*Jurnal Ekonomi dan Bisnis Indonesia*
*Vol. 15, No. 4, 2000, 471 - 485*

# THE ADEQUACY OF CPAs' UNDERSTANDING OF THE RELATIVE SERIOUSNESS OF ALPHA AND BETA RISKS IN STATISTICAL AUDIT SAMPLING

**Bambang Sudibyo**
Universitas Gadjah Mada

***ABSTRACT***

*The issue raised in this article is the adequacy of US CPAs' understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. The objective of this study is to seek empirical evidence on the issue. Empirical data was collected in 1984 using the method of mail survey, with members of the American Institute of Certified Public Accountants (AICPA) serving as the target population. The author concluded that CPAs, in general, did not have an adequate understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. The author also concluded that the understanding of the issue differed across various groups of CPAs. Big eight CPAs, in general, had an adequate understanding of the issue, while CPAs of the other groups, in general, did not. Comparisons between the scores of these groups Indicate that 1) big-eight CPAs scored higher than non-big-eight CPAs; 2) academic CPAs scored higher than non-academic CPAs; and 3) contrary to expectation, practitioners in general did not score significantly higher than non-practitioners.*

## THE PHENOMENON OF INCREASING AUDIT RISK

One of the most interesting phenomena in the profession of public accounting in the US in the last three or four decades has been the increasing audit risk. Public accountants in the US today are living in a litigious environment. Any audit engagement brings with it a risk that the accountant may be sued, which may result in a substantial damage being awarded against him. According to Jaenicke (1977), until 1967 suits against accountants by third parties had been almost universally unsuccessful: but three cases in the late 1960s - Fisher v. Klets, Escott v. Barchris, and United States v. Simon - triggered forces that dramatically changed the situation. Since then the number of court cases brought by third parties against auditors has substantially increased reaching a peak in 1975 and early 1976. Even though the number of court cases involving accountants currently may not be as large as that of the mid 1970s, the extent of litigation is still high, particularly if it is compared to the almost zero base of the mid 1960s. This adverse situation has substantially increased insurance costs (Jaenicke, 1977).

## THE RELATIVE SERIOUSNESS OF ALPHA AND BETA RISK IN AUDIT SAMPLING

The increasing popularity of statistical sampling in auditing and the nature of today's audit suggest that the risk associated with statistical sampling should be carefully investigated. There are two statistically determinable risks in sampling: 1) alpha risk, the risk of committing a type I error, which is that of rejecting a true null hypothesis; and 2) beta risk, the risk of committing a type II error, which is that of accepting as true a false null hypothesis.

In auditing, there are two areas that usually involve sampling: compliance and substantive tests. In a compliance test, an auditor is interested in determining the level of reliance that s/he will put on a certain internal accounting control procedure. In a substantive test, the auditor is concerned with deciding whether a book value is fairly presented, or in estimating the dollar value of a certain accounting parameter of interest.

Alpha risk in a compliance test is the risk associated with erroneously concluding that "the sample does not support the auditor's planned degree of reliance on the control when the true compliance rate supports such reliance"[1] (AICPA, 1961). If a type I error is committed, the auditor will underrate the system and lower the degree of reliance below the planned degree, and increase substantive tests by unnecessarily extending them. As a consequence extra cost will be incurred for the extended audit procedures, which must be borne by either the auditor or the client or a combination.

In a substantive test, alpha risk is the risk associated with committing the error of concluding that a book value is not fairly presented, when in fact it is.[2] If this happens, it is very likely that the client will resist the adjustment proposed by the auditor. This resistance, together with the need to correct related account details, will likely convince the auditor to enlarge the sample. The larger sample will cause the auditor either to alter or to reaffirm the original conclusion. In the case of the auditor's altering the original conclusion, i.e. from rejecting to accepting the book value as fairly presented, the adverse consequences of the apparent initial type I error would be the incremental audit cost related to the larger sample size. If, on the other hand, the auditor is convinced by additional information that the original conclusion was correct, the alpha risk still exists and at the same level, but the sampling error has been reduced. In such a situation the adverse consequences associated with the initial type I error would again be the incremental audit cost of the enlarged sample. Thus, regardless of the initial level of alpha risk, a rejection of a fairly presented book value would likely result in the enlargement of sample size until either persuasive evidence is obtained suggesting that a type I error has not been committed, or the sampling error is reduced to a level acceptable to both auditor and client.

Beta risk in a compliance test is the risk associated with erroneously concluding that "the sample supports the auditor's planned degree of reliance on the control when the true compliance rate does not justify such reliance"[3] (AICPA, 1981). This error causes the auditor to overrate the system and set the extent of the related substantive tests too low. In this situation the auditor runs the risk of expressing an opinion on financial statements without having obtained sufficient evidence. In terms of Generally Accepted Auditing Standards (GAAS), the audit is substandard, because of failure to meet the standard of fieldwork #3.[4] If the type II error were really

[1] ***Statement on Auditing Standards #39 calls this risk the "risk of underreliance." The above sentence implies the following null hypothesis: the true compliance rate is equal to or greater than the rate used as the base (or determining the planned degree of reliance. All discussions in this paper that relate to statistical hypothesis testing are presented in terms of the positive approach to hypothesis testing.***

[2] ***Statement on Auditing Standard #39 calls this risk "the risk of incorrect rejection." The above sentence implies the following null hypothesis: the book value is fairly presented, or the total error in the book is not material***

[3] ***Statement on Auditing Standards #39 calls this risk the "risk of over reliance." The above sentence implies the same null hypothesis indicated in footnote #1.***

[4] ***This will happen if the insufficient evidence obtained from the substantive tests is not compensated for by other evidence obtained from other audit procedures.***

committed, then the audit report would contain misleading information.

