Thompson, James H.

ESTABLISHING ALPHA AND BETA RISK LEVELS IN THE EVALUATION OF SUBSTANTIVE STATISTICAL AUDITING TESTS THAT EMPLOY HYPOTHESIS TESTING

The University of Oklahoma PH.D. 1982

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ESTABLISHING α AND β RISK LEVELS IN THE EVALUATION OF SUBSTANTIVE STATISTICAL AUDITING TESTS THAT EMPLOY HYPOTHESIS TESTING

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

JAMES H. THOMPSON

Norman, Oklahoma

1982

ESTABLISHING α AND β RISK LEVELS IN THE EVALUATION OF

SUBSTANTIVE STATISTICAL AUDITING TESTS

THAT EMPLOY HYPOTHESIS TESTING

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEGEMENTS

I would like to express my sincere appreciation to each member of my dissertation committee for their cooperation, assistance, and advice throughout this research effort. Dr. Bart Ward, the chairman of my committee, deserves special recognition for the invaluable hours of thought, guidance, and effort devoted to my dissertation project. Other committee members, Dr. Rob Owens, Dr. Homer Brown, Dr. Ed Crim, and Dr. Bob Foote, all deserve thanks for their technical advice and editorial comments.

Appreciation is expressed to the School of Accounting of the University of Oklahoma, especially Dr. Ted Herrick, for the financial assistance received during my doctoral program. Thanks are given to the three accounting firms participating in this study. Without their cooperation in the pretesting and data collection stages, this study could not have been accomplished. Thanks are also given to the many friends established during my stay at the University of Oklahoma for their encouragement and advice and to Dawn Hyde for her effort in typing the Manuscript.

Finally, special thanks are given to my wife, Annet; our boys, Kevin, Kyle, and Alex; and my parents, Mr. and Mrs. Roland Thompson for their support and sacrifices during this project. Without their loving support, this undertaking would not have been accomplished.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ESTABLISHING α AND β RISK LEVELS IN THE EVALUATION OF SUBSTANTIVE STATISTICAL AUDITING TESTS THAT EMPLOY HYPOTHESIS TESTING

CHAPTER I

INTRODUCTION

The authoritative auditing literature has long recognized and permitted two general approaches to audit sampling: nonstatistical and statistical. Indeed, Statement on Auditing Standards (SAS) No. 39, paragraph .03 contains the following statements recognizing the two approaches:

> There are two general approaches to audit sampling: nonstatistical and statistical. Both approaches require that the auditor use professional judgment in planning, performing, and evaluating a sample and in relating the evidential matter produced by the sample to other evidential matter when forming a conclusion about the related account balance or class of transactions. The guidance in this Statement applies equally to nonstatistical and statistical sampling.

Paragraph .04 continues with the following statements supporting the use of either approach:

> The third standard of field work states, "Sufficient competent evidential matter is to be obtained through inspection, observation, inquiries, and confirmations to afford a reasonable basis for an opinion regarding the financial statements under examination." Either approach to audit sampling, when properly applied, can provide sufficient evidential matter.

In expounding upon the use of audit sampling, SAS No. 39, paragraph .10 defines sampling risk as the possibility that, when a compliance or a substantive test is restricted to a sample, the auditor's conclusions may be different from the conclusions he would reach if the test were applied in the same way to all items in the account balance or class of transactions. In paragraph .12, SAS No. 39 identifies two aspects of sampling risk about which the auditor must exercise professional judgement when performing substantive tests:

(1) The risk of incorrect acceptance (also known as β risk or Type II error) is the risk that the sample supports the conclusion that the recorded account balance or class of transactions is not materially misstated when it is materially mistated, and

(2) The risk of incorrect rejection (also known as α risk or Type I error) is the risk that the sample supports the conclusion that the reported account balance or class of transactions is materially misstated when it is not materially misstated.

The auditor is advised that the former risk relates to the effectiveness of the audit and that the latter risk relates to the efficiency of the audit.

SAS No. 39 suggests that statistical sampling is sometimes useful because it allows the auditor to quantify these aspects of sampling risk to levels that he can consider acceptable. Though suggesting that statistical theory is sometimes useful in the context of audit sampling, SAS No. 39 does not stipulate quantitative restrictions with respect to these risks. However, by assumption, it is desirable to have both aspects of sampling risk small. On the other hand, reduction of one necessarily leads to the increase in the other given a certain amount of evidence. Thus, the auditor must trade off the two risks (risk of incorrect acceptance, risk of incorrect rejection) in accordance with a preference function U indicating the level of satisfaction. This preference function, U, reflects the attitude of the auditor about the relative importance of each risk. If the auditor regards one of the two risks as relatively more important than the other, then his preference function would weight that attitude accordingly. For example, if the auditor experiences a situation in which an increased level of at least one of the two risks must be assumed, he would tolerate a larger increase in the less important risk than in the more important risk. On the other hand, he would not be satisfied with a result that increases the level of the more important risk by an amount larger than the increase in the level of the less important risk. This study is undertaken to theoretically and empirically investigate and compare alternative approaches which can be used to structure and resolve the tradeoff dilemma.

Role of Hypothesis Testing in Auditing

In order to achieve his overall objective in carrying out the attest function, an auditor frequently relies on statistical hypothesis testing (Elliot and Rogers, 1972 and Kaplan, 1975). In auditing, two testable hypotheses exist that permit decisions relevant for auditing use. These hypotheses are:[1]

H_o: The financial statement amount is correct,[2]

H_a: The financial statement amount is misstated by a material amount.

Statistical Hypothesis Testing

The general idea underlying statistical hypothesis testing is that the decision-maker begins with a hypothesized value for a parameter, such as the total book value of a population, and then "tests" this hypothesized value by collecting a random sample and comparing the appropriate sample statistic, such as an extended estimate of a sample mean, to the hypothesized value. If the statistic is "close" to the hypothesized value, the decision-maker accepts the hypothesized value; if the sample statistic is so different from the hypothesized value that such a result is unlikely to occur by chance when the parameter has the hypothesized value, he rejects the hypothesized value (Kazmier, 1979). Thus, the objective of hypothesis testing is to accept or reject the hypothesized value as being correct on the basis of sample data that are collected.

The first step in a statistical hypothesis test is to formulate the null and alternate hypotheses (Kazmier, 1979). The null hypothesis (designated H_o) specifies the parameter value to be tested; the alternate hypothesis (designated H_a) specifies the parameter value(s) which is accepted if the null hypothesis is rejected.

The second step is to specify the level of significance, the standard used as the basis for rejecting the null hypothesis (Kazmier, 1979). For example, if a 5 percent level of significance is specified, the null hypothesis is rejected only if the sample result is so different from the parameter value that a difference of such magnitude could only occur by chance with a probability of 5 percent or less. If a 5 percent level of significance is used as the basis for rejecting the null hypothesis, it follows that the probability of rejecting the null hypothesis when it is true is 5 percent. This risk is referred to as Type I error, which is always equal to the level of significance used as a standard for rejecting (or not rejecting) the null hypothesis, H_o. For example, in the audit context, H_o is usually formulated as: the financial statement amount is correct. The probability of a Type I error is often designated by "α" (Kazmier, 1979). In contrast to Type I error, a Type II error is the risk of accepting a false null hypothesis. The probability of a Type II error is designated by "β" (Kazmier, 1979). For example, if a 5 percent β risk level is specified, the chance that the sample result fails to be sufficiently different from the parameter value when there is material misstatement is 5 percent or less. Type II error gives an indication of the ability of the test to discriminate between the null hypothesis and its alternative. This ability of a test is referred to as its power (Kazmier, 1979). Numerically, the power of a test is taken to be one minus the probability of a Type II error. In general, assuming less risk of committing a Type I error exposes the decision-maker to more risk of committing a Type II error and vice versa given a certain amount of evidence (Kazmier, 1979).

Selecting the test statistic is the third step in statistical hypothesis testing (Kazmier, 1979). The test statistic is the value, based on the sample, used to determine whether the null hypothesis should be accepted or rejected. For example, the extended estimate of a sample mean can serve as a test statistic for the hypothesized value of the total value of a population. Generally, the test statistic is the sample estimator of the population parameter being tested. In a typical audit case, the extended estimate of a sample mean is the sample estimator of the true audit value of the population.

Next, the decision-maker must determine the critical value(s) of the test statistic (Kazmier, 1979). There may be one or two critical values depending on whether a one-tailed or a two-tailed test is appropriate. For either type of test, a critical value is in the same unit of measurement as the test statistic and identifies the threshhold value(s) of the test statistic that would lead to rejection of the null hypothesis at the given level of significance. The critical values form the boundaries of the critical region--the interval within which the test statistic will support acceptance of the null hypothesis. A sample test statistic outside the critical region will indicate rejection of the null hypothesis and thus acceptance of the alternative hypothesis.

At this point, the decision-maker determines the sample value of the test statistic by collecting a random sample from the population of interest (Kazmier, 1979). He then compares the value of the test statistic with the previously established critical value(s). If the sample value of the test statistic falls within the critical region, he can accept the null hypothesis; otherwise, he rejects the null hypothesis and accepts the alternative as though it were correct.

Audit Application

To design a suitable hypothesis test to verify the correctness of the financial statement amount (H_o), the auditor must specify several parameters: T, the amount of tolerable error; IC, the auditor's assessment of the risk that, given that errors equal to tolerable error have occurred, the system of internal accounting control would fail to detect it; AR, the auditor's assessment of the risk that analytical review and other relevant auditing procedures would fail to detect errors equal to tolerable error, given that such errors have occurred and were not detected by the system of internal accounting control; and α, the risk of incorrect rejection (American Institute of Certified Public Accountants, 1981). The hypothesis test can be converted into an equivalent test stated in terms of reliability and precision. The decision rule is that the book value will be accepted as being correct if the estimated audited value is included within the decision interval, book value $\pm$ precision, but rejected otherwise. The conversion can be expressed as follows:

$RL = 1 - \alpha$, (the complement of level of significance)

$A = T/(1 + \frac{z_\beta}{z_{\alpha/2}})$, and

$DI = BV \pm A$ where

RL represents the reliability assigned to the substantive statistical test, (referred to as TD in the appendix to SAS No. 39)

DI represents the decision interval within which the estimated audited value supports acceptance of the book value,

$z_{\alpha/2}$ is the reliability coefficient of the two-tailed α risk level,

z_β is the one-tailed reliability coefficient for the given β (risk of incorrect acceptance) risk level, BV represents the book value asserted by the financial statement under audit, and

A represents precision (distance in dollars between the parameter and the limit of the critical region).

For example, if the auditor wants to have no more than a 5 percent chance of rejecting a given balance if it is correct, and no more than a 5 percent chance of accepting the balance if it is in error by $1,000 (tolerable error) or more, he should use a reliability level for the statistical test of 95 percent (1 - .05) and a precision of $544 (A = 1,000/(1 + 1.645/1.96)). The hypothesis test approach permits the auditor to recognize and control the α and β risk levels, provides a correct decision rule for using the statistical results, and states the problem in the most relevant audit terms (Elliott and Rogers, 1972).

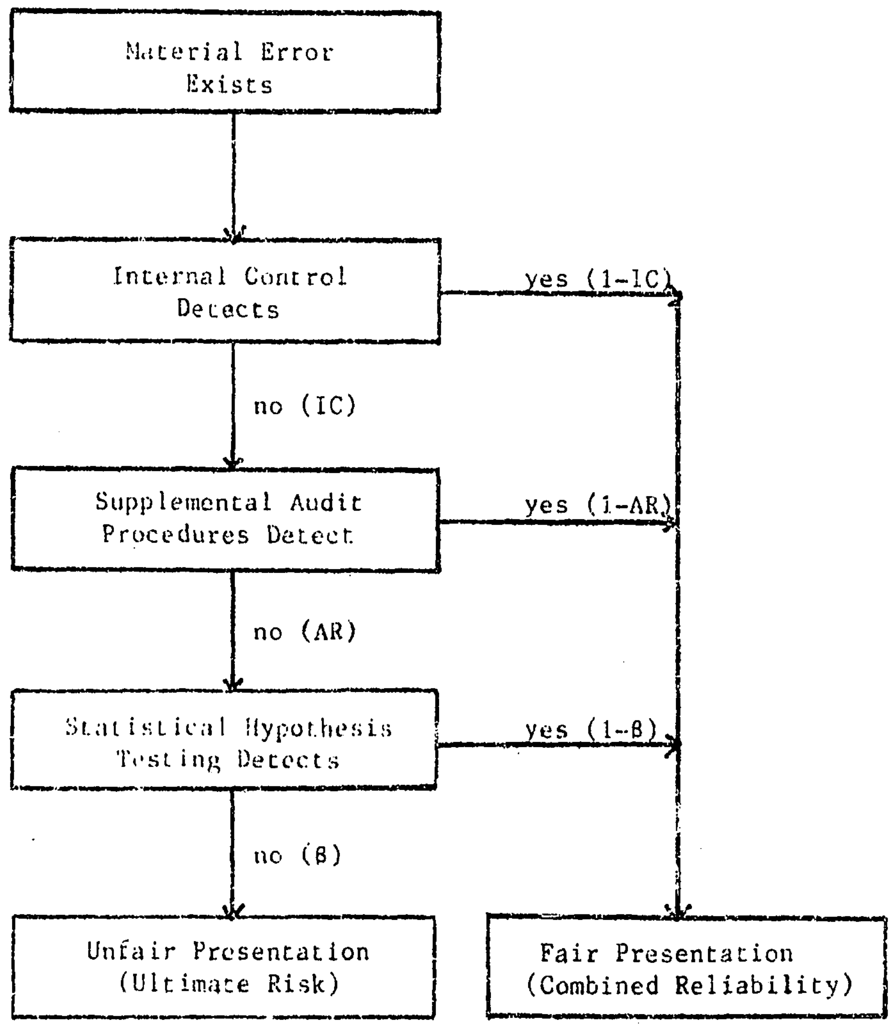
Thus, it is apparent that the level of significance of the hypothesis test is directly related to the risk of incorrect rejection (α). On the other hand, precision is a function of the risk of incorrect rejection, the risk of incorrect acceptance (β), and the amount of tolerable error. Establishing the risk of incorrect acceptance (β risk) is not an independent or isolated decision. Rather, this risk level is a function of the specified values of IC and AR. Figure I-1 illustrates this relationship. One can observe from Figure I-1 that the ultimate risk (UR) of unfair presentation is determined by β, IC, and AR. Symbolically, ultimate risk may be expressed as follows:

$$UR = IC \times AR \times \beta .$$

By restating the above equation into the following form, it becomes apparent that establishing β is not an independent or isolated decision:

FIGURE I-1

ULTIMATE RISK AND COMBINED RELIABILITY



$$\beta = UR/(IC \times AR).$$

Paragraph .06 of the appendix to Statement on Auditing Standards (SAS) No. 39 confirms this observation. When selecting the appropriate risks of incorrect rejection and acceptance (α and β risks, respectively), the auditor must consider the consequences of making an incorrect decision as in Table I-1.

The above discussion of the risks inherent in using substantive statistical tests was confined to the planning stage of a statistical sampling application. To summarize, the planning stage includes the following steps:

1. Selection of a reliability level for the substantive statistical test based on an acceptable level of α risk,
2. Judgmentally determining the risks of internal control (IC) and analytical review and other relevant auditing procedures (AR) failure and calculating an acceptable risk of incorrect acceptance (β),
3. Judgmentally determining T (tolerable error) and calculating the planned precision, A, and
4. Calculating the planned sample size using the planned precision, reliability, and an estimate of the population standard deviation.[3]

After the planning stage is completed, the auditor must execute the sampling application. The steps in the execution stage include:

1. Selecting the sample,
2. Performing the audit procedures, and
3. Obtaining the statistical estimate of the total audited value and the precision of that estimate.

TABLE I-1

CONSEQUENCES OF INCORRECT AUDIT DECISIONS

Actions	States: Amount is Correctly Stated	States: Amount of Misstatement Exceeds Tolerable Error
Accept the book value	No error	Type II error[b] (β risk)
Reject the book value	Type I error[a] (α risk)	No error

[a]Consequences of this incorrect decision may include the cost of unnecessarily extending audit procedures, needlessly requiring adjustment to the balance, or needlessly rendering a qualified or adverse opinion.

[b]Consequences of this incorrect decision may include the cost of potential lawsuits (both direct and indirect) and loss of professional reputation.

Once audit procedures are performed on each item in the sample, the results must be evaluated--both qualitatively and quantitatively. The qualitative evaluation includes an analysis of the sample evidence for information about the nature and causes of any differences noted in the sample to determine their audit impact (Ernst & Whinney, 1980). This part of the evaluation stage is very important since it may cause the validity of a previous judgment--for example, the assessed risk of internal control failure--to be questioned. Or, this phase may result in a postponement or abandonment of the statistical evaluation. For example, the qualitative evaluation may reveal evidence of an intentional irregularity. Quantitative evaluation may then be inappropriate until an additional investigation is completed.

Finally, a quantitative evaluation of the statistical results is used to determined whether the sample evidence supports acceptance of the book value. If the sample evidence does not support acceptance of the book value, the auditor must decide whether the sample evidence is conclusive[4] enough to propose an adjustment (Ernst & Whinney, 1980). If the sample evidence is deemed conclusive enough to propose an adjustment, the auditor must determine the amount of the necessary adjustment; if not, he should not propose an adjustment based solely on the sample. (Ernst & Whinney, 1980). If the sample evidence does support the acceptance of the book value, the auditor can accept the book value as correctly stated (Ernst & Whinney, 1980). The quantitative phase of the evaluation is the focus of this study.

Need for the Study

When statistical sampling is used as the basis for determining whether to accept or reject an account balance, one of three situations will exist depending on the difference between the book value and the estimated audited value as depicted in Figure I-2. Based on an examination of Figure I-2, it is apparent that when the estimated audited value, X_1, differs radically from the book value (BV), the statistical evidence fails to support acceptance of the book value regardless of whether decision interval one or two is used. Similarly, when the estimated audited value, X_3, closely approximates the book value, the statistical evidence supports acceptance of the book value regardless of whether decision interval one or two is used. In either of these situations, the decision to accept or reject the book value is not very sensitive to changes in the width (amount of precision) of the decision interval. Thus, in these situations, construction of the decision interval endpoints is not a very critical step.

On the other hand, when the estimated audited value, X_2, differs by more than a small amount but by less than a material (M) amount, the decision to accept or reject the book value turns on what decision interval is used (in this example, interval one or two). In this situation, the decision to accept or reject the book value is sensitive to changes in the width (amount of precision) of the decision interval. Thus, in closecall decisions, determining the decision interval endpoints is a critical step. It is this situation that this study will examine: in closecall decisions, what approach do auditors take in the construction of decision intervals when evaluating the sam-

FIGURE I-2

SENSITIVITY OF THE DECISION TO ACCEPT OR REJECT THE BOOK VALUE

TO CHANGES IN THE WIDTH OF THE DECISION INTERVAL

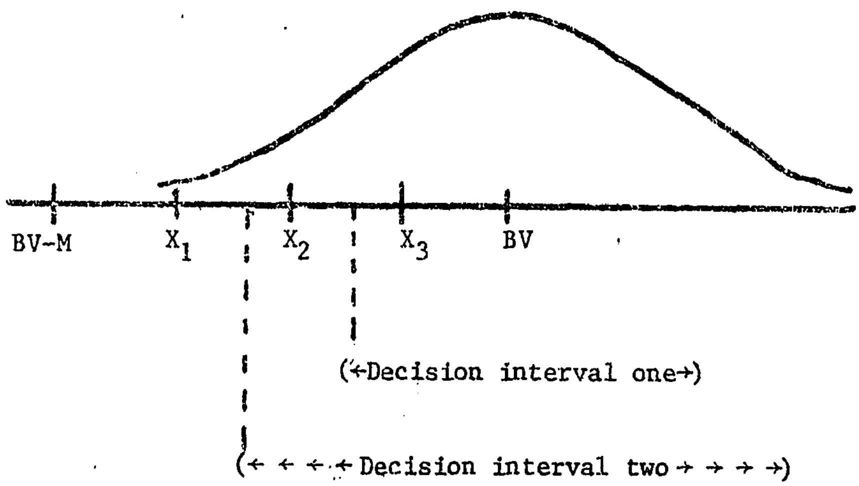


TABLE I-2

COMPARISON OF APPROACHES 1, 2, AND 3 AT PLANNING STAGE WITH RESPECT TO INITIAL CONSIDERATION IN PLANNING α AND β RISK LEVELS

	Approaches 1 and 2	Approach 3
Analysis:	Balanced consideration of α and β	Unbalanced consideration of α and β
Explanation:	α and β risk levels considered in light of relevant audit factors	β risk level considered in light of relevant audit factors
Risk Levels:	α risk influenced by: -frequency of expected errors (ER) -relative difficulty of extending the test (EXT) -tolerance of management (MGT)	(α risk not a consideration)
	β risk influenced by: -the risk that internal control fails to detect tolerable error (IC) -the risk that analytical review fails to detect tolerable errors (AR) -the ultimate risk of unfair presentation (UR)	β risk influenced by: -the risk that internal control fails to detect tolerable error (IC) -the risk that analytical review fails to detect tolerable error (AR) -the ultimate risk of unfair presentation (UR)

ple results of substantive tests? In order to answer this question, this study considers statistical sampling risks, α and β, inasmuch as control of these risks determines the amount of precision and thus the width of the decision interval.

