.C. Curve Sampling Sequences

normal approximation to the binomial distribution.

The confidence band was computed for three example O.C. curves: for NS = 25, NT = 1000, (Figure 19), for NS = 25, NT = 500, (Figure 20), and for NS = 100, NT = 200, (Figure 21).

It is known that the standard deviation of a binomial random variable with parameter p is

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

On the assumption that the binomial random variable $\hat{\beta}$ is normally distributed about its sample average $\hat{\beta}$ with standard deviation

$$\sigma_{\hat{\beta}} = \sqrt{\frac{\hat{\beta}(1-\hat{\beta})}{NT}} \quad ,$$

the following 95% confidence band results:

$$Pr(\beta_1 < \beta < \beta_2) = .95$$

where

$$\beta_1 = \hat{\beta} - 1.96\sigma_{\hat{\beta}}$$

and

$$\beta_2 = \hat{\beta} + 1.96\sigma_{\hat{\beta}}$$

Comparison of Figures 20 and 21 shows the logical justification for reducing the number of trials as sample size becomes larger. Because the curve in Figure 21 is so much steeper, the actual area enclosed in the 95% confidence band on is smaller than in Figure 20, although the ordinates of the confidence band are larger due to the smaller number of trials.

These confidence bands are only approximate, and are somewhat wider than the true bands would be, for the following reasons. There