A consideration of the alternative approaches available to the auditor for establishing the risks of incorrect rejection (α) and incorrect acceptance (β) for use in evaluating substantive statistical sampling tests reveals a considerable amount of diversity. The diversity exists as a result of different attitudes about the preliminary specification of those two risks. On one extreme, one might require that the risk of incorrect rejection (α) be maintained at the planned level. Because of the inverse relationship between the two risk levels, maintaining α at the planned level statistically requires that the risk of incorrect acceptance (β) be allowed to vary as necessary. Similarly, but on the other extreme, one might require that β be maintained at the planned level; thus, α must be allowed to vary as necessary. Between those two extremes, both α and β risk levels might be allowed to vary to some degree. Tables I-2, I-3a, and I-3b present three approaches that are considered in this study:

Approach 1 (classical statistical approach):

β is maintained at the planned level and α is free to vary as needed,[5]

Approach 2 (tradeoff approach):

Both α and β are varied in a manner that maintains the planned ratio of α to β,[6]

Approach 3 (dollar-unit sampling approach):

TABLE I-3a

COMPARISON OF APPROACHES 1, 2, AND 3 WITH RESPECT TO RESTATEMENT OF α AND/OR β RISK LEVELS AT THE EVALUATION STAGE AS A RESULT OF AN ACHIEVED STANDARD DEVIATION DIFFERENT FROM THE EXPECTED STANDARD DEVIATION WHEN THE ESTIMATED AUDITED VALUE, X, IS CONTAINED WITHIN THE DECISION INTERVAL IMPLIED BY α'' AND β''

	Approach 1	Approach 2	Approach 3
Analysis:	Unbalanced consideration in favor of β	Balanced consideration of α and β	Unbalanced consideration in favor of β
Explanation:	α risk is adjusted so as to restate β risk to the planned level	α and β risk levels increased (decreased) so as to maintain planned ratio of α to β $\left(\frac{\alpha''}{\beta''} = \frac{\alpha_p}{\beta_p}\right)$	β risk restated so as to maintain planned level; α risk not considered
Risk Levels:	$\alpha'' = \alpha_p$	$\alpha'' > \alpha_p$ and $\beta'' > \beta_p$	$\beta'' = \beta_p$
		or	
	$\beta'' = \beta_p$	$\alpha'' < \alpha_p$ and $\beta'' < \beta_p$	α'' not computed

α_p represents the planned level of α risk

α'' represents α risk after restatement

β_p represents the planned level of β risk

β'' represents β risk after restatement

TABLE I-3b

SIMILARITIES OF APPROACHES 1, 2, AND 3 WITH RESPECT TO RESTATEMENT OF

α AND/OR β RISK LEVELS AT THE EVALUATION STAGE AS A RESULT OF AN

ACHIEVED STANDARD DEVIATION DIFFERENT FROM THE EXPECTED

STANDARD DEVIATION WHEN THE ESTIMATED AUDITED VALUE, X,

IS NOT CONTAINED WITHIN THE DECISION INTERVAL

IMPLIED BY α'' AND β''

	Approaches 1 and 2	Approach 3
Analysis:	Balanced Consideration of α and β	Unbalanced Consideration in favor of β
Explanation:	α risk level set at the level required for the statistical test	α risk not considered
	β risk set equal to ultimate risk; no reliance on internal control nor analytic review as they were assessed as failing to detect misstatements that exceed tolerable error	β risk set equal to ultimate risk; no reliance on internal control nor analytic review as they were assessed as failing to detect misstatements that exceed tolerable error
Risk Levels:	$\alpha''' = .05$ or less	α''' not computed
	$\beta''' = f(UR) = .05$ or less	$\beta''' = f(UR) = .05$ or less

α''' represents α risk meeting conclusiveness criteria
β''' represents β risk meeting conclusiveness criteria
UR represents ultimated risk

β is maintained at the planned level and α is not considered.[7] These tables analyze the balanced or unbalanced treatment of α and β risk, explain the reasoning behind the adjustments, and describe the statistical adjustment required for both α and β risk at the planning (Table I-2) and evaluation (Tables I-3a and I-3b) stages.

At the planning stage, α and β risk levels are initially established in light of relevant audit factors as described in Table I-2. These factors are related to the auditor's expectations about the frequency of substantive errors in the account balance or transaction group being sampled. For example, as the frequency of error expectation increases, α risk would normally be established at a lower level by the auditor and vice versa.

At the evaluation stage, α and β risk levels must be reconsidered when the achieved (sample) standard deviation differs from that initially estimated. Once these two risk levels are reconsidered, the auditor must determine whether the resulting decision interval implied by the restated α and β risk levels contains the estimated audited value. When the estimated audited value falls within this decision interval, the auditor is able to accept the estimated audited value having at least implicitly controlled both α and β risk levels. This circumstance is described in Table I-3a.

On the other hand, when the estimated audited value is not contained within the decision interval implied by the restated α and β risk levels, the auditor must determine if the sample evidence is sufficiently conclusive to propose an audit adjustment. Conclusiveness criteria must provide for carefully controlled α and β risk levels--α must be controlled to limit the risk of rejecting a fairly stated book value (except in Approach 3), and β must be controlled to limit the risk of accepting an adjusted book value that is materially misstated. This latter circumstance is described in Table I-3b.

Analysis of Tables I-2, I-3a, and I-3b reveals the following about the three approaches:

1. Approach 1 begins with a balanced consideration of α and β risk levels at the planning stage when α_p and β_p are established. However, in the evaluation stage, a balanced consideration of the two risk levels is initially abandoned in order to set the adjusted β risk level, β'', equal to the planned level, β_p. Finally, in the evaluation stage when the estimated audited value is not contained within the decision interval implied by α'' and β'', approach 1 returns to a balanced consideration of the two risk levels when establishing conclusive alpha, α''', and beta, β''', risk levels.
2. Approach 2 begins with and consistently employs a balanced consideration of α and β risk levels throughout both the planning and evaluation stages.
3. Approach 3 begins with and consistently employs an unbalanced consideration of α and β risk levels in favor of β risk throughout both the planning and evaluation stages.

The importance of a balanced or unbalanced consideration of α and β risk levels depends on whether and to what extent the sample size or the statistical results and subsequent audit conclusions are affected by this characteristic.

In spite of the relative diversity of approaches available to the auditor, a search of the auditing literature reveals no relevant research to investigate the relative benefits and shortcomings of these alternatives. If investigation of the preference function of auditors in establishing the aforementioned risk levels reveals the existence of and preference for an approach different from that which universities and auditing firms are teaching, then a reallocation of resources may be needed. Inasmuch as efficient resource allocation is important in any business endeavor, an investigation of auditors' preferences for establishing α and β risk levels in evaluating substantive sampling tests is appropriate.

Objectives of the Study

The principal objectives of this study are:

1. To determine whether auditors who are involved in the quantitative evaluation stage of sample results from substantive statistical tests use approach 1 (classical) to establishing the risk of incorrect rejection (α) and the risk of incorrect acceptance (β),
2. To determine the extent to which those auditors agree on an approach to establishing those risk levels,
3. To determine whether exposure to an alternative approach to establishing those risk levels affects the approach selected by those auditors,
4. To determine whether attitudes about overall satisfaction with the sample results of substantive statistical tests is affected by the

approach employed by those auditors when selecting the risks of incorrect rejection and acceptance for evaluating those tests.

Scope, Design, and Research Method of the Study

The scope of this research was limited by the selection of subjects, to data collected from distributing the research instrument to participating audit executives, and to a consideration of only two approaches for establishing α and β risk levels. Because the level of specialization required to meaningfully respond to the research instrument was so great, only audit executives who were thought to be qualified for participation were considered. Access to these subjects was dependent on cooperation with large, international accounting firms. Only the larger firms were asked to participate inasmuch as these firms represent the major users of statistical sampling tests (Bedingfield, 1975). Due to a potentially small number of qualified respondents and to limited access to those respondents, the number of actual respondents in this study was limited.

The research instrument, contained in Appendix A, was the device for gathering responses to be analyzed in this study. To enhance its content validity, the research instrument was pretested first with accounting graduate students and then with audit executives (see chapter III).

This study considers only two approaches to establishing α and β risk levels--an approach (later designated the "classical approach") hypothesized to be predominant in practice and an alternative approach (later designated the "tradeoff approach") hypothesized to be conceptually superior. These approaches are studied only within the hypothesis testing mode. Thus, whether the conclusions would differ if the approaches were studied within some other mode, such as calculation confidence intervals rather than performing hypothesis testing, is beyond the scope of this study.

The research method employed in this study consisted of eight steps. These steps were (1) formulating the research questions to be studied, (2) outlining the data analysis techniques to be used to interpret the results of the survey, (3) choosing the research design for the questions to be studied, (4) designing and pretesting the research instrument, (5) selecting the participants for the study, (6) conducting the survey, (7) analyzing the results, and (8) formulating conclusions based on the analysis of results.

Organization of the Study

This research study consists of four chapters in addition to Chapter I which introduces the study by providing the statement of the problem; discussing the role of hypothesis testing in auditing: and outlining the need for, objectives of, scope and research design and method of, and organization of this study. In Chapter II, the existence of an alternative approach to the auditor's selection of the risks of incorrect rejection and acceptance within the hypothesis testing mode for evaluating the sample results of substantive statistical tests is explored. This chapter also describes the approach that is hypothesized to be predominant in practice and provides a comparison of the two approaches. Chapter III provides a comprehensive discussion of the research design and method to be employed in the study. Chapter IV presents a detailed analysis and interpretation of each of the research questions investigated in the study. Finally, Chapter V offers conclusions and implications of the study.

Appendix A presents the case situations and questions that were distributed to the subjects and a summary of each subject's responses. Appendix B presents the debriefing questionnaire that was used to pretest the research instrument.

CHAPTER II

COMPARATIVE THEORETICAL DEVELOPMENT AND PHILOSOPHY OF ALTERNATIVE STATISTICAL SAMPLING STRATEGIES USING HYPOTHESIS TESTING

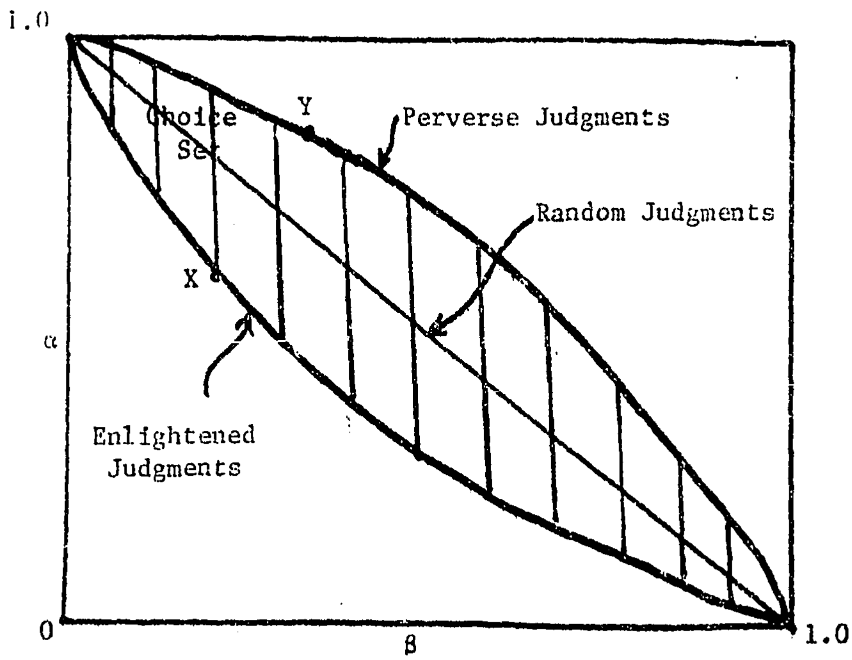
A number of alternative approaches to evaluating the sample results of substantive statistical tests are possible. Three possible approaches are a classical approach, a decision theory approach, and a tradeoff approach.

The theory of statistical hypothesis testing is generally concerned with defining procedures such that α and β[8] are well-defined. In auditing, a typical choice set for α and β is presented in Figure II-1.

The line running from the point ($\alpha = 1$, $\beta = 0$) to $\alpha = 0$, $\beta = 1$) represents the set of all feasible randomized pairs of risk descriptions when no evidence is gathered. For this case, statistical hypothesis testing would not be used inasmuch as flipping a coin would be a more efficient and equally effective strategy. The point ($\alpha = 0$, $\beta = 0$) represents perfect evidence. Clearly, this result is not possible in an audit situation and is thus excluded from the choice set. Similarly, the perfect error point ($\alpha = 1$, $\beta = 1$) is not available since if we could be sure of making an error, by doing the exact opposite, we could be sure of not making an error. The curve labelled "enlightened judgments" represents the best possible audit action based

FIGURE II-1

PROBABILITIES OF ERROR



H_o: The financial statement amount is correct,

H_a: The financial statement amount is misstated by a material amount,[9]

α : P(rejecting $H_o | H_o$ is true),

β : P(accepting $H_o | H_a$ is true).

Source: Leamer, 1978

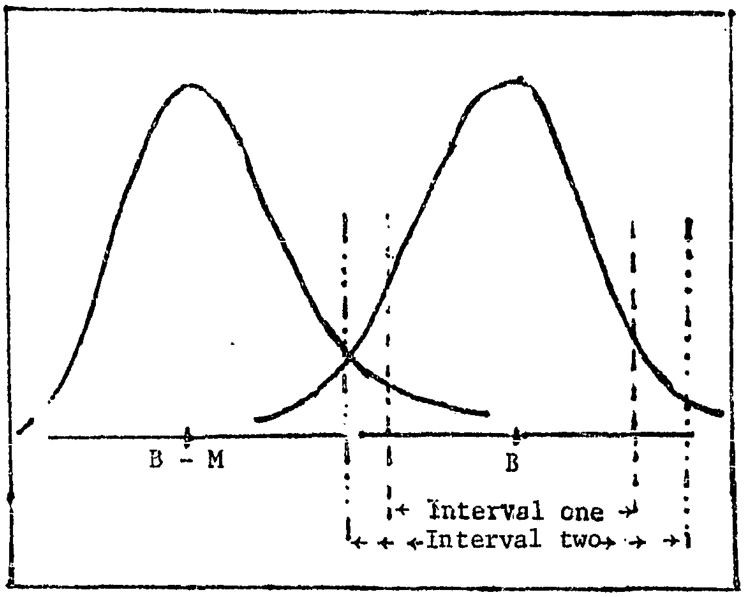
on the available evidence. The curve is composed of an infinite number of points each representing a statistically acceptable pair of risk descriptions, (α_i, β_i), based on the available evidence (Leamer, 1978). In the context of the auditor's problem, each of these pairs of risk decriptions relate to one of the infinite number of statistically acceptable decision intervals that the auditor may select when evaluating the sample results of substantive statistical tests. Figure II-2 illustrates two such decision intervals: interval one, in which α and β are approximately .10 and .025, respectively; and interval two, in which α and β are approximately .05 and .05, respectively.

The curve labelled "perverse judgments" in Figure II-1 is just the mirror image of the "enlightened judgments" curve. For example, Y (approximately α_Y = .8, β_Y = .4) is the mirror image of X (approximately α_X = .6, β_X = .2). Since X reflects lower risk levels for both α and β, X represents an "enlightened judgment" and Y represents a "perverse judgment" of available evidence.

The essential problem faced by an auditor in this situation is to select those levels of α and β resulting in the greatest level of satisfaction given available evidence. Theoretically, the optimum levels of α and β are determined by the point of intersection of the "enlightened judgments" curve (E_1) and the highest possible utility curve (U_2) as depicted in Figure II-3. Practically, the auditor must exercise professional judgment when considering the α, β tradeoff so as to select those levels of α and β risk that maximize satisfaction. As previously mentioned and as can be inferred from the above analysis, there is no obvious solution because, by assumption, it is desirable to

FIGURE II-2

TWO DECISION INTERVALS FOR A GIVEN TEST SETTING



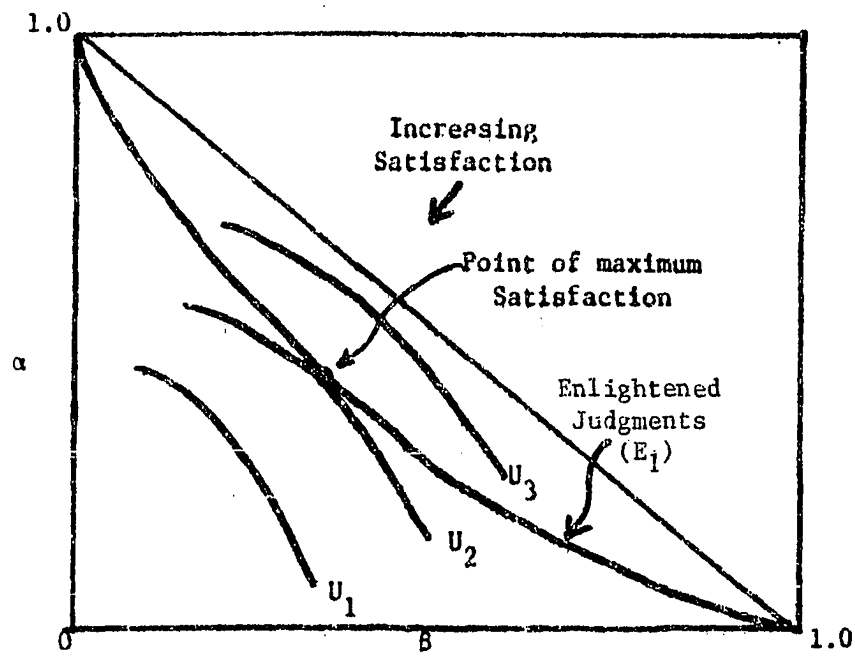
B: the book value

B - M: the book value less the amount of a material error

Source: Ernst & Whinney, 1980

FIGURE II-3

INDIFFERENCE CURVES OF A PREFERENCE FUNCTION



Note: Any other α, β pair along E_1 would result in lower satisfaction because that point would intersect a utility curve to the right of U_2 (i.e., utility decreases to the right).