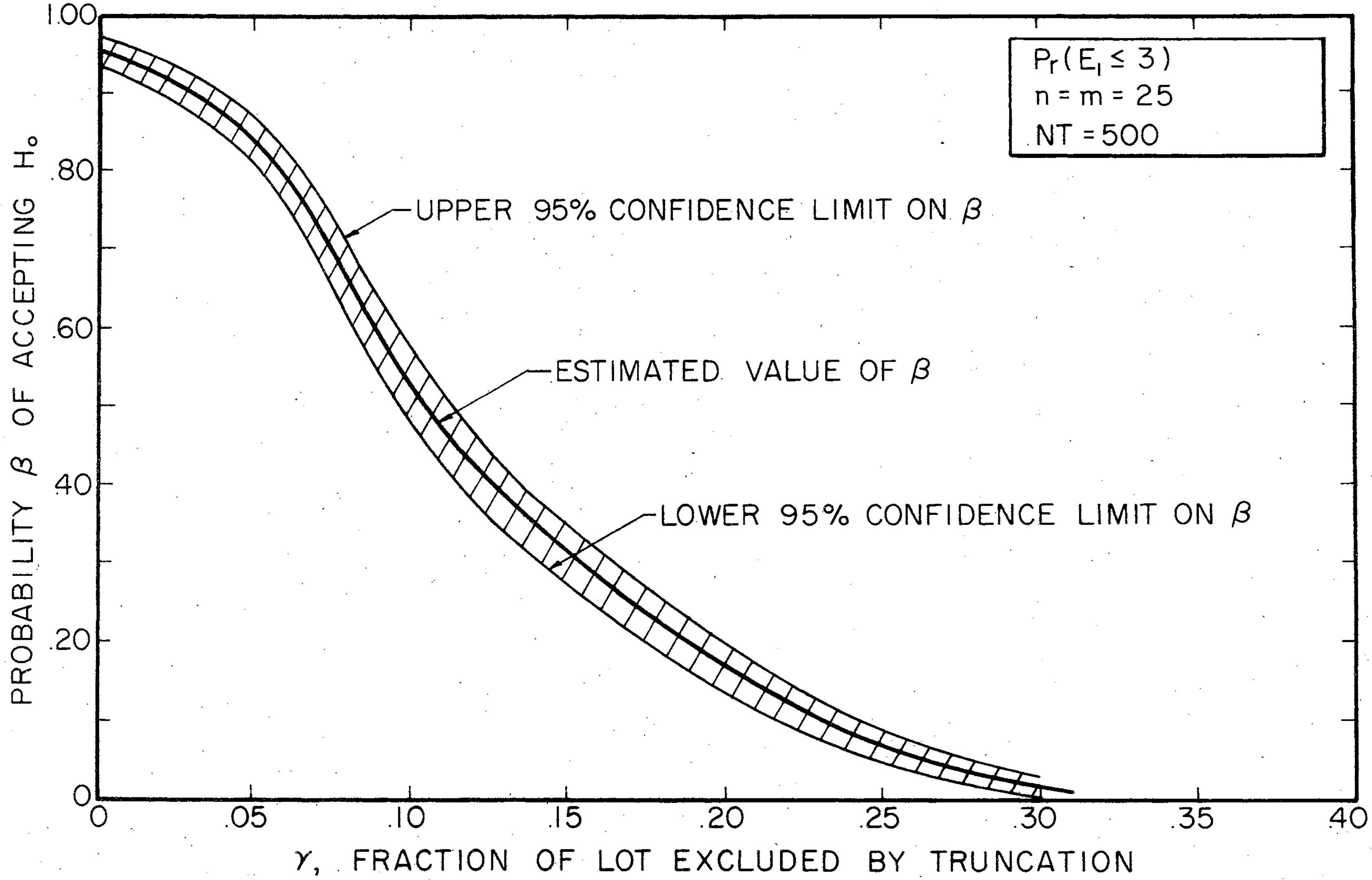

Figure 21. Confidence Band for 500 Trials, Sample Size of 25

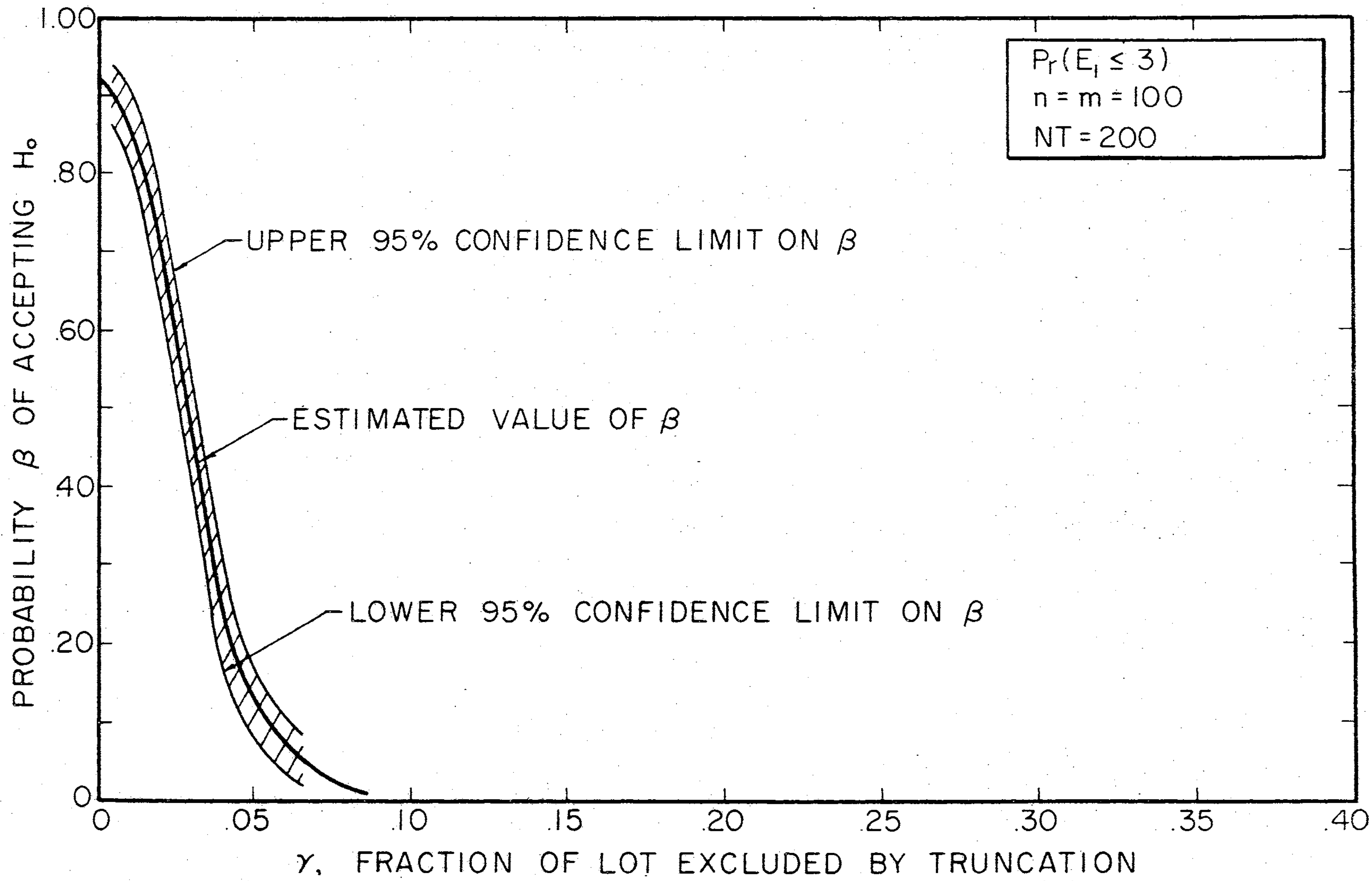

Figure 20. Confidence Band for 200 Trials, Sample Size of 100

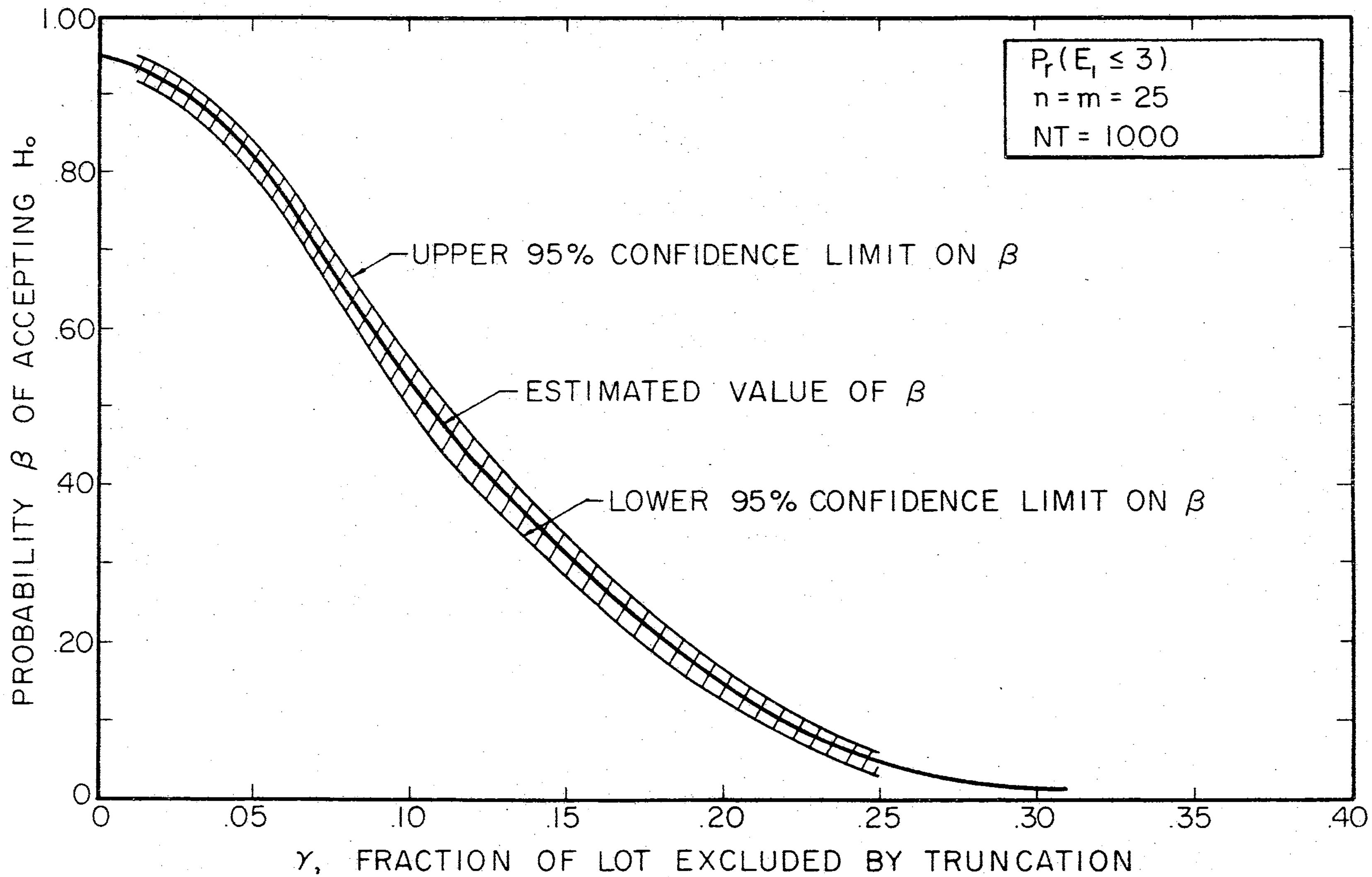

Figure 19. Confidence Band for 1000 Trials, Sample Size of 25

is no sampling variation at the point $\gamma = 0$ since α is known exactly at this point. Knowledge of this point also assists in fitting the curve through the next two or three points, and thus tends to minimize the sampling error near $\gamma = 0$. The conclusion from the interpretation of these figures is that the precision of the O.C. curves is adequate for the purposes of this study.

APPENDIX D

SIGNIFICANCE PROBABILITIES FOR THE EXCEEDANCE TEST