Source: Leamer, 1978

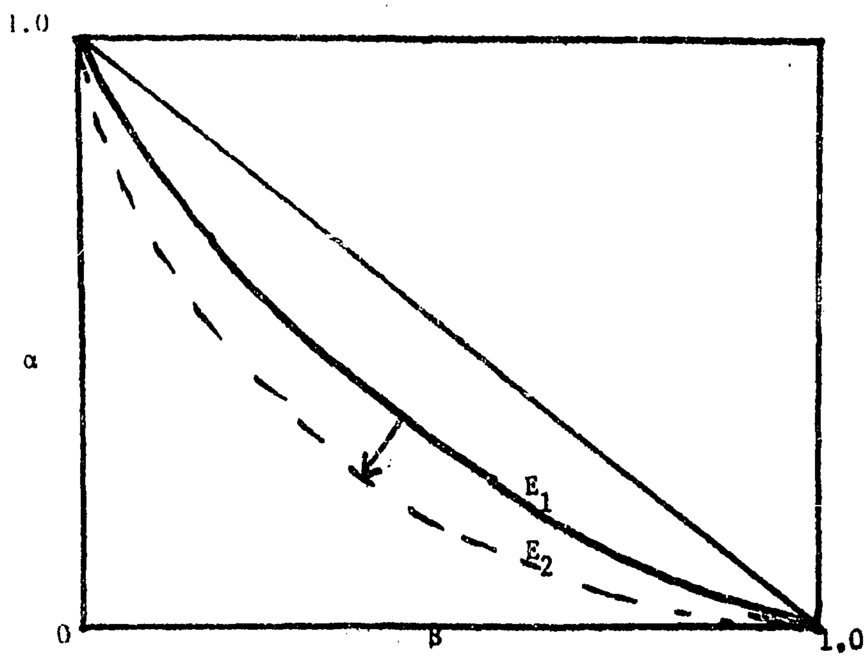
have both α and β small; yet, the decrease in one risk level necessarily increases the other given a certain amount of evidence. Furthermore, the solution is dynamic; that is, the auditor initially selects the levels of α and β risk on the basis of an expected population standard deviation. Once the sample is drawn and differences are noted, he must evaluate his earlier assessment on the basis of an observed standard deviation different from the estimated amount.

One way of alleviating the choice dilemma in some audit situations is to gather additional evidence (when achieved results would without additional sampling place the auditor on a lower utility curve because the sample is less powerful than the expected population standard deviation suggested). If that alternative is not feasible, the auditor must, as a result of observing a standard deviation greater than the planned amount, consider other alternatives such as adjusting the levels of α and/or β risk or abandoning the statistical procedure. Even in those situations in which the decision to gather additional evidence is initially a feasible alternative, the situation ultimately reduces to one in which the auditor must establish a tradeoff between the two risk levels. This tradeoff is inevitable because, as pointed out earlier, perfect evidence is never available.

Possession of additional evidence would permit a decrease in both α and β. As the amount of evidence increases, the "enlightened judgments" curve bows more in the direction of the point of perfect evidence as depicted by the dotted line (E_2) in Figure II-4. In terms of utility, the level of satisfaction increases as the amount of evi-

FIGURE II-4

EFFECT OF INCREASED INFORMATION ON α AND β RISK LEVELS



Adapted: Leamer, 1978

dence increases. This increase in satisfation is depicted in Figure II-5.

On the other hand, if the achieved standard deviation is less than the planned amount, the auditor's earlier assessment is more adequately satisfied. Thus, α and/or β risk levels could be reduced without the need to gather additional evidence.

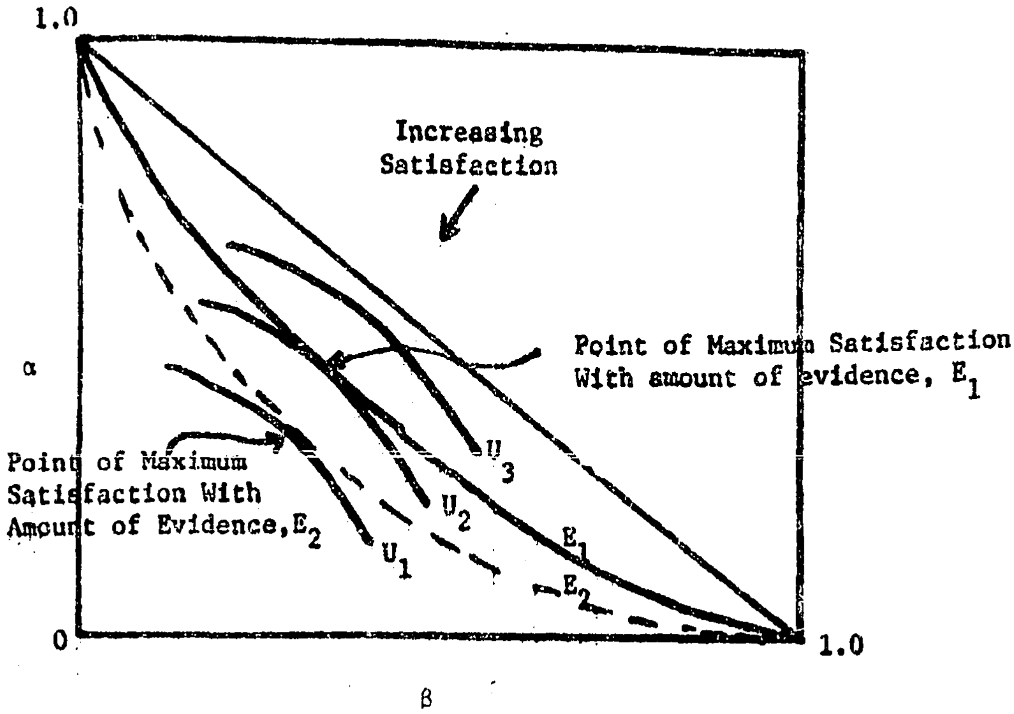
Classical Statistical Approach

Kern, Neyhart, and Hock report that auditors traditionally have adopted a reliability level of 95 percent (5 percent risk level) for planning substantive tests irrespective of factors indicating that perhaps a lower level of reliability is suitable for a particular segment of the audit. This practice occurs because there are no objective measurement criteria with which to judge the propriety of the auditor's selected level of reliability. Without effective, well-defined criteria, an auditor must judgmentally select a level of sample reliability based on his own knowledge of statistics, his firm's policies, reliability levels used in examples in the auditing literature, or reliability levels used in other disciplines. Because the auditing literature is replete with examples in which 95 percent reliability levels are used and because 95 percent reliability levels are frequently used in other disciplines, the tendency for the auditor to use 95 percent[10] reliability levels is encouraged (Kern, Neyhart, and Hock, 1973).

In the evaluation stage, Kern, Neyhart, and Hock indicate that the auditor can "control" both risks by altering the level of reli-

FIGURE II-5

EFFECT OF INCREASED INFORMATION ON SATISFACTION WITH EVIDENCE



Adapted; Leamer, 1978

ability (and consequently α risk). Thus, their approach reflects a decision to set β risk at the planned level.

Elliott and Rogers admit a theoretical optimum value for α that will minimize total cost. This value, they feel, is probably very low, say 0.1 or less. Further, they reason that since sample sizes are substantially larger when α is less than 0.05, an α risk level in the range 0.05 to 0.1 appears reasonable. Their contention is that solving for the optimal value of α is not practical. Rather, the most practical approach is for the auditor to select an acceptable α risk level and then use that pre-specified level for all statistical tests as a matter of policy (Elliott and Rogers, 1972).

Elliott and Rogers state that minimization of overall β risk is the reason for existence of the public accounting profession. Therefore, an auditor desires a great deal of assurance that he has not accepted materially misstated amounts. Thus, a low β risk level, say 0.05, is appropriate when the auditor has nothing but statistical evidence on which to rely. They contend that a reasonable upper limit for β is 0.5 since once the auditor decides to conduct a statistical test, he would not want to bother with any test having less than an even chance of discovering a material error should one exist. Thus, they conclude that a β risk ranging from 0.05 to 0.5 is generally appropriate. Within this range, Elliott and Rogers propose a point system to determine the precise β level. This point system is based on reliance to be placed on internal control and on other audit procedures (Elliott and Rogers, 1972).

Once the sample is drawn and audited, Elliott and Rogers point out that because the achieved standard deviation almost always will be different from the planned standard deviation, one or more of the parameters α, β, and M, the predetermined amount of a material error, must be adjusted. They suggest that holding β and M at the planned level and allowing α to vary over some reasonable range is the most logical approach since α is not as critical to the audit as β and M are. If the achieved standard deviation is found to be larger than planned, the test can still be considered complete as long as α remains at an acceptable level--say not greater than .10. If, as a result of holding β and M at the planned level, α increases to an unacceptable level, then additional sampling would be required to bring the three parameters to within the required limits (Elliott and Rogers, 1972).

Roberts suggests an approach[11] to statistical decision- making that also hypothesizes that the recorded amount is correct. The decision rule under this approach describes the circumstances under which statistical evidence is interpreted as supporting the correctness of the recorded amount:

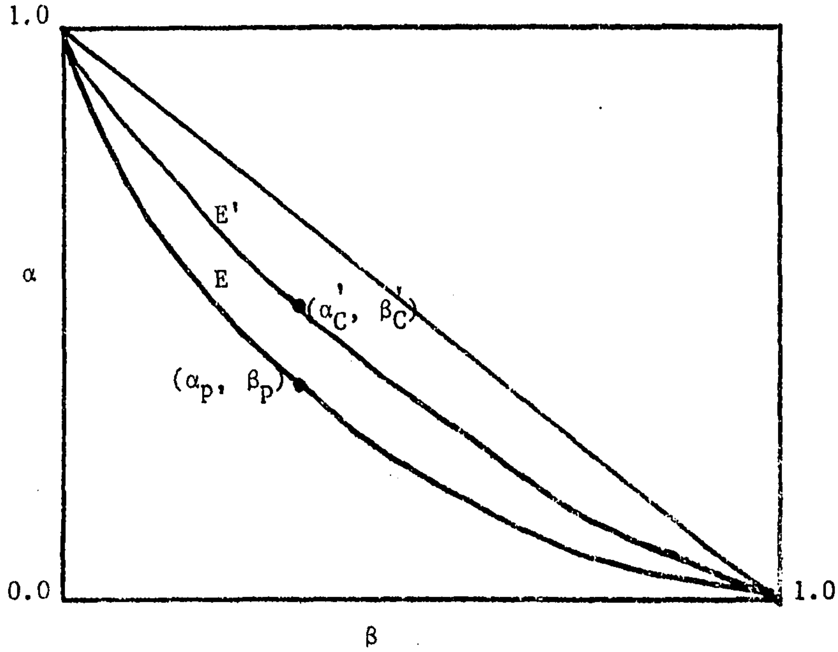
1. Compute the value of the estimated audited amount, together with the achieved precision at a specified interval reliability, and
2. Decide that the statistical evidence supports the correctness of the recorded amount if it is within the calculated precision interval; otherwise, decide that the statistical evidence fails to support the correctness of the recorded amount (Roberts, 1978).

Under this approach, risks of deciding incorrectly are controlled by selecting the values of precision and reliability. Roberts states that the probability of deciding that the statistical evidence supports the correctness of the recorded amount when in fact the recorded amount is materially misstated is the more important error (β risk). This risk is controlled by selecting precision in relation to the amount of a material misstatement. When the achieved standard deviation differs from the estimated standard deviation, the auditor can adjust precision to maintain an effective β risk at the planned level. However, the α risk level is allowed to vary as needed to fulfill the β risk requirement.

The positions taken by Kern, Neyhart, and Hock, by Elliott and Rogers, and by Roberts have a common characteristic in that β risk is not allowed to vary as a result of a decrease (increase) in the amount of evidence due to a higher (lower) standard deviation observed in the sample as compared to the estimated amount; rather α risk is allowed to vary as needed. A graphic representation of this adjustment when the amount of evidence decreases (E') appears in Figure II-6. In Figure II-6, the planned levels of α and β risk, (α_p, β_p), and the adjusted levels of α and β risk, (α_p', β_p') are identified. Note that $\alpha_p < \alpha_C'$ but $\beta_p = \beta_C'$. In general, as the amount of information decreases (increases), the point of maximum satisfaction under the classical approach travels up (down) the line parallel to the vertical axis and passing through the point identified by (α_p, β_p) as illustrated by Figure II-7.

FIGURE II-6

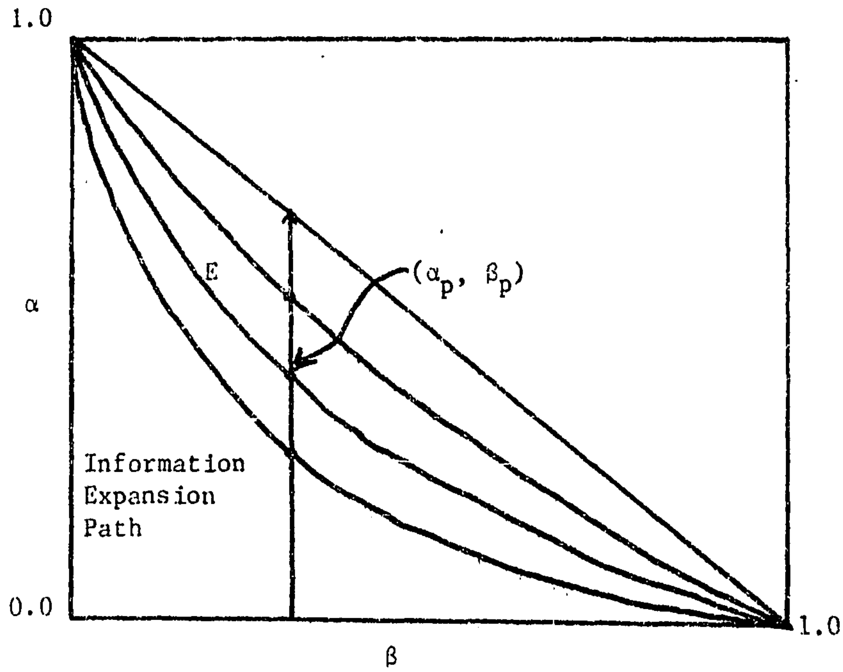
RESPONSE TO DECREASE IN EVIDENCE: CLASSICAL APPROACH



Adapted: Leamer, 1978

FIGURE II-7

INFORMATION EXPANSION PATH: CLASSICAL APPROACH



Adapted: Leamer, 1978

Tradeoff Approach

Figure II-8 presents a graphic solution of the tradeoff approach as compared to the classical approach when the amount of evidence decreases (E'). Note that the adjustment under the tradeoff approach is to (α_T', β_T') where $\alpha_p < \alpha_T' < \alpha_C'$ and $\beta_p = \beta_C' < \beta_T'$. In general, as the amount of information decreases (increases), the point of maximum satisfaction under the tradeoff approach travels away from (toward) the origin tracing out an information expansion path as indicated in Figure II-9.

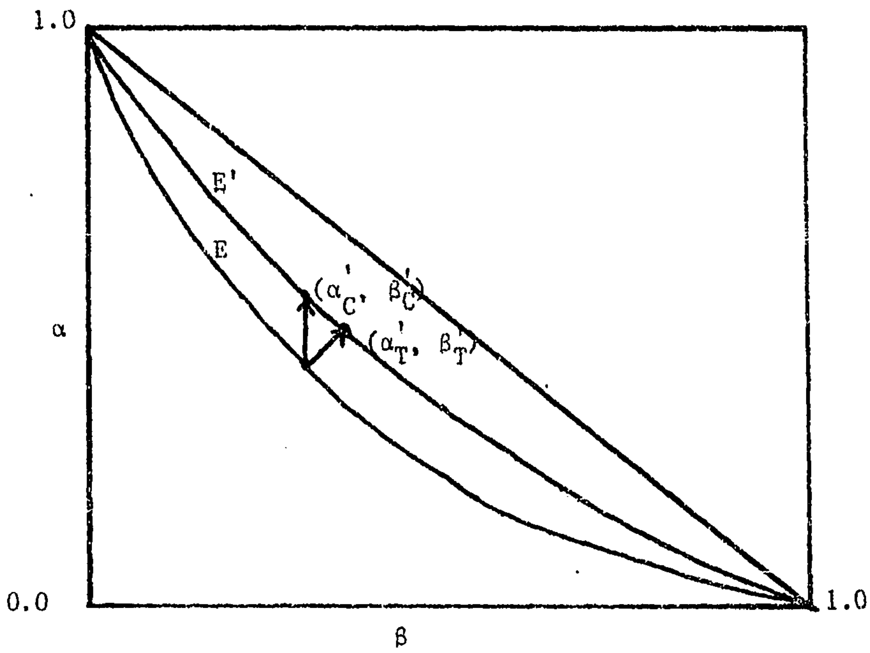
Decision Theory Approach

A decision theory approach to the sampling problem has been proposed by Kinney. According to Kinney, the traditional approach bases α and β on an internal control system design (IC) and on the probability of management override of the internal control system (MO) (Kinney, 1975). The function relating IC and MO to α and β is presumably based on expected cost considerations over all accounts and clients. The traditional approach results in a sample size based on pre-specified levels of α and β, an estimated standard deviation, and the amount of a material error. In contrast to the traditional approach, the decision theory approach does not require pre-specification of α and β. Rather, the optimum sample size is derived from (1) explicit consideration of the probability that financial statement amounts are correctly presented, (2) anticipated costs associated with Type I and Type II errors as well as costs of sampling, (3) an estimate of the population standard deviation, and (4) the amount of a material error (Kinney, 1975).

FIGURE II-8

COMPARISON OF ADJUSTMENT FOR CLASSICAL AND TRADEOFF APPROACHES TO

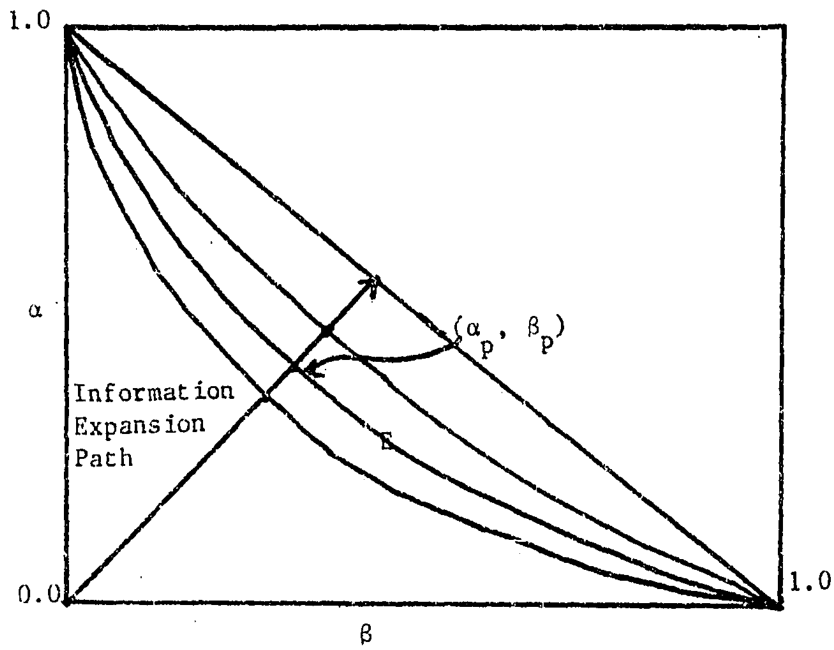
DECREASE IN EVIDENCE



Adapted: Leamer, 1978

FIGURE II-9

INFORMATION EXPANSION PATH: TRADEOFF APPROACH



Source: Leamer, 1978

The optimal sample size calculated using a decision theory approach implies optimal α and β which are merely products of the sample size and not determinants of it. However, this approach, while avoiding the need to pre-specify α and β, creates two new problems--making probability estimates regarding the likelihood that financial statement amounts are correct and making cost estimates associated with making Type I and Type II errors. Regarding the former problem, Corless' study raises considerable doubt about the quality of such probability estimates (Corless, 1972). Felix found that subject-auditors in his study appeared to agree quite closely on their assessment of prior probabilities; however, he refrained from drawing inferences from this result because it could be the result of unintentional bias in the content of his case study (Felix, 1974). Chesley, in his study of elicitation of subjective probabilities, found differences to exist among the responses of his student-subjects, but he hypothesized that training and explanation would remove the differences he found (Chesley, 1976). Thus, the ability of auditors to make the subjective probability estimates needed to apply this approach is an unsettled issue.

The latter of the two problems introduced by use of a decision theory approach is unique to this approach not because costs associated with Type I and Type II errors are considered but rather because of the need to precisely quantify these costs. The costs associated with making Type I or Type II errors are indirectly considered by other approaches in the sense that β risk is generally regarded as the more important error (Elliott, 1973). Thus, the relative degree to which auditors are willing to allow β risk to vary is not as great as the degree to which α risk is allowed to vary.

Consistent with the objective of minimizing expected loss, Leamer identifies an approach similar to the decision theory approach in which decreases (increases) in the amount of evidence affect the levels of α and β risk as depicted in Figure II-10 when the loss associated with a Type I error is three times as great as the loss associated with a Type II error. In Figure II-10, the information expansion path (E) is relatively flat because the loss associated with a Type I error is relatively large as compared to the loss associated with a Type II error (Leamer, 1978). If the loss associated with a Type II error is large relative to the loss associated with a Type I error, the information expansion path would be transposed as in Figure II-11.

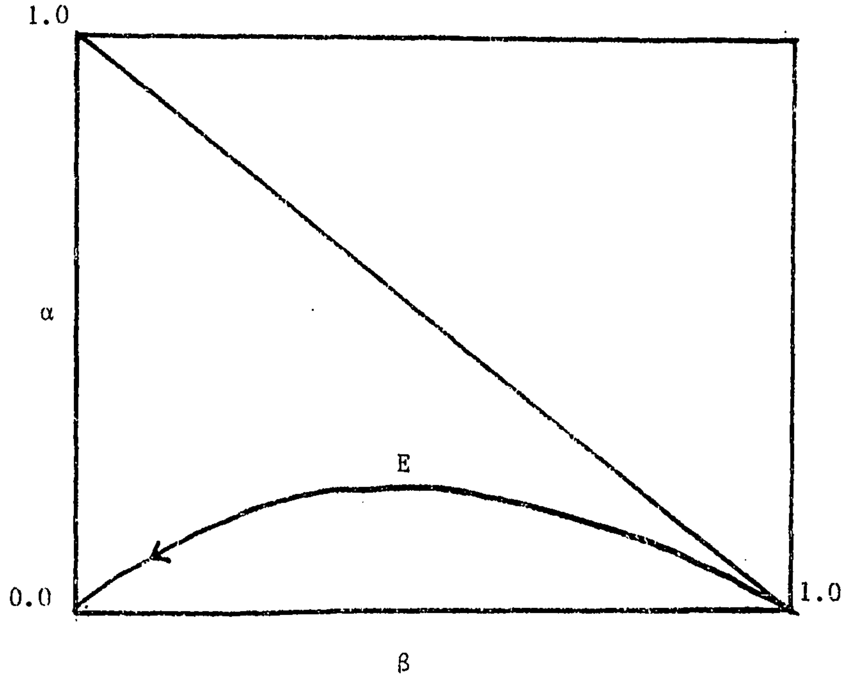
Relation of Approaches to this Study

The decision theory approach, as previously identified, directly considers the probability that financial statement amounts are correct and the costs associated with making a Type I or Type II error but does not require *ex ante* specification of α and β risk levels. By contrast, the classical approach and the tradeoff approach require prespecification of α and β risk levels and only indirectly consider the probability that financial statement amounts are correct and the costs associated with a Type I or Type II error. Because of this difference in orientation of the decision theory approach and because the decision theory approach has not been widely accepted by practitioners, this approach will not be considered further in this study.

How do auditors practically resolve this problem of establishing α and β risk levels when evaluating sample results of substantive

FIGURE II-10

DECISION THEORY APPROACH RESPONSE TO CHANGES IN THE AMOUNT OF EVIDENCE WHEN $L = 3\alpha + \beta$



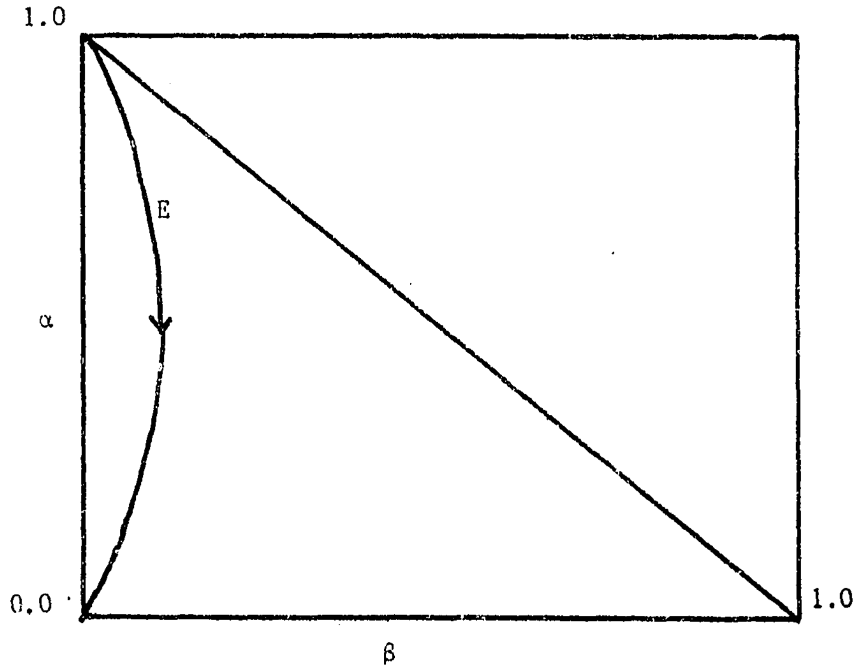
L: loss associated with an incorrect decision

Source: Leamer, 1978

FIGURE II-11

INFORMATION EXPANSION PATH FOR DECISION THEORY APPROACH

WHEN $L = \alpha + 3\beta$



L: loss associated with an incorrect decision

Adapted: Leamer, 1978

tests when maintaining the planned levels of both α and β is infeasible? This question represents the essence of this study. Two possible approaches are:

1. Set $\alpha = 0.05$ and minimize β. Implicitly, this approach assumes that the Type I error is the more important error. To make this approach relevant to the audit context, this rule may be restated: set β at the planned level, and minimize α. This restatement results in the assumption that the Type II error is the more important error--an assumption more appropriate in the audit context (Elliott, 1973).

   (This approach will be alluded to as the "classical approach.")

2. Minimize the maximum of $L_1\alpha$ and $L_2\beta$ where:

   L_1 is the loss associated with rejecting a correctly stated amount (a Type I error) such as the cost of unnecessarily extending audit procedures, needlessly requiring adjustment to the balance, or needlessly rendering a qualified or adverse opinion.

   L_2 is the loss associated with accepting a materially misstated amount (a Type II error) such as the cost of potential lawsuits (both direct and indirect) and loss of professional reputation.

   As the amount of evidence increases (decreases), both α and β are decreased (increased). The relative probabilities are $\alpha/\beta = L_2/L_1$.

   (This approach will be alluded to as the "tradeoff approach.")

The "classical approach" has frequently been applied in auditing. Characteristically, the classical approach allows α to vary as needed in order to maintain the planned level of β risk. On the other hand, the tradeoff approach characteristically results in adjustment to both risk levels so as to maintain the ratio of α to β equal to the ratio of L_1 to L_2.

The classical approach, by using an expansion path which fixes β risk at the planned level for evaluation purposes, is insensitive to different levels of α risk. If this insensitivity were reflected in the auditor's strategy at the planning stage, then for sampling efficiency, the sample size, n, would be established at a nominal level. Thus, the α risk level would be a function of the sample size, the estimated population standard deviation, the population size, the amount of a material error, and the β risk level. However, such a strategy is not employed during the planning stage using the classical approach. Rather, the α risk level is independently established in light of relevant audit factors. Therefore, for the classical approach there exists a discontinuity between the planning stage and the evaluation stage with respect to the level of α risk employed (i.e., α risk is independently established for the planning stage but is functionally dependent in the evaluation stage).

To overcome this discontinuity, an audit decision strategy should adjust α and β risk levels such that $\frac{U(\alpha)}{U(\beta)}$ is a constant for all α, β risk descriptions along the expansion path dictated by that strategy. The tradeoff approach comes closer to achieving that relationship than does the classical approach. The tradeoff approach characteristically dictates adjustments to α and β risk levels such that $\frac{\alpha}{\beta}$ is a constant. This adjustment procedure produces a ratio of $U(\alpha)$ to $U(\beta)$ that may approximate a constant so long as $U(\alpha)$ and $U(\beta)$ vary proportionately with changes in α and β risk, respectively.

On the other hand, the classical approach characteristically produces a ratio of $U(\alpha)$ to $U(\beta)$ that decreases as the level of α

risk increases and vice versa (assuming that there is lower utility associated with higher levels of α risk). That is, since $U(\alpha)$ is decreasing for all points on the expansion path such that α exceeds α_p and $U(\beta)$ remains constant (because the level of β risk is maintained at the planned level), then $\frac{U(\alpha)}{U(\beta)}$ decreases. Thus the classical approach results in adjustment to the α risk level such that the ratio of $U(\alpha)$ to $U(\beta)$ does not approximate a constant.

Conclusion

Several approaches to measuring and controlling auditors' α and β risk levels with respect to substantive testing exist. The more traditional approaches, those that do not allow α risk to vary, fail to actively control for α risk. In effect, α risk is ignored in order to maintain the pre-specified level of β risk. This result seems to indicate an inefficient use of resources in that evidence obtained in the audit have no bearing on the amount of β risk that is assumed.

A decision theory approach requires consideration of both α and β risk levels when evaluating the sample results of substantive tests. However, this approach requires probability assignments which require subjective input that has proven to be a source of considerable disagreement among auditors. Moreover, the decision theory approach requires absolute measure of the losses associated with each risk whereas the tradeoff approach requires only a relative measure of the two losses. Thus, use of this approach overcomes the shortcoming of the traditional approach but creates other problems in its place.

The "tradeoff approach" avoids the shortcoming of the classical approach and the need for subjective probability specifications of a decision theory approach. In general, this approach should be preferred on the basis of its conceptual merit in that shortcomings of available approaches are avoided and that both α and β risk levels are determined in a manner that considers all of the available evidence. However, no approach is necessarily optimal in all situations.

CHAPTER III

RESEARCH DESIGN AND METHOD

A series of four case situations was developed to collect data regarding approaches used to establish α and β risk levels in the evaluation of substantive statistical tests. The purpose of the survey was to empirically determine the attitudes of auditors who are involved in the quantitative phase of the evaluation of sample results. Specifically, auditor attitudes about two possible approaches to establishing α and β risk levels at the evaluation stage were of interest. This chapter explains the research design and method that was employed in the conduct of this study.

Selection of Research Strategy

Empirical research may conform to one of several research strategies. Six empirical research strategies suggested by Stone include: the laboratory experiment, simulation, the field experiment, the field study, sample survey, and the case study. The major determinant in the selection of a research strategy is the nature of the problem under study (Stone, 1978). Here, data for testing the four stated hypotheses was required. Therefore, the opportunity to obtain under controlled conditions auditor decisions concerning the evaluation of substantive tests was necessary.

With the nature of this research in mind, the six research strategies were compared by considering seven dimensions along which research strategies differ (Stone, 1978):

1. Naturalness of setting,
2. Generalizability,
3. Control,
4. Artifacts,
5. Study induced changes in the researcher,
6. Strength and range of studied variables, and
7. Cost.

Consideration of these seven dimensions revealed that none of the six strategies had "favorable" ratings on all dimensions. Rather, the writer made tradeoffs according to those dimensions most necessary in the research.

In this research, the potential for control and manipulation of variables and the potential for testing causal hypotheses was important. These strengths were needed in order to provide internal validity for this research. However, in order to obtain these strengths, the research sacrificed to some degree naturalness of setting and generalizability.

Analysis of the six research strategies in light of these desired strengths revealed that:

1. The laboratory experiment and simulation offer a high potential to manipulate independent variables and a high potential for testing causal hypotheses. These favorable characteristics are present because measurement in the laboratory/simulation takes place under

highly controlled conditions. However, as a result of maintaining these controls, the generalizability of results may be limited since laboratory/simulation settings may lack realism.

2. The field experiment and field study offer a high potential for generalizing the results of the experiment/study. These favorable characteristics result as a consequence of the fact that phenomena are studied in natural settings. However, because the experiment/study occurs in a natural setting, these research strategies offer a low potential (none in field studies) to manipulate independent variables and for testing causal hypotheses. These unfavorable characteristics are present because the degree of control over variables in the field is very low.

3. The sample survey offers a high potential for generalizing the results of the survey because of large sample sizes and low sampling error and because data are collected in natural settings. However, because they are conducted in natural settings, sample surveys have no potential for manipulating independent variables; as a result, causal inference from sample survey generated data are difficult to justify.

4. The case study offers no basis for manipulating variables, testing causal hypotheses, or generalizing from the results of the study. Rather, this strategy is better suited to generation of hypotheses than to testing of hypotheses.

From the above analysis, the laboratory experiment and simulation appear to be more appropriate strategies than the field experiment, field study, sample survey, and case study because the former yield more favorable ratings along those dimensions desired (potential for manipulating independent variables and testing of causal hypotheses) than the latter strategies. Comparing the laboratory experiment and simulation, the nature and timing of events in a laboratory experiment is completely determined by the researcher whereas events and their timing in a simulation are determined by both the researcher-established simulation rules and the behavior of the simulation participants. Because of time and cost limitations and of limited access to subject-participants, the laboratory experiment was selected rather than simulation as the research strategy for this study.

Choice of Research Design

This laboratory experiment was based upon a post-test-only control group design. A diagram of this design is (Campbell and Stanley, 1963)

Control group	R O
Treatment group	R X O

As indicated above, subjects were randomly assigned (R) into two groups: (1) the control group who did not receive exposure to the theory of a tradeoff between α and β risk levels and (2) the treatment group who did receive exposure (X) to the theory. Observations (O) consisted of subject responses to a series of audit situations in which an evaluation of sample results of substantive tests was necessary.

The post-test-only control group design was selected partly because this design controls for all of the following threats to internal validity (Campbell and Stanley, 1963)

history: This threat refers to events other than X (exposure) which may be responsible for observed differences between two different observations, O_1 (before exposure and), O_2 (after exposure). This threat is eliminated in this study by the choice of a post-test-only control group design. This research design does not utilize a pretest, thus there is only a single observation.

maturation: This threat refers to effects that occur systematically with the passage of time. In this study, maturation is a potential threat inasmuch as subjects are asked to read and respond to four similar decision cases. To control for this threat, the ordering of the decision cases was varied.

testing: This threat refers to differences between O_1 (observation before exposure) and O_2 (observation after exposure) due to measurement at two different periods in time. For example, interest measured at O_2 may be greater than at O_1. Choice of the post-test-only control group design eliminates this threat because no pretest is administered; that is, only one observation is made.

instrumentation: This threat refers to the chance that the ability of a measure to accurately index the measured variable changes over time. In this study, instrumentation is a potential threat because each subject is asked to rate his involvement in evaluating the sample results of substantive statistical tests. To control for this threat, the research instrument provides objective guidelines for assessing degree of involvement. Pretesting of the research instrument revealed no difficulty in making this assessment.

statistical regression: This threat refers to the likelihood that subjects who are assigned to treatments because of extreme scores on a less than perfectly reliable pretest have true scores that are nearer the mean score of all measured subjects. In this study, there is no pretest for the purpose of assigning subjects; rather, subjects are randomly assigned to treatments.

selection: This threat refers to the chance that differences noted in subject responses may be due to initial differences that existed between the two groups and not to effects of the treatment. To control for this threat, subjects were randomly assigned to treatment groups. Campbell and Stanley point out that the most adequate all-purpose assurance of lack of initial biases between groups is randomization.

mortality: This threat refers to observed differences between O_1 (observation before exposure) and O_2 (observation after exposure) due to individuals dropping out of a study between pretest and post-test periods. Choice of the post-test-only control group design eliminates this threat in this study since there is no pretest.

interactive effects: This threat refers to observed differences between O_1 (observation before exposure) and O_2 (observation after exposure) due to interaction of two or more of the above-mentioned phenomena. Choice of the post-test-only control group design eliminates this threat in this study since there is no pretest. Also, all of the above-mentioned threats are either eliminated by the choice of research design or are controlled.

An experiment is internally valid when one is able to conclude that an experimental treatment has indeed had an effect. To reach such a conclusion, the decision-maker must employ an experimental design that allows him to rule out the effects of the confounding factors discussed above (Campbell and Stanley, 1963).

In addition, the post-test-only control group design involved only two groups of subjects. Thus, this design requires fewer subjects than other experimental designs, such as the Solomon Four Group Design. This feature was most attractive in this study since the number of qualified subjects available was extremely limited (as discussed later). Because only two groups were used, no pretest was administered to the subjects. Although the pretest is a concept deeply imbedded in the thinking of researchers, its use is not essential to true experimental designs. Within the limits of confidence stated by tests of significance, randomization is sufficient assurance that the two groups are "equal" (Campbell and Stanley, 1963).

Factors Manipulated

Two factors that may cause subjects (in this case auditors) to interpret sample evidence differently are the severity of the risk of accepting a materially misstated amount and the amount of evidence available. Each of these factors was studied at two levels. To permit study of these factors and levels, the experimental design employed a 2 X 2 factorial plan. That is, observations consisted of all combinations of the two factors and levels that could be formed:

1. Lower risk of accepting a materially misstated amount and smaller amount of evidence,
2. Lower risk of accepting a materially misstated amount and larger amount of evidence,
3. Higher risk of accepting a materially misstated amount and smaller amount of evidence, and
4. Higher risk of accepting a materially misstated amount and larger amount of evidence.

Selection of Subjects

Subjects in this laboratory experiment were audit executives from three of the largest public accounting firms. Auditors whose backgrounds were likely to include involvement in the quantitative phase of the evaluation stage of substantive statistical sampling applications were selected for participation for two reasons. This group of auditors was desired because auditors who are involved in the quantitative evaluation stage of substantive statistical applications that employ hypothesis testing represent the population to which inference is desired. Also, selecting subjects whose backgrounds were likely to include the involvement desired was a means of insuring that the task to be performed in this experiment was a realistic one; that is, adequate surrogation was more likely to be achieved by selecting only those auditors who routinely in the field made decisions similar to the ones required in this experiment.

Experimental Task

The subjects were presented with a series of audit situations in which an evaluation of substantive statistical samples was necessary. Their task was to express preferences from among alternative ways of interpreting sample evidence. Subjects were informed in each case that the hypothetical client to be considered was a medium-sized manufacturer, that test extension was infeasible[12] with respect to the account in question, and that the firm has rendered an unqualified opinion on the client's financial statements for each of the last five years. In addition, subjects were provided with planning specifications, including (1) reliability assigned to internal control, (2) reliability assigned to supplemental audit procedures, (3) the amount of a material error, (4) the risk of rejecting a correctly stated amount, and (5) the estimated population standard deviation. They were also given achieved sample data, including the standard deviation of sample items, and the estimated audited value of the population.

Before responding to the decision cases, subjects were asked to rate their degrees of sampling involvement with respect to substantive tests as extensive, moderate, limited, or zero. Because subjects were selected on the basis of involvement in the quantitative phase of the evaluation stage of substantive statistical applications, responses to this self-classification were expected to yield mostly "extensive" ratings. Interjudge consistency in self-classification was vouchsafed by providing the following scale:

	Description of degrees of sampling involvement with respect to substantive testing		
Degree of Involvement	Involved in Execution At Least Once	Involved in Planning At Least Once	Involved in Analysis At Least Once
Zero	No	No	No
limited	Yes	No	No
moderate	Yes	Yes	No
extensive	Yes	Yes	Yes

Finally, subjects were asked to consider their expressed interpretation of the statistical evidence in each case and the subsequent audit action to accept or reject the account balance. Specifically, subjects were asked to indicate whether they would base their accept/reject decision on their interpretation of the statistical evidence or whether they would deem the statistical evaluation "inconclusive" and thus make their accept/reject decision judgmentally.