The significance probabilities associated with the exceedance test are presented here. Frequently used sample sizes between 2 and 100 are included, and all sample sizes considered in this dissertation are included. The tables include all critical values required in order to bracket the range of significance probabilities from 0.01 to 0.10.

In Table VII the significance probabilities for the E_1 test are tabulated. These results are based on the equations 4-15 and 4-19. In Table VIII the significance probabilities for the E_2 test are presented, based upon equations 4-20 and 4-25.

TABLE VII

SIGNIFICANCE PROBABILITIES, ONE-SIDED EXCEEDANCE TEST

$$\alpha = Pr(E_1 > k \mid F(x) = G(y))$$

n ↓ k →	1	2	3	4	5	6
2	.16667					
4	.21429	.07143	.01428	0		
6	.22727	.09091	.03030	.00758		
8		.10000	.03846	.01282	.00350	
9		.10294	.04118	.01471	.00452	
10		.10526	.04334	.01625	.00542	
12		.10870	.04658	.01863	.00686	
14		.11111	.04889	.02037	.00797	
16		.11290	.05061	.02169	.00884	
18		.11429	.05194	.02273	.00953	
20		.11538	.05301	.02356	.01001	
22		.11628	.05389	.02425	.01057	.00445
25		.11735	.05493	.02508	.01114	.00481
30		.11864	.05620	.02610	.01186	.00527
35		.11956	.05711	.02682	.01238	.00561
36		.11972	.05726	.02694	.01247	.00567
40		.12025	.05778	.02737	.01277	.00587
45		.12079	.05831	.02780	.01308	.00607
49		.12113	.05865	.02808	.01328	.00621
50		.12121	.05873	.02814	.01333	.00624
55		.12155	.05907	.02842	.01353	.00637
60		.12184	.05936	.02866	.01370	.00649
64		.12204	.05956	.02882	.01382	.00657
65		.12209	.05960	.02886	.01385	.00659
70		.12230	.05981	.02903	.01398	.00667
75		.12248	.05999	.02917	.01408	.00675
80		.12264	.06015	.02930	.01418	.00681
81		.12267	.06018	.02933	.01420	.00683
90		.12290	.06041	.02952	.01434	.00692
100		.12311	.06062	.02969	.01446	.00701