Research Instrument Construction and Pre-testing

In order to maximize the response rate of subjects (in this case, auditors) and, more importantly, to insure content validity of the decision situations, a carefully constructed research instrument was critical. In recognition of this need, the content of the decision situations was refined through a series of revisions.

First, a review of auditing textbooks and firm statistical sampling manuals served to identify the decision criteria and appropriate setting needed in the cases. That review combined with preconceptions of the writer and consultation with the dissertation committee chairman provided the basis for the first draft of the research instrument.

After its initial construction, several accounting graduate students having a background consisting of undergraduate and graduate auditing courses and, in most cases, experience with large, national accounting firms were asked to provide a general critique of the decision situations. After interviewing each of the participants, the decision situations were modified on the basis of suggestions received during the interviews. In addition, the research instrument, included in the writer's dissertation proposal, was presented at a workshop conducted by the accounting faculty of the University of Oklahoma for its faculty and graduate students. From comments and suggestions received during this presentation, the research instrument was further modified.

This revised instrument was then distributed to audit executives at the conclusion of a national auditing workshop conducted by a leading accounting firm for its auditing staff. These executives were asked to complete the revised research instrument and a debriefing questionnaire (see appendix B). The purpose of the debriefing questionnaire was to insure that certain terms and concepts used in the decision situations were interpreted as they were designed, to identify any other terms causing confusion, and to identify any other weaknesses present in the decision situations.

The research instrument, revised to reflect comments and suggestions received from the audit executives who were attending the national workshop, was subjected to a final review by two audit executives of a large, national accounting firm who were involved in field decisions similar to those decisions required by the research instrument. Those executives, interviewed by the writer, were asked to critique the decision situations with particular emphasis on the realism achieved in the cases. This critique provided suggestions that were incorporated into the final revision of the research instrument that was printed for distribution to subjects in this study. A copy of the research instrument appears in Appendix A.

Research Questions

Three specific research questions were to be investigated in this study. A discussion of each of those questions follows.

Research Question 1

What approach do auditors who are not exposed to the theory of a tradeoff between α and β risk levels employ when establishing those risk levels in the evaluation of substantive statistical tests? One hypothesis of this study was that auditors who were not exposed to this theory would employ the "classical" approach. This result is expected because auditing practitioners and researchers have suggested that the most appropriate approach to establishing α and β risk levels in these circumstances is to maintain β at the planned level and allow α to vary as needed (Elliott and Rogers, 1972). This condition in the auditing environment is likely to result in a psychological set as described by Kagan and Havemann (Kagan and Havemann, 1976). They define a psychological set as a preparatory readiness to make a particular response to a given stimulus. In this case, auditors are expected to have a preparatory readiness to maintain β risk at the planned level and allow α to vary as needed.

Research Question 2

Does exposure to the theory of a tradeoff between α and β risk levels affect the approach that auditors employ when establishing those risk levels in the evaluation of sample results of substantive statistical tests? It is hypothesized that exposure to the theory would affect the approach that auditors employ. This result is expected because the theory of a tradeoff between α and β risk levels has merit; that is, this theory offers the opportunity for auditors to maximize their degree of satisfaction. Its persuasiveness should be sufficient to sensitize the experimental group subjects to the benefits of considering a tradeoff between α and β risk levels. As Chang and Birnberg report, forms of inertia such as a psychological set can be overcome if events surrounding the task sensitize the subjects to the issue (Chang and Birnberg, 1977).

Research Question 3

Is rejection of the classical approach by auditors who have been exposed to the theory of a tradeoff between α and β risk levels a result of dissatisfaction with the classical approach or of benefits derived from using the approach offered by this theory? In this study, it is hypothesized that rejection of the classical approach would be the result of benefits derived from using the tradeoff approach. Study of this question provides insight as to why auditors who rejected the classical approach after exposure to the theory of a tradeoff between α and β risk levels did so. This insight may determine whether further research into application of this theory in auditing is warranted or whether further research into sample evaluation should be directed elsewhere.

Data Analysis

Data analysis in this study will be conducted using nonparametric statistical tests. Nonparametric tests are used in preference to parametric tests because the distributions of subject responses to the decision cases are not thought to resemble the normal curve. The presence of any of these conditions suggests that nonparametric tests are appropriate (Daniel, 1978). Random-based techniques will be used for data analysis although the sample was not randomly selected. Rather, subjects were selected based on their involvement in the evaluation of substantive statistical tests (as explained earlier in this chapter). For each statistical test performed in this study, the level of significance is .05. The specific statistical methods to be used to analyze the data in this study are discussed in the following paragraphs.

Research Question 1

Research question one (determining whether auditors consistently employ the classical approach) will be studied by observing responses of control group subjects to the decision cases. The decision cases require that subjects select from a series of decision intervals that reflect use of either the classical approach or the tradeoff approach. The distribution of responses by subjects who are not exposed to the theory is hypothesized to be stacked on the choice reflecting use of the classical approach. Because the participants in this study were not randomly selected, statistical inference with respect to the proportion of auditors who use the classical approach cannot be made. Accordingly, with respect to research question one, this study will merely report on a survey basis the frequencies as in Table III-1 that are observed in the sample.

Research Question 2

Research question two (whether exposure to the theory of a tradeoff between α and β affects the approach that auditors employ) will be studied by performing the Fisher exact test. This test is an extremely useful nonparametric technique for analyzing either nominal or ordinal data from two independent samples--particularly when the two samples are small in size. The test's use is appropriate when the data all fall into one or the other of two mutually exclusive classes (Daniel, 1978). For example, such data are obtained when two treatments are compared and subjects are classified as exhibiting or not exhibiting some characteristic. In this case, the two independent samples are the "control group" (those not exposed to the tradeoff approach) and the "experimental group" (those who are exposed to the tradeoff approach). The two mutually exclusive classes are "those employing the classical approach" and "those not employing the classical approach." The test determines whether the two groups differ in the

TABLE III-1

FREQUENCY OF CLASSICAL APPROACH USAGE BY CONTROL GROUP SUBJECTS

Category	Observed frequency
control group subjects employing the classical approach	a
control group subjects not employing the classical approach	b

a represents the number of control group subjects employing the classical approach,
b represents the number of control group subjects not employing the classical approach.

Adapted: Daniel, 1978.

proportion in which they fall into the two classes (Daniel, 1978).

Using the Fisher exact test, the frequencies from the two samples can be displayed in a 2 X 2 contingency table of the form in Table III-2. In this instance, the two hypotheses are:

H_o: The proportion of auditors employing the classical approach is the same for the control group and the experimental group.

$$(\text{i.e. } \frac{a}{A} = \frac{b}{B})$$

H_a: The proportion of auditors employing the classical approach is greater in the control group than in the experimental group.

$$(\text{i.e. } \frac{a}{A} > \frac{b}{B})$$

The alternative hypothesis is a one-tailed one in this case because the hypothesized result to investigation of research question two is that exposure to the tradeoff approach will cause auditors to respond differently to the decision situations. The conduct of the test requires that the test statistic, b (from Table III-2), be compared with the critical value of b. The decision rule is that H_o is rejected and thus H_a is accepted if the observed value of b is equal to or less than the critical value of b at the given level of significance. Otherwise, H_o is accepted at that level of signficance.

Research Question 3

Research question three (determining why auditors reject the classical approach) will also be studied by performing the Fisher exact test. Like research question two, the data are drawn from two indepen-

TABLE III-2

2 X 2 CONTINGENCY TABLE TO DISPLAY FREQUENCIES OF CONTROL GROUP AND EXPERIMENTAL GROUP SUBJECTS WHO EMPLOY THE CLASSICAL APPROACH VERSUS THOSE WHO DO NOT EMPLOY THE CLASSICAL APPROACH

Group	+	-	Total
Control	a	A-a	A
Experimental	b	B-b	B
Total	a+b	A+B-a-b	A+B

+ means the subject employs the classical approach,
- means the subject does not employ the classical approach,
a represents the number of subject in the control group employing the classical approach,
b represents the number of subjects in the experimental group employing the classical approach,
A represents the total number of subjects in the control group, and
B represents the total number of subjects in the experimental group.

Source: Daniel, 1978.

dent samples, and are classified as exhibiting or not exhibiting some characteristic. In this case, the two samples are "experimental group subjects employing the classical approach" and "experimental group subjects not employing the classical approach." The two classes are "those experimental group subjects for whom statistical evaluation was the basis for audit action" and "those experimental group subjects for whom nonstatistical evaluation was the basis for audit action." The test determines whether the two subsets of the experimental group differ in the proportion in which they fall into the two classes (Daniel, 1978).

Using the Fisher exact test, the frequencies from the two samples can be displayed in a 2 X 2 contingency table of the form in Table III-3. In this study, the two hypotheses are:

H_o: The proportion of auditors for whom statistical evaluation is the basis for audit action is the same for the two subsets of the experimental group.

$$(\text{i.e. } \frac{j}{J} = \frac{k}{K})$$

H_a: The proportion of auditors in the experimental group for whom statistical evaluation is the basis for audit action is greater for that subset of the experimental group which does not employ the classical approach than for that subset which does employ the classical approach.

$$(\text{i.e. } \frac{j}{J} > \frac{k}{K})$$

Because this study hypothesizes that rejection of the classical approach results from a greater degree of satisfaction in the statistical

TABLE III-3

2 X 2 CONTINGENCY TABLE TO DISPLAY FREQUENCIES OF EXPERIMENTAL GROUP SUBJECTS WHO EMPLOY THE CLASSICAL APPROACH VERSUS THOSE WHO DO NOT EMPLOY THE CLASSICAL APPROACH

Group	+	-	Total
experimental group subjects employing the classical approach	j	J-j	J
experimental group subjects not employing the classical approach	k	K-k	K
Total	j+k	J+K-j-k	J+K

+ means that statistical evaluation was the basis for audit action,
- means that non-statistical evaluation was the basis for audit action,
j represents the number of experimental group subjects who employ the classical approach and for whom statistical evaluation is the basis for audit action,
k represents the number of experimental group subjects who do not employ the classical approach and for whom statistical evaluation is the basis for audit action,
J represents the number of experimental group subjects employing the classical approach, and
K represents the number of experimental group subjects not employing the classical approach.

Source: Daniel, 1978.

evaluation when the tradeoff approach is employed, a one-tailed alternative hypothesis is appropriate. The conduct of the test requires that the test statistic, k (from Table III-3), be compared with the critical value of k. The decision rule is that H_o is rejected and thus H_a is accepted if the observed value of k is equal to or less than the critical value of k at the given level of significance. Otherwise, H_o is accepted at that level of significance (Daniel, 1978).

CHAPTER IV

DATA ANALYSIS

Data analyzed in this study are taken from responses to the series of decision cases. The twenty-four subjects are all auditors of one of three large international public accounting firms--10 from one firm, 2 from a second firm, and 12 from a third firm. All of the subjects participating in this study are involved in the evaluation of substantive statistical sampling tests. The subjects were not randomly selected but rather were judgmentally selected because they possess the desired attribute of involvement. This procedure of selecting subjects is deemed necessary to insure that the task to be performed in the study is a realistic one. However, subjects were randomly assigned to either the control or experimental group.

The four decision cases were developed so as to permit investigation of the three research questions formulated in Chapter III. The cases differ only with respect to two levels (one high, one low) of two factors. The two factors, "amount (n) of evidence" and "reliability assigned to internal control (IC)," are presented in varying combinations in the decision cases so that their effect on subject responses can be determined. The factor combinations present in the study are identified in Table IV-1. The remainder of Chapter IV sum-

TABLE IV-1

FACTOR COMBINATIONS APPEARING IN DECISION CASES

Case	Factor Combinations	
	Amount of Evidence	Reliability Assigned to Internal Control
A	n = 196	IC = .70
B	n = 100	IC = .70
C	n = 196	IC = none
D	n = 100	IC = none

marizes the data collected for each case and analyzes the results of investigation of the three research questions.

For each case presented in the research instrument, the subject's response can be classified according to whether he received exposure to the tradeoff approach, whether his audit conclusion reflects reliance or abandonment of the statistical evaluation and whether his response reflects use of the classical or tradeoff approach. Table IV-2 presents an overall breakdown with respect to this classification. Investigation of the three research questions in this study entails analysis of relevant subsets of this data.

Research Question 1

The propensity of auditors who are involved in the quantitative phase of evaluation of sample results of substantive statistical tests to consistently employ the classical approach is investigated in research question one. Data summarizing responses by control group subjects (those not exposed to the tradeoff approach) are presented in Table IV-3. As the Table shows, 80 percent of those auditors participating in the study employed the classical approach.

Research Question 2

Research question two considers whether exposure to the theory of a tradeoff between α and β risk levels affects the approach that auditors employ. Data summarizing usage of the classical approach (-)[13] or usage of an approach different from the classical approach (+) by control group subjects (those not exposed to the tradeoff approach) and experimental group subjects (those exposed to the tradeoff approach) are presented in Table IV-4.

TABLE IV-2

GENERAL DATA BREAKDOWN WITH RESPECT TO TREATMENT GROUP, STATISTICAL SAMPLING APPROACH USED, AND RELIANCE ON THE STATISTICAL EVALUATION

	Case A (n = 196, IC = .70)		Case B (n = 100, IC = .70)		Case C (n = 196, IC = .00)		Case D (n = 100, IC = .00)	
Subject Classifications	Classical Approach Response	Tradeoff Approach Response	Classical Approach Response	Tradeoff Approach Response	Classical Approach Response	Tradeoff Approach Response	Classical Approach Response	Tradeoff Approach Response
Experimental Group Subjects Relying on Statistical Evaluation	2	3	4	3	3	5	3	5
Experimental Group Subjects Abandoning Statistical Procedure	8	1	7	0	6	0	6	0
Control Group Subjects Relying on Statistical Evaluation	4	1	4	1	3	2	3	2
Control Group Subjects Abandoning Statistical Procedure	4	1	4	1	5	0	5	0

TABLE IV-3

USE OF THE CLASSICAL APPROACH BY CONTROL GROUP SUBJECTS

Group	Observed Frequency Case A (n = 196, IC = .70)	Case B (n = 100, IC = .70)	Case C (n = 196, IC = .00)	Case D (n = 100, IC = .00)
Control group subjects employing the classical approach	8	8	8	8
Control group subjects not employing the classical approach	2	2	2	2

TABLE IV-4

OBSERVED FREQUENCIES OF CLASSICAL APPROACH AND

NON-CLASSICAL APPROACH USE

	Case A (n = 196, IC = .70)			Case B (n = 100, IC = .70)			Case C (n = 196, IC = .00)			Case D (n = 100, IC = .00)		
Group	+	−	Total	+	−	Total	+	−	Total	+	−	Total
Experimental	4	10	14	3	11	14	5	9	14	5	9	14
Control	2	8	10	2	8	10	2	8	10	2	8	10
Total	6	18	24	5	19	24	7	17	24	7	17	24
	A = 14			A = 14			A = 14			A = 14		
	B = 10			B = 10			B = 10			B = 10		
	a = 4			a = 3			a = 5			a = 5		
	b = 2			b = 2			b = 2			b = 2		

+ means that the subject employs an approach different from the classical approach
− means that the subject employs the classical approach
A represents the number of subjects in the experimental group
B represents the number of subjects in the control group
a represents the number of subjects in the experimental group employing an approach different from the classical approach
b represents the number of subjects in the control group employing an approach different from the classical approach

To test the null hypothesis that the proportion of auditors employing the classical approach is the same for the control group and the experimental group, the Fisher exact test[14] requires that the test statistic, b, the observed frequency in the control group[15] that employed the classical approach, be compared to the critical value of b at the given level of signficance. Table IV-5 summarizes the critical b values at the .05 level of significance, computed test statistics, and decisions to accept or reject H_o for each case. As Table IV-5 indicates, no significant differences are found to exist between the two groups in any of the cases. Thus, the null hypothesis for each case that the proportion of auditors employing the classical approach is the same for the control group and the experimental group cannot be rejected. Because the null hypothesis cannot be rejected for any of the cases, the factors "amount of evidence" and "reliability assigned to internal control" apparently do not significantly affect the auditor's selection of an approach to establish α and β risk levels during sample evaluation of substantive statistical tests. Although no significant difference is detected, there is a noticeable directional effect. That is, the proportions (4/14, 3/14, 5/14, and 5/14 in Cases A, B, C, and D, respectively) of subject-auditors in the experimental group not employing the classical approach is greater than the proportions (2/10, 2/10, 2/10, and 2/10 in Cases A, B, C, and D, respectively) of subject-auditors in the control group not employing the classical approach. The null hypothesis in cases A, B, C and D could be rejected at the .335, .385, .260 and .260 levels of significance, respectively.

TABLE IV-5

CRITICAL VALUES, TEST STATISTICS, AND ACCEPT/REJECT DECISIONS FOR DETERMINING WHETHER EXPOSURE TO THE THEORY OF A TRADEOFF BETWEEN α AND β RISK LEVELS AFFECTS THE APPROACH THAT AUDITORS EMPLOY IN THE EVALUATION OF SUBSTANTIVE STATISTICAL SAMPLING TESTS

Case	Critical Value[16]	Test Statistic	Accept/reject H_o
A	–	2	Accept H_o
B	–	2	Accept H_o
C	0	2	Accept H_o
D	0	2	Accept H_o

Research Question 3

Determining why auditors who have been exposed to the tradeoff approach reject the classical approach is investigated in research question three. Data summarizing responses by experimental group subjects (those exposed to the tradeoff approach) are presented in Table IV-6. One characteristic of the data in Table IV-6 is that J, the number of experimental group subjects employing the classical approach, varies among the four cases. This result is apparently due to the decision by participating auditors to switch from classical to non-classical form of evaluation, or vice versa, as different levels of internal control reliance and/or evidence were observed.

To test the null hypothesis that the proportion of auditors for whom statistical evaluation is the basis for audit action is the same for the two subsets, "users of the classical approach (those subjects who selected a decision interval in which β was different from the planned level)" and "nonusers of the classical approach (those subjects who selected the decision interval in which β was maintained at the planned level)," of the experimental group, the Fisher exact test[17] requires that the test statistic k be compared with the critical value of k at the given level of significance. The test statistic in this situation represents the observed frequency of those auditors in the experimental group not employing the classical approach for whom statistical evaluation is the basis for audit action. Table IV-7 summarizes the critical k values at the .05 level of significance, the test statistics, and the decisions to accept or reject H_o for each of the four cases.

TABLE IV-6

OBSERVED FREQUENCIES OF BASES FOR AUDIT ACTION

FOR EXPERIMENTAL GROUP SUBJECTS

Group	Observed frequencies											
	Case A (n = 196, IC = .70)			Case B (n = 100, IC = .70)			Case C (n = 196, IC = .00)			Case D (n = 100, IC = .00)		
	+	−	Total	+	−	Total	+	−	Total	+	−	Total
Experimental group subjects employing the classical approach	8	2	10	7	4	11	6	3	9	6	3	9
Experimental group subjects not employing the classical approach	1	3	4	0	3	3	0	5	5	0	5	5
Total	9	5	14	7	7	14	6	8	14	6	8	14

	Case A	Case B	Case C	Case D
	J = 10	J = 11	J = 9	J = 9
	K = 4	K = 3	K = 5	K = 5
	j = 8	j = 7	j = 6	j = 6
	k = 1	k = 0	k = 0	k = 0

+ means that non-statistical evaluation was the basis for audit action
− means that statistical evaluation was the basis for audit action
J represents the number of subjects in the experimental group employing the classical approach
K represents the number of subjects in the experimental group not employing the classical approach
j represents the number of subjects in the experimental group employing the classical approach and for whom non-statistical evaluation was the basis for audit action
k represents the number of subjects in the experimental group not employing the classical approach and for whom non-statistical evaluation was the basis for audit action

TABLE IV-7

CRITICAL VALUES, TEST STATISTICS, AND ACCEPT/REJECT DECISIONS REGARDING H_o FOR DECISION CASES TO DETERMINE WHETHER SATISFACTION WITH THE STATISTICAL EVALUATION IS AFFECTED WHEN AUDITORS USE THE TRADEOFF APPROACH RATHER THAN THE CLASSICAL APPROACH

Case	Critical Value[18]	Test Statistic	Accept/reject H_o
A	0	1	Accept H_o
B	-	0	Accept H_o
C	0	0	Reject H_o
D	0	0	Reject H_o

As Table IV-7 indicates, significant differences are found to exist between the two subsets of the experimental group in Cases C (n = 196 and IC = .00) and D (n = 100 and IC = .00). Thus, the null hypothesis for Cases C and D that the proportion of auditors for whom statistical evaluation is the basis for audit action is the same for those two subsets can be rejected, but the null hypothesis cannot be rejected for Cases A (n = 196 and IC = .70) and B (n = 100 and IC = .70). Cases C and D differ from both Cases A and B in that the reliability assigned to internal control (IC) is low for the former cases and high for the latter cases. Thus, a low reliability assigned to internal control appears to significantly affect the degree of satisfaction with the tradeoff approach. On the other hand, the amount of evidence appears not to significantly affect the degree of satisfaction with the tradeoff approach since both Cases C and D had different levels of this factor; yet, the null hypothesis was rejected in each case. Apparently, rejection of the classical approach is then influenced by the presence of a poor system of internal control but not by the amount of available evidence. Thus, auditors appear to be satisfied with classical statisical results only when the client's system of internal control is reliable. The null hypothesis in both Cases C and D could be rejected only at a significance level lower than .028. The null hypothesis in Case A could be rejected at any significance level higher than .09 but the null hypothesis in Case B could not be rejected at any level of significance.

In summary, the findings of this study are:

1. Participating auditors who are involved in the quantitative phase of evaluation of sample results of substantive

statistical tests apparently have a preparatory readiness to maintain β risk at the planned level and allow α risk to vary as needed inasmuch as those auditors were found to consistently employ the classical approach,

2. Participating auditors exhibit a tendency toward inertia with respect to consideration of alternative approaches to the evaluation of sample results of substantive statistical tests since those auditors even when exposed to the tradeoff approach continue to consistently employ the classical approach,
3. Participating auditors who are involved in the quantitative phase of evaluation of sample results of substantive statistical tests and who reject the classical approach after receiving exposure to the tradeoff approach apparently reject the classical approach in the presence of a poor system of internal control.

Based on these findings, Chapter V presents the conclusions and recommendations of this study.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

This study represents the first effort to explore the theory of a tradeoff between α and β risk levels in auditing. This research differs from previous auditing research on three important points:

(1) The approach to establishing α and β risk levels has received little attention in the auditing literature,

(2) The theory of a tradeoff that exists between α and β risk levels has not been explored in an auditing context, and

(3) The existence of a psychological set and the presence of inertia have not been explored in an audit sampling context.[19]

In the past, use of the tradeoff theory has been overlooked or dismissed as being impractical to apply. As a means of testing this theory's relevance in the evaluation of sample results of substantive statistical tests, this research provides insights into three areas:

(1) Whether auditors who are involved in the evaluation of sample results of substantive statistical tests consistently employ the classical approach when evaluating those sample results,

(2) Whether exposure to the theory of a tradeoff between α and β risk levels affects the approach that those auditors employ when evaluating the sample results of substantive statistical tests, and

(3) Whether future research into the theory of a tradeoff between α and β risk levels is warranted in auditing with respect to the evaluation of sample results of substantive statistical tests.

With respect to these three areas, the findings of this study are that:

(1) Eighty percent of those auditors participating in this study who were not exposed to the tradeoff approach employed the classical approach when evaluating the sample results of substantive statistical tests,

(2) Exposure to the theory of a tradeoff between α and β risk levels did not significantly affect the approach that subject-auditors employed when evaluating the sample results of substantive statistical tests,

(3) Even though exposure to the theory of a tradeoff between α and β risk levels did not significantly affect the approach that subject-auditors employed when evaluating the sample results of substantive statistical tests, those subject-auditors employing the tradeoff approach exhibited a significantly greater propensity to rely upon the statistical evaluation in situations where internal control was judged to be poor than did their counterparts who relied upon the classical approach.

The degree to which the findings of this study are generalizable to actual auditing situations is limited by several conditions. These conditions include:

(1) The experimental setting in which the audit decisions were made and recorded,

(2) The small number of subject-auditors who were available for participation in the study,

(3) The consideration of only two approaches, both of which were studied only within the hypothesis testing mode, for evaluating the sample results of substantive statistical sampling tests,

(4) The consideration of only two factors that could affect the approach that an auditor might employ.

The choice of the laboratory experiment as the research strategy for this study was made, as Chapter III explains, primarily to control for threats to internal validity. Because the laboratory experiment is structured to control for those potentially confounding variables, the degree to which the experimental setting mirrors the intended natural setting is limited. For example, in this study the experimental setting did not permit the auditor to "gather additional evidence" before making his audit decision. Also, the study gave little if any information which would be of assistance to the auditor wishing to establish qualitative dimensions for evaluating sample results. In a real audit situation, that constraint ordinarily does not exist. Thus, as this point illustrates, generalizability is limited because creation of a natural setting was not completely achieved.

Because only a limited number of auditors were available for participation in the study, the findings may not be entirely representative. The nature of the problem under investigation required that only subject-auditors experienced in the evaluation of sample results of substantive statistical tests be considered; otherwise, the external validity would have been extremely limited because many subjects would have been required to complete an unfamiliar and, hence, an unrealistic task. Thus, a large number of auditors who did not possess this specialized experience were not considered for participation in the study.

Another matter that contributed to the low number of qualified subject-auditors available for participation was the hypothesis-testing orientation of the research instrument. When accounting firms were solicited for participation, some declined on the basis that their firms' approach was so different from the approach reflected in the decision cases that their employees would be unable to effectively deal with the decision requested. Thus, the number of qualified subject-auditors was further reduced.

From this reduced number of qualified and available subject-auditors, the participating subject-auditors were selected. Since these participants may or may not be representative of the population of all auditors who are involved in the quantitative phase of the evaluation stage of substantive statistical tests that employ hypothesis testing, the generalizability of this study's findings is limited.

This study considers only two approaches out of a number of approaches that are potentially available for use. Thus, even though no significant difference was detected between use of the two approaches considered in this study, that finding does not ensure that significant differences could not be detected if alternate approaches were considered. Hence, generalizability of this study's findings are further limited to the extent that other potentially relevant methods were not considered.

Only the two factors "amount of evidence" and "reliability assigned to internal control" were considered as affecting the selection of an approach for evaluating the sample results of substantive statistical tests. There may be additional factors that could influence the selection process. To the extent that other factors not considered influence the selection process, the generalizability of this study's findings is even further limited.

Though there are limitations concerning the generalizability of the findings of this study, the results nevertheless have some relevance to the evaluation of sample results of substantive statistical sampling tests. Strengths of the study such as (1) the controls for internal validity realized through use of the post-test-only control group design, (2) the close correspondence between the experimental tasks and the real-world tasks performed by the subject-auditors achieved by selecting only auditors who are involved in similar decisions in the field, and (3) the careful development of the research instrument by a series of critiques and pretests by qualified academic colleagues and audit executives warrant consideration of the results of this study.

The most important finding of this study is that satisfaction with the tradeoff approach was greater than satisfaction with the classical approach. That is, subject-auditors were more willing to base their audit decision on the statistical evidence when the tradeoff approach is used. This finding was found significant at the 5 percent level of significance in this study when the reliability assigned to internal control was low (.00 in this case). To the extent that this result can be replicated in future studies, its implication is that the auditing profession should consider providing support as well as guidelines for use of the tradeoff approach provided the approach is found to be theoretically appropriate vis-à-vis the philosophy of audit inference and decision. Furthermore, the curricula of auditing education and content of staff training programs might be modified to include efforts to ensure that auditors develop an understanding of this theory and acquire a knowledge of how it can be applied. Although the proportion of subject-auditors not employing the classical approach in the experimental group exceeded the proportion of those subject-auditors in the control group, no significant difference at the 5 percent level between use of the tradeoff approach and use of the classical approach was detected.

Even though no significant difference was detected, the absence of such a difference does not necessarily imply that subject-auditors do not personally prefer the tradeoff approach to the classical approach. The subject-auditor's use of the classical approach might be due to factors other than personal preference. For example, subject-auditors may have felt compelled to use their firm's approach regardless of personal preference. Also, subject-auditors may not have felt that generally accepted auditing standards would permit use of the tradeoff approach; that is, because authoritative auditing literature is replete with examples illustrating 5 percent α risk levels, auditors might view this occurrence as an implicit preference for if not a requirement to employ this level of α risk. In addition, the exposure given in the study to the tradeoff approach may not have sufficiently sensitized many of the subject-auditors to the benefits of this approach. Finally, the subject-auditor may not have opted for a preferred approach but rather opted for the classical approach because there was not the time nor perhaps the ability to learn about the tradeoff approach to use it correctly. In effect, the choice was made out of knowledge and not out of preference.

On the other hand, the directional effect noted above could be due to a potential bias in the methodology. This bias is potentially created by explaining the advantages and operation of the tradeoff approach to the experimental group subjects and then asking them to immediately respond to a set of decision cases which provide an opportunity to exercise this newly acquired knowledge. That is, the subject may respond in a manner consistent with the tradeoff approach because he feels that he is expected to use this "other" approach.

Not unexpectedly, when not exposed to the tradeoff approach, a high percentage (80 percent) of auditors employed the classical approach. A high rate of usage of the classical approach was expected because of the emphasis on this approach in auditing education, staff training programs, and auditing literature.

Because of the study's implications, further research into alternative approaches to evaluating the sample results of substantive tests seems appropriate. The tradeoff approach, which was found to give greater satisfaction to subject-auditors in this study, should be among those methods considered.

Since this study only considered two factors that might affect the auditor's interpretation of sample evidence, further research should consider whether other potentially relevant factors exist. Also, expanding the number of levels of each factor presented might be appropriate. For example, the finding in research question three was that auditors are better satisfied with the statistical evaluation when the tradeoff approach, as opposed to the classical statistical approach, is employed in a situation in which the system of internal control is poor. Such a finding was not observed in situations in which the system of internal control is excellent. In further research, the effect on the auditor's satisfaction with the statistical evaluation in the presence of moderately effective systems of internal control should be considered, and an explanation of the phenomenon actually observed should be formulated and tested.

Another idea for further research is to consider the effect of geographical location and/or firm affiliation. Auditors in certain sections of the country may be educated in or exposed to different methods of statistical evaluation. In addition, different accounting firms may train their staff to apply different techniques. Accordingly, responses to case situations similar to those presented in this study's decision cases may vary because of these characteristics.

Further research is also needed to establish the degree of agreement by auditors in establishing α and β risk levels when using the tradeoff approach. In other auditing research areas, such as elicitation of judgmentally-assessed prior probabilities in a decision theoretic setting, evidence has indicated that a considerable amount of diversity existed in responses among subjects as to specification of the judgmentally-assessed variable. Thus, investigation seems appropriate in this area.

Finally, additional research is needed to determine whether use of the tradeoff approach, or of other alternatives to the classical approach, results in different audit actions from those taken when the classical approach is employed. If the same audit conclusions are reached regardless of the approach employed, then there is little reason to further consider alternative approaches. Conversely, if difference audit conclusions are reached when different methods of statistical evaluation are employed, investigation of the relative benefits of each alternative is needed.

FOOTNOTES

[1]Which parameter value is designated as the null hypothesis is important because, in the process of hypothesis testing, the null hypothesis is given the benefit of the doubt. In statistics, the null hypothesis is frequently designated as the parameter to be rejected; that is, in this setting, rejection is the primary concern. This orientation is ordinarily appropriate because the decision-maker can never be certain of what he accepts based on statistical results, but he can be certain of rejection. However, consistent with the positive approach in auditing, the decision-maker is motivated to accept the book value; thus, the null hypothesis is designated as the parameter to be accepted. Kazmier, Leonard J., Basic Statistics for Business and Economics, p. 146. and Elliott, Robert K. and Rogers, John R., "Relating Statistical Sampling to Audit Objectives," Journal of Accountancy (July 1972), p. 46.

[2]Strictly speaking, the auditor does not assert that the financial statement amount is correct. Rather, when his statistical evaluation supports acceptance of the book value, he accepts the book value as being free from material misstatement.

[3]In a dollar-unit sampling approach, the sample size depends on the number of expected errors, the book value of the population, and the level of materiality. Guy, Dan M. Statistical Sampling in Auditing, p. 177.

[4]Conclusiveness criteria are needed when the book value is not supported by the statistical evidence. At this point in sampling procedures, both α and β risk must be carefully controlled. Alpha (α) must be controlled to limit the risk of rejecting a fairly stated book value. Beta (β) must be controlled to protect against accepting an adjusted book value that is materially misstated. Conclusiveness critera is the general term describing these maximum levels of α and β risk. (Ernst & Whinney, Audit Sampling, p. 158.

[5]Ernst & Whinney, Audit Sampling, p. 154.

[6]Leamer, Edward E., Specification Searches, p. 76.

[7]Guy, Dan M., Statistical Sampling in Auditing, p. 180.

[8] As indicated in Chapter I, α represents the risk of incorrect rejection, and β represents the risk of incorrect acceptance. "Risk of incorrect rejection" and "risk of incorrect acceptance" are terms used in SAS No. 39. American Institute of Certified Public Accountants, Statement on Auditing Standards, No. 39, p. 6. Throughout the study, the notation α and β will be used to refer to these terms.

[9] The term "material" amount is used throughout the auditing literature. SAS No. 39 introduces another term, "tolerable" amount. "Tolerable" amount is a planning concept and is related to the auditor's preliminary estimate of materiality. American Institute of Certified Public Accountants, Statement on Auditing Standards, No. 39, p. 7. The term "material" amount will be used throughout the remainder of this study.

[10] Some auditors feel that lower reliability levels are often appropriate. Arens, Alvin A. and Loebbecke, James K., Auditing: An Integrated Approach, p. 415.

[11] Roberts identifies different approaches that an auditor may take--a positive approach or a negative approach. In the positive approach, the auditor starts with the proposition that the recorded amount is correct and uses statistical evidence to support or reject that proposition. Following the negative approach, the auditor begins with a proposition that the recorded amount reflects a material misstatement; when the statistical evidence renders the proposition implausible, the auditor decides that the recorded amount is not materially misstated. Roberts further states that the two approaches are equivalent in the sense that when the desired risks are the same, the two approaches will result in the same decision. The positive approach will be employed throughout this study. Roberts, Donald M., Statistical Auditing, pp. 41, 45.

[12] The assumption that test extension is infeasible is not an unrealistic one. For example, to audit inventory, an observation of the inventory count is usually necessary. However, subsequent to the date of the physical inventory count, sales and purchases of inventory destroy the sampling frame that previously existed at the date of the count. In such circumstances, test extension is infeasible.

[13] The contingency table must be arranged such that $A \geq B$ and $a|A \geq b|B$. Finney, D. J., "The Fisher-Yates Test of Significance in 2 X 2 Contingency Tables," Biometrika (May 1948), p. 146.

[14] The justification for choosing the Fisher Exact Test is explained in Chapter III.

[15] The contngency table must be arranged such that $A \geq B$ and $a|A \geq b|B$. Finney, D. J., "The Fisher-Yates Test of Significance in 2 X 2 Contingency Tables," Biometrika (May 1948), p. 146.

[16]A dash, or absence of any entry, for some combination of A, B, and a indicates that no contingency table in that class is significant at the given level of significance. Finney, D. J., "The Fisher-Yates Test of Significance in 2 X 2 Contingency Tables," Biometrika (May 1948), p. 146.

[17]The justification for choosing the Fisher Exact Test is explained in Chapter III.

[18]A dash, or absence of any entry, for some combination of A, B, and a indicates that no contingency table in that class is significant at the given level of significance. Finney, D. J., "The Fisher-Yates Test of Significance in 2 X 2 Contingency Tables," Biometrika (May 1948), p. 146.

[19]Some exploration of the effects of environmental components on the periods preceding the audit, during the audit, and after the audit has been undertaken. Gibbins, Michael and Wolf, Frank M., "Auditors' Subjective Decision Environment--The Case of a Normal External Audit," The Accounting Review (January 1982), p. 121.

BIBLIOGRAPHY

American Institute of Certified Public Accountants, Statement on Auditing Standards, No. 39. New York: AICPA, 1981.

Arens, Alvin A. and Loebbecke, James K. Auditing An Integrated Approach, Second Edition. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1980.

Bedingfield, James P. "The Current State of Statistical Sampling and Auditing," Journal of Accountancy, December, 1975, pp. 48-55.

Birnberg, Jacob G. and Nath, Raghu. "Laboratory Experimentation in Accounting Research," The Accounting Review, January, 1968, pp. 38-45.

Campbell, Donald T. and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research, Third Printing. Chicago: Rand McNally & Company, 1963.

Chang, Davis L. and Birnberg, Jacob G. "Functional Fixity in Accounting Research: Perspective and New Data," Journal of Accounting Research, Autumn, 1977, pp. 300-312.

Chesley, G. R. "Elicitation of Subjective Probabilities: A Laboratory Study in an Accounting Context," Journal of Accounting Research, Spring, 1976, pp. 27-48.

Cochran, William G. and Cox, Gertrude M. Experimental Designs. New York: John Wiley & Sons, Inc., 1957.

Corless, John C. "Assessing Prior Distributions for Applying Bayesian Statistics in Auditing," The Accounting Review, July, 1972, pp. 556-566.

Daniel, Wayne W. Applied Nonparametric Statistics. Boston: Houghton Mifflin Company, 1978.

Dickhaut, J. L. and Watson, D. "On the Use of Surrogates in Behavioral Experimentation," Supplement to The Accounting Review, 1972, pp. 455-471.

Edwards, Allen L. Techniques of Attitude Scale Construction. New York: Appleton-Century-Crofts, Inc., 1957.

Elliott, Robert K. "Disagreement with Conclusions in article 'Statistical Risk and Sample Reliability,'" The CPA Journal, November, 1973, pp. 953.

Elliott, Robert K. and Rogers, John R. "Relating Statistical Sampling to Audit Objectives," The Journal of Accountancy, July, 1972, pp. 46-55.

Ernst and Whinney. Audit Samping. 1980.

Felix, William L. Jr. "A Decision Theory View of Auditing," in Contemporary Auditing Problems. (Lawrence, Kansas: University of Kansas Printing Service, 1974): 63-71.

Finney, D. J. "The Fisher Rates Test of Significance in 2 X 2 Contingency Tables," Biometrika, May, 1948, pp. 145-156.

Gibbins, Michael and Wolf, Frank M. "Auditors' Subjective Decision Environment--The Case of a Normal External Audit," The Accounting Review, January, 1982, pp. 105-124.

Gibbons, Jean Dickinson. Nonparametric Methods for Quantitative Analysis. New York: Holt, Rinehart and Winston, 1976.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, Inc., 1954.

Guy, Dan M. Statistical Sampling in Auditing. New York: John Wiley & Sons, Inc., 1981.

Hirsch, Maurice L. Jr. "Disaggregated Probabilistic Accounting Information: The Effect of Sequential Events on Expected Value Minimization Decisions," Journal of Accounting Research, Autumn, 1978, pp. 254-269.

Holscher, Harry H. Simplified Statistical Analysis. Boston: Cahners Publishing Co., Inc., 1971.

Jaedicke, Robert K. Ijiri, Yuji and Nielsen, Oswald. "The Effect of Accounting Alternatives on Management Decisions," in Research in Accounting Measurement. (Menasha, Wisconsin: American Accounting Association, 1966): 186-199.