TABLE VIII

SIGNIFICANCE PROBABILITIES, TWO - SIDED EXCEEDANCE TEST

$\alpha = Pr(E_2 > k \mid F(x) = G(y))$

n ↓ k →	1	2	3	4	5	6	7	8
2	.50000	0						
4		.24286	.07143	0				
6		.27273	.12121	.04004	.00758			
8			.14102	.05944	.02028	.00505		
9			.14706	.06561	.02489	.00761		
10			.15170	.07043	.02864	.00988		
12			.15839	.07748	.03432	.01360	.00471	
14			.16296	.08237	.03841	.01643	.00638	
16			.16629	.08596	.04146	.01863	.00775	
18			.16883	.08871	.04384	.02038	.00888	
20			.17082	.09088	.04574	.02180	.00982	
22			.17243	.09264	.04729	.02297	.01062	.00467
25			.17434	.09473	.04914	.02440	.01161	.00529
30			.17662	.09725	.05139	.02616	.01284	.00609
35			.17824	.09904	.05300	.02742	.01375	.00669
36			.17851	.09933	.05327	.02764	.01391	.00679
40			.17943	.10036	.05420	.02838	.01444	.00716
45			.18035	.10139	.05513	.02913	.01499	.00753
49			.18095	.10205	.05574	.02961	.01534	.00777
50			.18109	.10220	.05587	.02972	.01543	.00783
55			.18168	.10286	.05648	.03021	.01579	.00808
60			.18218	.10342	.05699	.03062	.01609	.00829
64			.18252	.10379	.05733	.03090	.01630	.00844
65			.18260	.10388	.05741	.03097	.01635	.00847
70			.18295	.10428	.05778	.03127	.01658	.00863
75			.18326	.10462	.05810	.03152	.01677	.00876
80			.18353	.10492	.05837	.03175	.01694	.00888
81			.18358	.10498	.05842	.03179	.01697	.00890
90			.18398	.10542	.05883	.03213	.01722	.00908
100				.10582	.05920	.03243	.01745	.00924

VITA

Roger Wayne Berger

Candidate for the Degree of

Doctor of Philosophy

Thesis: STATISTICAL TESTS FOR TRUNCATION IN VALIDATING THE RESULTS OF VARIABLES SAMPLING INSPECTION

Major Field: Industrial Engineering

Biographical:

Personal Data: Born in Hastings, Nebraska, May 3, 1935, the son of Mr. and Mrs. Fred H. Berger. Married Lucille Kay Jackson June 7, 1958. Children are Angela, 8; Randall, 4; and Dustin, 0.

Education: Graduated from Hastings High School, Hastings, Nebraska, May, 1953. Received the degree Bachelor of Science in Mechanical Engineering from the University of Nebraska in June 1958. Attended University of California at Los Angeles, 1960. Received the degree Master of Science in Industrial Engineering from Kansas State University in August, 1962. Attended University of New Mexico, 1963. Completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1968.

Professional Experience: Line officer, United States Navy, 1958-1961. Industrial Engineer, Hercules Powder Co., Summer 1961. Graduate Research Assistant, Kansas State University, 1961-1962. Technical Staff Member, Sandia Corporation, 1962-1965, Summer 1966, Summer 1967. Adjunct Professor, University of New Mexico, 1963-1965. Assistant Professor, West Virginia University, 1967-1968.

Professional Activities: Associate Member, American Institute of Industrial Engineers; Member, American Statistical Association; Member, American Society for Engineering Education; Associate Member, Sigma Xi; Member, Phi Kappa Phi; Member, Alpha Pi Mu; Member, Pi Tau Sigma.