Jensen, Donald R., Beus, G. B. and Storm, George. "Simultaneous Statistical Tests on Categorical Data," The Journal of Experimental Education, Summer, 1968, pp. 46-56.

Kagan, J. and Havemann E. Psychology: An Introduction. New York: Harcourt Brace Jovanovich, 1976.

Kazmier, Leonard J. Basic Statistics for Business and Economics. New York: McGraw-Hill, 1979.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1964.

Kern, Charles L. Neyhart, Charles A. Jr. and Hock, Clayton A. "Statistical Risk and Sample Reliability," The CPA Journal, January, 1973, pp. 55-60.

Kinney, William R. Jr. "A Decision-Theory Approach to the Sampling Problem in Auditing," Journal of Accounting Research, Spring, 1975, pp. 117-132.

Knoblett, James A. "The Applicability of Bayesian Statistics in Auditing," Decision Sciences, 1970, pp. 423-439.

Kyburg, Henry E. Jr. and Smokler, Howard E. Studies in Subjective Probability. New York: John Wiley & Sons, Inc., 1964.

Kyburg, Henry E. Classical Statistical Theory. New York: John Wiley & Sons, Inc.

Leamer, Edward E. Specification Searches. Los Angeles: John Wiley & Sons, Inc., 1978.

Li, Jerome C. R. Statistical Inference, Volume 1. Ann Arbor, Michigan: Edwards Brothers, Inc., 1964.

Likert, Rensis. "A Technique for the Measurement of Attitudes," in Archives of Psychology (New York, 1932): 5-55.

Miller, Rupert G. Jr. Simultaneous Statistical Inference. New York: McGraw-Hill Book Company, 1966.

Moore, Samuel A. II. "Organizational Inertia and Resistance to Change," The Educational Forum, November 1976, pp. 33-36.

Murphy, Gardner and Likert, Rensis. Public Opinion and the Individual. New York: Harper & Brothers Publishers, 1937.

Roberts, Donald M. Statistical Auditing. New York: American Institute of Certified Public Accountants, Inc., 1978.

Siegel, Sidney. Nonparametric Statistics. New York: McGraw-Hill Book Company, 1956.

Sorenson, James E. "Bayesian Analysis in Auditing," The Accoutning Review, July, 1969, pp. 555-561.

Stone, Eugene. Research Methods in Organizational Behavior. Santa Monica, California: Goodyear Publishing Company, Inc., 1978.

Appendix A

Research Instrument and Summary

of Subjects Responses

Evaluation of Sample Results of Substantive Statistical Sampling Tests in Auditing

When statistical sampling is used as the basis for determining whether to accept or reject an account balance, one of three situations will exist depending on the difference between the book value and the estimated audited value as depicted in Figure 1.

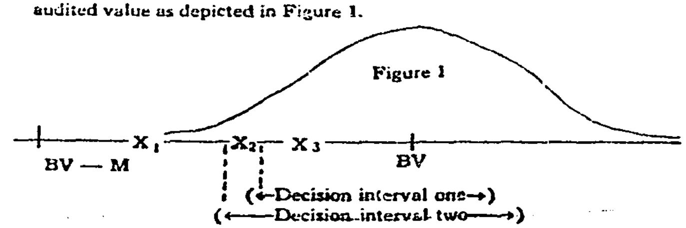


Figure 1

When the estimated audited value, X_1, differs radically from the book value (BV), the statistical evidence fails to support acceptance of the book value regardless of whether decision interval one or two is used. Similarly, when the estimated audited value, X_3, closely approximates the book value, the statistical evidence supports acceptance of the book value regardless of whether decision interval one or two is used. In either of these situations, the decision to accept or reject the book value is not very sensitive to changes in the width (amount of precision) of the decision interval. Thus, in these situations, construction of the decision interval endpoints is not a very critical step.

On the other hand, when the estimated audited value, X_2, differs by more than a small amount but by less than a material (M) amount, the decision to accept or reject the book value turns on what decision interval is used (in this exemple, interval one or two). In this situation, the decision to accept or reject the book value is sensitive to changes in the width (amount of precision) of the decision interval. Thus, in closecall decisions, determining the decision interval endpoints is a critical step. It is this situation that this study will examine: in closecall decisions, what approach do auditors take in the construction of decision intervals when evaluating the sample results of substantive tests? In order to answer this question, we will consider statistical sampling risks, α and β, inasmuch as control of these risks determines the amount of precision and thus the width of the decision interval.

(In the research instrument distributed to control group subjects, the above discussion was omitted.)

EVALUATION OF SAMPLE RESULTS WHEN THE ACHIEVED STANDARD DEVIATION EXCEEDS THE PLANNED STANDARD DEVIATION

The essential problem faced by an auditor when evaluating the sample results of substantive statistical tests is to select those levels of α and β risk that are best given available evidence. In this context, α risk (a type I error) is the probability of rejecting a correctly stated amount, and β risk (a type II error) is the probability of accepting a materially misstated amount. When selecting the appropriate levels of α and β risk, the auditor must consider the consequences of making an incorrect decision as follows:

Actions	Amount is Correctly Stated	Amount is Materially Misstated
Accept the book value	No error	Type II error[2] (β risk)
Reject the book value	Type I error[1] (α risk)	No error

[1] Consequences of this incorrect decision may include the cost of unnecessarily extending audit procedures, needlessly requiring adjustment to the balance, or needlessly rendering a qualified or adverse opinion.

[2] Consequences of this incorrect decision may include the cost of potential lawsuits (both direct and indirect) and loss of professional reputation.

The auditor must exercise professional judgement when considering the tradeoff between α and β so as to select those levels of α and β that are most appropriate in the circumstances. The solution is not obvious because it is desirable to have both α and β small; yet, a decrease in one risk level necessarily increases the other given a certain amount of evidence when the achieved standard deviation exceeds the planned standard deviation.

One approach to resolving this problem at the evaluation stage is to set β at the planned level and allow α to vary as necessary (we call this the classical approach). However, other solutions can be calculated which are also statistically acceptable. We want to make you aware of an approach to selecting α and β risk levels which is different from the classical approach. We will then solicit your response to a series of case situations in which you will be asked to express your preference among various statistically acceptable evaluations of sample data.

The tradeoffs between α and β that are available to the auditor can be depicted as in Figure 2.

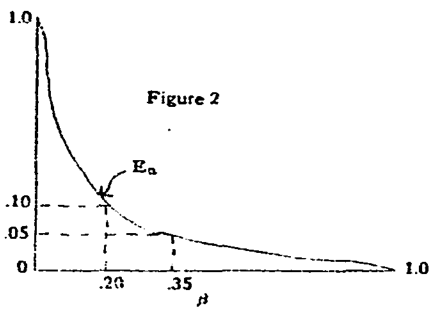


Figure 2

The curved line (E_n) represents the tradeoffs between α and β that are available when a given amount of evidence (a particular sample size, n) is available. Any combination of α and β levels lying on the line E_n is a statistically acceptable interpretaton of the available evidence. For example, in Figure 1 two acceptable interpretations are: (1) $\alpha = .05$ and $\beta = .35$ and (2) $\alpha = .10$ and $\beta = .20$.

An alternative to the classical approach to selecting α and β risk levels is an approach which attempts to balance the consequences of both types of risk (we call this the tradeoff approach). Using the tradeoff approach, an auditor would accept some increase in both risk levels rather than fixing one risk level (β) and allowing the other (α) to increase as necessary. Each risk level would be increased proportionately using the tradeoff approach so as to maintain the planned ratio of α to β.

In summary, the classical approach represents one way to resolve the problem of how to establish α and β risk levels. However, this approach has the shortcoming that α risk is not controlled in spite of the possible consequences of both kinds of risk. An approach that considers consequences of both kinds of risk, the tradeoff approach, avoids this shortcoming by permitting controlled increases in both α and β risk levels in such a way that the planned ratio of α to β is maintained.

INSTRUCTIONS

This exercise is a part of an audit sampling research study conducted by James H. Thompson, a doctoral student at the University of Oklahoma. Your participation and sincere effort are greatly appreciated.

Enclosed you will find four cases concerning statistical evaluation of sample results. The content of the cases differs only with respect to the effectiveness of internal controls and sample size. In each case you are asked to indicate how you would interpret the statistical evidence by choosing from among the choices that are provided.

Before beginning the exercise, rate your involvement in statistical sampling with respect to *substantive* tests as zero, limited, moderate, or extensive using the following guidelines:

Guidelines concerning degree of involvement with *substantive* statistical sampling tests

Involved in Execution At Least Once	Involved in Planning At Least Once	Involved in Analysis At Least Once	Degree of Involvement (check one)	
No	No	No	zero	____
Yes	No	No	limited	____
Yes	Yes	No	moderate	____
Yes	Yes	Yes	extensive	____

In addition, please supply the following additional information:

Years of experience in auditing ____

Current position in firm ____

Griffin, Inc., an electronics firm operating out of the midwest, manufactures component parts used in the cable television industry. As your firm's client, Griffin has requested an audit of its financial statements. Your firm has audited Griffin's financial statements for each of the last five years. In each of those years, Griffin has received an unqualified opinion on its statements.

Planning Data

You are responsible for determining the reasonableness of the accounts receivable balance. Financial data projected at the end of the current year are:

Accounts receivable balance	$ 2,500,000
Total assets	14,000,000
Equity	3,500,000
Net income	875,000

It has already been verified, based on this year's study and evaluation of internal controls, including tests of compliance utilizing attributes sampling techniques, that internal controls for the accounts receivable system are excellent. Accordingly, the reliance assigned to internal control is .70. The reliance assigned to supplemental procedures (such as analytical review of significant ratios and trends and resulting investigation of unusual fluctuations and questionable items) is .20. Thus, for a desired combined reliability of .95, the planned risk of accepting a materially misstated amount (β risk) is .2083.

The amount of a material error with respect to the accounts receivable balance has been determined to be $175,000. Based on this materiality threshhold, a planned β risk of .2083, and a planned risk of rejecting a correctly stated amount (α risk) of .05, desired precision is $123,827.

Based on desired precision of $123,827, a planned α risk of .05, and an estimated standard deviation of differences in accounts included in the accounts receivable balance of $1,768.95 (based on last year's achieved standard deviation of differences), a simple random sample (stratification was not practicable) of size 196 was calculated. Positive confirmations were obtained for most accounts. For those accounts in which replies to confirmation requests were not received, the amounts were verified by examining relevant contracts, shipping documents, and evidence that the buyer had acknowledged receipt of the shipment and by vouching transactions for a period around the balance sheet date.

The difference method of estimation was selected (as has been in previous audits) to determine the estimated audited value of the account. The difference method of estimation was selected because a substantial number of differences were expected (based on previous audits of accounts receivable) and because these differences have not appeared to be related to the magnitude of the book value (based on previous audits of accounts receivable). These differences have not been systematic nor material in the aggregate in the past and are not expected to be systematic nor material in this audit.

Execution

The estimated audited value of $2,385,000 was determined as follows:

d = (Sum of observed differences/n) = $45,080/196 = $230 (overstatement)

D = N · d = 500 · $230 = $115,000 (overstatement)

AV = BV + D = $2,500,000 — $115,000 = $2,385,000

where:

d = the mean difference of accounts in the sample from the book value

N = the number of accounts in the population of accounts receivable

D = the estimated total difference between book and audited values

AV = the estimated audited value

BV = the reported book value

In determining the mean difference of accounts in the sample from the book value (d), 118 accounts had zero differences and 78 accounts had non-zero differences ranging from an understatement of $200 to an overstatement of $1,250. The pattern of non-response to confirmation requests was as follows:

	Percentage of Sample Accounts Not Responding
Accounts with no observed differences	8%
Accounts with non-zero differences observed	10%

The observed differences were due to one of the following errors: improperly pricing invoices, misfooting invoices, billing for taxes to tax exempt organizations, billing for freight on sales with terms F.O.B. destination, improperly figuring sales discounts, or transposing digits in posting. No systematic differences were detected.

Evaluation

With the passage of time, extensive effort would be required to select and conduct tests of more accounts. The cost of this effort is thought to be prohibitive; thus, test extension is not a viable alternative.

After analyzing the sample data, a revised estimate of the standard deviation of differences of $2,122.75 was made. From the achieved sample results, an infinite number of decision intervals can be constructed the purpose of which is to determine whether the statistical evidence supports the correctness of the book value. Five such intervals are given below. Select from among those decision intervals that are given the *one* that best reflects your interpretation of the statistical evidence in this case:
(Note that there is not a "correct answer" to this exercise. Rather you are asked to select from the five decision intervals that are given the one that best reflects *your* interpretation of the statistical evidence. All of the decision intervals have been properly computed, and each represents a statistically acceptable solution. Although it is not necessary to verify the computations, calculations to support these five decision intervals are given in Appendix A at the end of this exercise.

___Decision interval: $2,351,407 to $2,648,593 with 95.0% reliability (Book value of $2,500,000 ± precision of $148,593); β risk is .3632

___Decision interval: $2,365,054 to $2,634,946 with 92.5% reliability (Book value of $2,500,000 ± precision of $134,946); β risk is .2981

___Decision interval: $2,375,667 to $2,624,333 with 90.0% reliability (Book value of $2,500,000 ± precision of $124,333); β risk is .2514

___Decision interval: $2,382,491 to $2,617,509 with 88.0% reliability (Book value of $2,500,000 ± precision of $117,509); β risk is .2236

___Decision interval: $2,386,281 to $2,613,719 with 86.6% reliability (Book value of $2,500,000 ± precision of $113,719); β risk is .2083

Consider your expressed interpretation of the statistical evidence above in terms of the decision to accept the account balance as being correct or to reject the account balance as being materially misstated. Which of the following positions better describes the course you would take:

Deem the statistical evaluation "inconclusive" and thus make the accept/reject decision judgementally

Base the accept/reject decision on your statistical evaluation expressed above

Markley, Inc., an electronics firm operating out of the midwest, manufactures component parts used in the cable television industry. As your firm's client, Markley has requested an audit of its financial statements. Your firm has audited Markley's financial statements for each of the last five years. In each of those years, Markley has received an unqualified opinion on its statements.

Planning Data

You are responsible for determining the reasonableness of the accounts receivable balance. Financial data projected at the end of the current year are:

Accounts receivable balance	$ 2,500,000
Total assets	14,000,000
Equity	3,500,000
Net income	875,000

It has already been verified, based on this year's study and evaluation of internal controls, including tests of compliance utilizing attributes sampling techniques, that internal controls for the accounts receivable system are excellent. Accordingly, the reliance assigned to internal control is .70. The reliance assigned to supplemental procedures (such as analytical review of significant ratios and trends and resulting investigation of unusual fluctuations and questionable items) is .20. Thus, for a desired combined reliability of .95, the planned risk of accepting a materially misstated amount (β risk) is .2083.

The amount of a material error with respect to the accounts receivable balance has been determined to be $175,000. Based on this materiality threshhold, a planned β risk of .2083, and a planned risk of rejecting a correctly stated amount (α risk) of .05, desired precision is $123,827.

Based on desired precision of $123,827, a planned α risk of .05, and an estimated standard deviation of differences in accounts included in the accounts receivable balance of $1,263.54 (based on last year's achieved standard deviation of differences), a simple random sample (stratification was not practicable) of size 100 was calculated. Positive confirmations were obtained for most accounts. For those accounts in which replies to confirmation requests were not received, the amounts were verified by examining relevant contracts, shipping documents, and evidence that the buyer had acknowledged receipt of the shipment and by vouching transactions for a period around the balance sheet date.

The difference method of estimation was selected (as has been in previous audits) to determine the estimated audited value of the account. The difference method of estimation was selected because a substantial number of differences were expected (based on previous audits of accounts receivable) and because these differences have not appeared to be related to the magnitude of the book value (based on previous audits of accounts receivable). These differences have not been systematic nor material in the aggregate in the past and are not expected to be systematic nor material in this audit.

Execution

The estimated audited value of $2,385,000 was determined as follows:

d = (Sum of observed differences/n) = $23,000/100 = $230 (overstatement)

$D = N \cdot d$ = 500 · $230 = $115,000 (overstatement)

$AV = BV + D$ = $2,500,000 — $115,000 = $2,385,000

where:

d = the mean difference of accounts in the sample from the book value

N = the number of accounts in the population of accounts receivable

D = the estimated total difference between book and audited values

AV = the estimated audited value

BV = the reported book value

In determining the mean difference of accounts in the sample from the book value (d), 60 accounts had zero differences and 40 accounts had non-zero differences ranging from an understatement of $120 to an overstatement of $960. The pattern of non-response to confirmation requests was as follows:

	Percentage of Sample Accounts Not Responding
Accounts with no observed differences	8%
Accounts with non-zero differences observed	10%

The observed differences were due to one of the following errors: improperly pricing invoices, misfooting invoices, billing for taxes to tax exempt organizations, billing for freight on sales with terms F.O.B. destination, improperly figuring sales discounts, or transposing digits in posting. No systematic differences were detected.

Evaluation

With the passage of time, extensive effort would be required to select and conduct tests of more accounts. The cost of this effort is thought to be prohibitive; thus, test extension is not a viable alternative.

After analyzing the sample data, a revised estimate of the standard deviation of differences of $1,516.25 was made. From the achieved sample results, an infinite number of decision intervals can be constructed the purpose of which is to determine whether the statistical evidence supports the correctness of the book value. Five such intervals are given below. Select from among those decision intervals that are given the one that best reflects your interpretation of the statistical evidence in this case:

(Note that there is not a "correct answer" to this exercise. Rather you are asked to select from the five decision intervals that are given the one that best reflects *your* interpretation of the statistical evidence. All of the decision intervals have been properly computed, and each represents a statistically acceptable solution. Although it is not necessary to verify the computations, calculations to support these five decision intervals are given in Appendix A at the end of this exercise.

__ Decision interval: $2,351,407 to $2,648,593 with 95.0% reliability (Book value of $2,500,000 ± precision of $148,593); β risk is .3632

__ Decision interval: $2,365,054 to $2,634,946 with 92.5% reliability (Book value of $2,500,000 ± precision of $134,946); β risk is .2981

__ Decision interval: $2,375,667 to $2,624,333 with 90.0% reliability (Book value of $2,500,000 ± precision of $124,333); β risk is .2514

__ Decision interval: $2,382,491 to $2,617,509 wth 88.0% reliability (Book value of $2,500,000 ± precision of $117,509); β risk is .2236

__ Decision interval: $2,386,281 to $2,613,719 with 86.6% reliability (Book value of $2,500,000 ± precision of $113,719); β risk is .2083

Consider your expressed interpretation of the statistical evidence above in terms of the decision to accept the account balance as being correct or to reject the account balance as being materially misstated. Which of the following positions better describes the course you would take:

__ deem the statistical evaluation "inconclusive" and thus make the accept/reject decision judgementally

__ base the accept/reject decision on your statistical evaluation expressed above

Thaxton, Inc., an electronics firm operating out of the midwest, manufactures component parts used in the cable television industry. As your firm's client, Thaxton has requested an audit of its financial statements. Your firm has audited Thaxton's financial statements for each of the last five years. In each of those years, Thaxton has received an unqualified opinion on its statements.

Planning Data

You are responsible for determining the reasonableness of the accounts receivable balance. Financial data projected at the end of the current year are:

Accounts receivable balance	$ 2,500,000
Total assets	14,000,000
Equity	3,500,000
Net income	875,000

It has already been verified, based on this year's study and evaluation of internal controls, including tests of compliance utilizing attributes sampling techniques, that internal controls for the accounts receivable system are poor. Accordingly, the reliance assigned to internal control is .00. The reliance assigned to supplemental procedures (such as analytical review of significant ratios and trends and resulting investigation of unusual fluctuations and questionable items) is .20. Thus, for a desired combined reliability of .95, the planned risk of accepting a materially misstated amount (β risk) is .0625.

The amount of a material error with respect to the accounts receivable balance has been determined to be $175,000. Based on this materiality threshhold, a planned β risk of .0625, and a planned risk of rejecting a correctly stated amount (α risk) of .05, desired precision is $98,281.

Based on desired precision of $98,281, a planned α risk of .05, and an estimated standard deviation of differences in accounts included in the accounts receivable balance of $1404.01 (based on last year's achieved standard deviation of differences), a simple random sample (stratification was not practicable) of size 196 was calculated. Positive confirmations were obtained for most accounts. For those accounts in which replies to confirmation requests were not received, the amounts were verified by examining relevant contracts, shipping documents, and evidence that the buyer had acknowledged receipt of the shipment and by vouching transactions for a period around the balance sheet date.

The difference method of estimation was selected (as has been in previous audits) to determine the estimated audited value of the account. The difference method of estimation was selected because a substantial number of differences were expected (based on previous audits of accounts receivable) and because these differences have not appeared to be related to the magnitude of the book value (based on previous audits of accounts receivable). These differences have not been systematic nor material in the aggregate in the past and are not expected to be systematic nor material in this audit.

Execution

The estimated audited value of $2,416,000 was determined as follows:

d = (Sum of observed differences/n) = $32,928/196 = $168 (overstatement)

D = N · d = 500 · $168 = $84,000 (overstatement)

AV = BV + D = $2,500,000 — $84,000 = $2,416,000

where:

d = the mean difference of accounts in the sample from the book value

N = the number of accounts in the population of accounts receivable

D = the estimated total difference between book and audited values

AV = the estimated audited value

BV = the reported book value

In determining the mean difference of accounts in the sample from the book value (d), 118 accounts had zero differences and 78 accounts had non-zero differences ranging from an understatement of $150 to an overstatement of $1,050. The pattern of non-response to confirmation requests was as follows:

	Percentage of Sample Accounts Not Responding
Accounts with no observed differences	8%
Accounts with non-zero differences observed	10%

The observed differences were due to one of the following errors: improperly pricing invoices, misfooting invoices, billing for taxes to tax exempt organizations, billing for freight on sales with terms F.O.B. destination, improperly figuring sales discounts, or transposing digits in posting. No systematic differences were detected.

Evaluation

With the passage of time, extensive effort would be required to select and conduct tests of more accounts. The cost of this effort is thought to be prohibitive; thus, test extension is not a viable alternative.

After analyzing the sample data, a revised estimate of the standard deviation of differences of $1,684.81 was made. From the achieved sample results, an infinite number of decision intervals can be constructed the purpose of which is to determine whether the statistical evidence supports the correctness of the book value. Five such intervals are given below. Select from among those decision intervals that are given the one that best reflects your interpretation of the statistical evidence in this case:

(Note that there is not a "correct answer" to this exercise. Rather you are asked to select from the five decision intervals that are given the one that best reflects your interpretation of the statistical evidence. All of the decision intervals have been properly computed, and each represents a statistically acceptable solution. Although it is not necessary to verify the computations, calculations to support these five decision intervals are given in Appendix A at the end of this exercise.

__Decision interval: $2,382,063 to $2,617,937 with 95.0% reliability (Book value of $2,500,000 ± precision of $117,937); β risk is .17[illegible]

__Decision interval: $2,392,894 to $2,607,106 with 92.5% reliability (Book value of $2,500,000 ± precision of $107,106); β risk is .129[illegible]

__Decision interval: $2,401,318 to $2,598,682 with 90.0% reliability (Book value of $2,500,000 ± precision of $ 98,682); β risk is .1026

__Decision interval: $2,409,742 to $2,590,258 with 86.7% reliability (Book value of $2,500,000 ± precision of $ 90,258); β risk is .0793

__Decision interval: $2,416,963 to $2,583,037 with 83.2% reliability (Book value of $2,500,000 ± precision of $ 83,037); β risk is .0625

Consider your expressed interpretation of the statistical evidence above in terms of the decision to accept the account balance as being correct or to reject the account balance as being materially misstated. Which of the following positions better describes the course you would take:

- deem the statistical evaluation "inconclusive" and thus make the accept/reject decision judgementally
- Base the accept/reject decision on your statistical evaluation expressed above

104

Winston, Inc., an electronics firm operating out of the midwest, manufactures component parts used in the cable television industry. As your firm's client, Winston has requested an audit of its financial statements. Your firm has audited Winston's financial statements for each of the last five years. In each of those years, Winston has received an unqualified opinion on its statements.

Planning Data

You are responsible for determining the reasonableness of the accounts receivable balance. Financial data projected at the end of the current year are:

Accounts receivable balance	$ 2,500,000
Total assets	14,000,000
Equity	3,500,000
Net income	875,000

It has already been verified, based on this year's study and evaluation of internal controls, including tests of compliance utilizing attributes sampling techniques, that internal controls for the accounts receivable system are poor. Accordingly, the reliance assigned to internal control is .00. The reliance assigned to supplemental procedures (such as analytical review of significant ratios and trends and resulting investigation of unusual fluctuations and questionable items) is .20. Thus, for a desired combined reliability of .95, the planned risk of accepting a materially misstated amount (β risk) is .0625.

The amount of a material error with respect to the accounts receivable balance has been determined to be $175,000. Based on this materiality threshhold, a planned β risk of .0625, and a planned risk of rejecting a correctly stated amount (α risk) of .05, desired precision is $98,281.

Based on desired precision of $98,281, a planned α risk of .05, and an estimated standard deviation of differences in accounts included in the accounts receivable balance of $1,002.87 (based on last years achieved standard deviation of differences), a simple random sample (stratification was not practicable) of size 100 was calculated. Positive confirmations were obtained for most accounts. For those accounts in which replies to confirmation requests were not received, the amounts were verified by examining relevant contracts, shipping documents, and evidence that the buyer had acknowledged receipt of the shipment and by vouching transactions for a period around the balance sheet date.

The difference method of estimation was selected (as has been in previous audits) to determine the estimated audited value of the account. The difference method of estimation was selected because a substantial number of differences were expected (based on previous audits of accounts receivable) and because these differences have not appeared to be related to the magnitude of the book value (based on previous audits of accounts receivable). These differences have not been systematic nor material in the aggregate in the past and are not expected to be systematic nor material in this audit.

Execution

The estimated audited value of $2,416,000 was determined as follows:

d = (Sum of observed differences/n) = $16,800/100 = $168 (overstatement)

$D = N \cdot d$ = 500 · $168 = $84,000 (overstatement)

$AV = BV + D$ = $2,500,000 — $84,000 = $2,416,000

where:

d = the mean difference of accounts in the sample from the book value

N = the number of accounts in the population of accounts receivable

D = the estimated total difference between book and audited values

AV = the estimated audited value

BV = the reported book value

In determining the mean difference of accounts in the sample from the book value (d), 60 accounts had zero differences and 40 accounts had non-zero differences ranging from an understatement of $100 to an overstatement of $690. The pattern of non-response to confirmation requests was as follows:

	Percentage of Sample Accounts Not Responding
Accounts with no observed differences	8%
Accounts with non-zero differences observed	10%

The observed differences were due to one of the following errors: improperly pricing invoices, misfooting invoices, billing for taxes to tax exempt organizations, billing for freight on sales with terms F.O.B. destination, improperly figuring sales discounts, or transposing digits in posting. No systematic differences were detected.

Evaluation

With the passage of time, extensive effort would be required to select and conduct tests of more accounts. The cost of this effort is thought to be prohibitive; thus, test extension is not a viable alternative.

After analyzing the sample data, a revised estimate of the standard deviation of differences of $1,203.44 was made. From the achieved sample results, an infinite number of decision intervals can be constructed the purpose of which is to determine whether the statistical evidence supports the correctness of the book value. Five such intervals are given below. Select from among those decision intervals that are given the one that best reflects your interpretation of the statistical evidence in this case:

(Note that there is not a "correct answer" to this exercise. Rather you are asked to select from the five decision intervals that are given the one that best reflects *your* interpretation of the statistical evidence. All of the decision intervals have been properly computed, and each represents a statistically acceptable solution. Although it is not necessary to verify the computations, calculations to support these five decision intervals are given in Appendix A at the end of this exercise.

__Decision interval: $2,382,063 to $2,617,937 with 95.0% reliability (Book value of $2,500,000 ± precision of $117,937); β risk is .1711

__Decision interval: $2,392,894 to $2,607,106 with 92.5% reliability (Book value of $2,500,000 ± precision of $107,106); β risk is .1292

__Decision interval: $2,401,318 to $2,598,682 with 90.0% reliability (Book value of $2,500,000 ± precision of $ 98,682); β risk is .1020

__Decision interval: $2,409,742 to $2,590,258 with 86.7% reliability (Book value of $2,500,000 ± precision of $ 90,258); β risk is .0793

__Decision interval: $2,416,963 to $2,583,037 with 83.2% reliability (Book value of $2,500,000 ± precision of $ 83,037); β risk is .0625

Consider your expressed interpretation of the statistical evidence above in terms of the decision to accept the account balance as being correct or to reject the account balance as being materially misstated. Which of the following positions better describes the course you would take:

_ deem the statistical evaluation "inconclusive" and thus make the accept/reject decision judgementally

_ base the accept/reject decision on your statistical evaluation expressed above

Appendix A—Calculations

Griffin case:

$\beta = \frac{1 - CR}{(1-IC)(1-SP)} = \frac{1 - .95}{(1 - .7)(1 - .2)} = .2083$

$A = \frac{M}{\left(1 + \frac{U_\beta}{U_R}\right)} = \frac{175{,}000}{1 + (.81/1.96)} = \$123{,}827$

$n = \left(\frac{S_{D_1} \cdot U_R \cdot N}{A}\right)^2 = \left(\frac{(1769.90)(1.96)(500)}{123{,}827}\right)^2 = 14^2 = 196$

Decision interval construction:

With 95.0% reliability: $A = \frac{(2122.75)(1.96)(500)}{14} = \$148{,}593$

With 92.5% reliability: $A = \frac{(2122.75)(1.78)(500)}{14} = \$134{,}946$

With 90.0% reliability: $A = \frac{(2122.75)(1.64)(500)}{14} = \$124{,}333$

With 88.0% reliability: $A = \frac{(2122.75)(1.55)(500)}{14} = \$117{,}509$

With 86.6% reliability: $A = \frac{(2122.75)(1.50)(500)}{14} = \$113{,}719$

Markley Case:

$\beta = \frac{1 - CR}{(1-IC)(1-SP)} = \frac{1 - .95}{(1 - .7)(1 - .2)} = .2083$

$A = \frac{M}{\left(1 + \frac{U_\beta}{U_R}\right)} = \frac{175{,}000}{1 + (.81/1.96)} = \$123{,}827$

$n = \left(\frac{S_{D_1} \cdot U_R \cdot N}{A}\right)^2 \quad \left(\frac{(1283.54)(1.96)(500)}{123{,}827}\right)^2 = 10^2 = 100$

Decision interval construction:

With 95.0% reliability: $A = \frac{(1516.25)(1.96)(500)}{10} = \$148{,}593$

With 92.5% reliability: $A = \frac{(1516.25)(1.78)(500)}{10} = \$134{,}946$

With 90.0% reliability: $A = \frac{(1516.25)(1.64)(500)}{10} = \$124{,}333$

With 88.0% reliability: $A = \frac{(1516.25)(1.55)(500)}{10} = \$117{,}509$

With 86.6% reliability: $A = \frac{(1516.25)(1.50)(500)}{10} = \$113{,}719$

For Griffin and Markley cases:

Achieved β: $U_\beta = \frac{(M - A)U_R}{A}$; β level then drawn from normal curve area table

For 1st decision interval: $U_\beta = \frac{(175{,}000 - 148{,}593)1.96}{148{,}593} = .35$; $\beta = .5 - .1368 = .3632$

For 2nd decision interval: $U_\beta = \frac{(175{,}000 - 134{,}946)1.78}{134{,}946} = .53$; $\beta = .5 - .2019 = .2981$

For 3rd decision interval: $U_\beta = \frac{(175{,}000 - 124{,}333)1.64}{124{,}333} = .67$; $\beta = .5 - .2486 = .2514$

For 4th decision interval: $U_\beta = \frac{(175{,}000 - 117{,}509)1.55}{117{,}509} = .76$; $\beta = .5 - .2764 = .2236$

For 5th decision interval: $U_\beta = \frac{(175{,}000 - 113{,}719)1.50}{113{,}719} = .81$; $\beta = .5 - .2910 = .2090$

Thaxton case.

$\beta = \frac{1 - CR}{(1-IC)(1-SP)} = \frac{1 - .95}{(1 - .0)(1 - .2)} = .0625$

$A = \frac{M}{\left(1 + \frac{U_\beta}{U_R}\right)} = \frac{175{,}000}{1 + (1.53/1.96)} = \$98{,}281$

$n = \left(\frac{S_{D_1} \cdot U_R \cdot N}{A}\right)^2 = \left(\frac{(1684.81)(1.96)(500)}{98{,}281}\right)^2 = 14^2 = 196$

Decision interval construction:

With 95.0% reliability: $A = \frac{(1684.81)(1.96)(500)}{14} = \$117{,}937$

With 92.5% reliability: $A = \frac{(1684.81)(1.78)(500)}{14} = \$107{,}106$

With 90.0% reliability: $A = \frac{(1684.81)(1.64)(500)}{14} = \$98{,}682$

With 86.7% reliability: $A = \frac{(1684.81)(1.50)(500)}{14} = \$90{,}258$

With 83.2% reliability: $A = \frac{(1684.81)(1.38)(500)}{14} = \$83{,}037$

Winston case:

$\beta = \frac{1 - CR}{(1-IC)(1-SP)} = \frac{1 - .95}{(1 - .0)(1 - .2)} = .0625$

$A = \frac{M}{\left(1 + \frac{U_\beta}{U_R}\right)} = \frac{175{,}000}{1 + (1.53/1.96)} = \$98{,}281$

$n = \left(\frac{S_{D_1} \cdot U_R \cdot N}{A}\right)^2 = \left(\frac{(1003.83)(1.96)(500)}{98{,}281}\right)^2 = 10^2 = 100$

Decision interval construction:

With 95.0% reliability: $A = \frac{(1203.44)(1.96)(500)}{10} = \$117{,}937\ 37$

With 92.5% reliability: $A = \frac{(1203.44)(1.78)(500)}{10} = \$107{,}106\ 06$

With 90.0% reliability: $A = \frac{(1203.44)(1.64)(500)}{10} = \$98{,}682\ 82$

With 86.7% reliability: $A = \frac{(1203.44)(1.50)(500)}{10} = \$90{,}258\ 58$

With 83.2% reliability: $A = \frac{(1203.44)(1.38)(500)}{10} = \$83{,}037\ 37$

For Thaxton and Winston cases:

Achieved β: $U_\beta = \frac{(M - A)U_R}{A}$; β level then drawn from normal curve area table

For 1st decision interval: $U_\beta = \frac{(175{,}000 - 117{,}937)1.96}{117{,}937} = .95$; $\beta = .5 - .3289 = .1711$

For 2nd decision interval: $U_\beta = \frac{(175{,}000 - 107{,}106)1.78}{107{,}106} = 1.13$; $\beta = .5 - .3708 = .1292$

For 3rd decision interval: $U_\beta = \frac{(175{,}000 - 98{,}682)1.64}{98{,}682} = 1.27$; $\beta = .5 - .3980 = .1020$

For 4th decision interval: $U_\beta = \frac{(175{,}000 - 90{,}258)1.50}{90{,}258} = 1.41$; $\beta = .5 - .4207 = .0793$

For 5th decision interval: $U_\beta = \frac{(175{,}000 - 83{,}037)1.38}{83{,}037} = 1.53$; $\beta = .5 - .4375 = .0625$

Symbols:

- CR [illegible]
- IC [illegible]
- SP [illegible]
- A [illegible]
- M [illegible]
- U_β [illegible]
- U_R [illegible]
- n [illegible]
- S_{D_1} [illegible]
- N [illegible]

Formulae:	Definition:
$n = \left(\frac{S_{D_1} \cdot U_R \cdot N}{A}\right)^2$	Sample size using difference method-with replacement-unstratified
$\beta = \frac{1 - CR}{(1-IC)(1-SP)}$	Planned β level
$A = \frac{M}{\left(1 + \frac{U_\beta}{U_R}\right)} =$	Planned precision
$A = \left(\frac{S_{D_1} \cdot U_R \cdot N}{\sqrt{n}}\right)$	Achieved precision at a given level of reliability
$U_\beta = \frac{(M - A)U_R}{A}$	β coefficient used to determine β risk at a given reliability level and achieved precision

TABLE A-1

RESPONSES BY CONTROL GROUP SUBJECTS TO SELF-CLASSIFICATION

EXERCISE AND TO DECISION CASES

Subject Number	Self-Classification of Extent of Involvement in Sample Evaluation	Decision Case Responses							
		Case A (IC = 70, n = 196)		Case B (IC = 70, n = 100)		Case C (IC = .00, n = 196)		Case D IC = .00, n = 100)	
		Approach Employed	Basis for Audit Action	Approach Employed	Basis for Audit Action	Approach Employed	Basis for Audit Action	Approach Employed	Basis for Audit Action
1	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
2	Moderate	Classical	Statistical	Classical	Statistical	Non-classical	Statistical	Non-classical	Statistical
3	Extensive	Classical	Statistical	Classical	Statistical	Classical	Statistical	Classical	Statistical
4	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
5	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
6	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
7	Extensive	Non-classical	Statistical	Non-classical	Statistical	Classical	Judgmental	Classical	Judgmental
8	Extensive	Classical	Statistical	Classical	Statistical	Classical	Statistical	Classical	Statistical
9	Extensive	Classical	Statistical	Classical	Statistical	Classical	Statistical	Classical	Statistical
10	Extensive	Non-classical	Judgmental	Non-classical	Judgmental	Non-classical	Statistical	Non-classical	Statistical

TABLE A-2

RESPONSES BY EXPERIMENTAL GROUP SUBJECTS TO SELF-CLASSIFICATION EXERCISE AND TO DECISION CASES

Subject Number	Self-Classification of Extent of Involvement in Sample Evaluation	Decision Case Responses: Case A (IC = .70, n = 196)		Case B (IC = .70, n = 100)		Case C (IC = .00, n = 196)		Case D (IC = .00, n = 100)	
		Approach Employed	Basis for Audit Action	Approach Employed	Basis for Audit Action	Approach Employed	Basis for Audit Action	Approach Employed	Basis for Audit Action
1	Extensive	Classical	Judgmental	Classical	Judgmental	Non-classical	Statistical	Non-classical	Statistical
2	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
3	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
4	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
5	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
6	Moderate	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
7	Extensive	Non-classical	Statistical	Non-classical	Statistical	Non-classical	Statistical	Non-classical	Statistical
8	Extensive	Non-classical	Judgmental	Classical	Statistical	Classical	Statistical	Classical	Statistical
9	Extensive	Classical	Judgmental	Classical	Statistical	Non-classical	Statistical	Non-classical	Statistical
10	Extensive	Nonclassical	Statistical	Nonclassical	Statistical	Non-classical	Statistical	Non-classical	Statistical
11	Extensive	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental	Classical	Judgmental
12	Extensive	Non-classical	Statistical	Non-classical	Statistical	Non-classical	Statistical	Non-classical	Statistical
13	Extensive	Classical	Statistical	Classical	Statistical	Classical	Statistical	Classical	Statistical
14	Extensive	Classical	Statistical	Classical	Statistical	Classical	Statistical	Classical	Statistical

Appendix B

Debriefing Questions

1. Describe the meaning of the word "substantive" as it is used in this study.

2. Describe the meaning of the term "decision interval" as it is used in this study.

3. List at least one task that would be included in each of the three phases of a statistical sampling application:

    execution

    planning

    evaluation

4. Were any of the instructions unclear? If so, which one(s)?

5. Were there any terms about which you were unsure of their meaning? If so, list the term and describe what you thought it meant.

6. List any suggestions that you have that you feel might improve this questionnaire.