In a substantive test, beta risk is the risk of concluding that a book value is fairly presented, when in fact it is not.[5] The related financial statements and audit report will then be misleading. The adverse consequences of this risk are those associated with the reliance that third parties place on the misleading financial statements and audit report. These consequences in recent years have led to an Increasing number of lawsuits with the attendant losses in money and reputation.

As previously noted, the adverse consequences of committing a type I error in compliance or substantive tests are the costs associated with the extended audit procedures. In the United States, where most audits are on an hourly fee basis, these costs are normally paid by clients, for they are incurred during the audit. Thus, to the auditor, at least in the short-run, the financial consequences associated with alpha risk are practically none.[6] In the long-run, however, frequent occurrences of type I error would serve as an indication of inefficient audits. Clients would not tolerate the continually excessive audit fees, and sooner or later they would substitute more efficient auditors for the inefficient ones.

In the case of type II error, the threats of financial loss come from third parties after the audit report has been signed and issued by the auditor. This loss can be great, and it must be paid by the auditor or the insurance carrier. Associated with this loss is another loss, the burden of which no one else can assume. This is the loss in reputation, which even in the short-run can bring some serious financial consequences.

Thus, either type I or type II error can lead to a serious financial loss. However, the litigious environment of today's auditors suggests that the consequences of making a type II error are of more immediate concern to the auditors than those of making a type I error. The Statement on Auditing Standard #39 (AICPA, 1981) implies that alpha risk relates to the efficiency of an audit, while beta risk relates to the effectiveness of an audit.[7] An inefficient audit may still be effective and therefore meet the objectives of attest function, though wastefully. On the other hand, an ineffective audit, though it may be efficient, fails to achieve the objectives of attest function, and therefore it represents a waste of resources. Thus, beta risk deserves more attention than does alpha risk because it directly touches the basic reason for the creation of a public accounting profession by a society. Beta risk is a professional risk, while alpha risk, the risk of being wasteful, is a general business risk present in any business undertaking

## ALPHA AND BETA RISKS IN STATISTICAL TEXTBOOKS

In statistical texts typically used by non-statistics majors of colleges and universities, alpha and beta risks are usually discussed in the chapters or sections dealing with hypothesis testing. The contents of the texts in general are still dominated by classical statistics, and the chapters or sections on Bayesian statistics typically represent a minor part of individual texts. Within the classical tradition there are two competing schools of thought that interpret the results of a test

***

[5] ***Statement on Auditing Standards #39 call this risk the "risk of incorrect acceptance." The above sentence implies the same null hypothesis indicated in footnote #2.***

[6] ***Some audits are on a fixed fee basis; In this case the costs associated with the extended audit procedures have to be paid by the auditor.***

[7] ***This sentence should be understood in the light of the preceding six paragraphs; otherwise, it can easily be confused with the notion of sampling efficiency of a plan with a single selection. In the latter case, beta risk is directly related to sample size, and therefore sample efficiency, because beta is a function of sample size; on the other hand, sample size, and therefore sample efficiency, cannot affect the magnitude of alpha, because alpha is set judgmentally ex-ante the test.***

differently.[8] The first school, hereinafter called the philosophy of science school, interprets the results of statistical hypothesis testing strictly from the point of view of the Popperian philosophy of science. This philosophy maintains that a hypothesis or a theory can never be proven true; on the other hand, it is possible to "prove" it wrong: a good theory or hypothesis is one that well-designed empirical tests done repeatedly fail to refute (Popper, 1962). Accordingly, inference from results of a statistical significance test can only be made when the test leads to a rejection of the underlying null hypothesis. In this case, it is philosophically valid to conclude that the null hypothesis is wrong. When the test fails to reject it, it is philosophically invalid to conclude that the null hypothesis is right or acceptable. In this case the statistician reserves his or her judgment.

The second school, hereinafter called the mathematical school bases the interpretation of the results of a hypothesis testing purely on the mathematical scheme of the test. Within this scheme acceptance and rejection are perfectly complementary, and both are valid inferences. Rejection of a null hypothesis means that it is false, and acceptance of it means that it is true.

This controversy has two implications. First, beta risk is not an important issue in the philosophy of science school because it does not make a substantive inference from a failure to reject a null hypothesis. Second, alpha and beta risks are of equal importance in the mathematical school because either rejection or acceptance of a null hypothesis is a valid conclusion about the results of a test.

Even though the influences of these two schools on the general presentation of hypothesis testing in statistical texts are mixed, the following observation supports a belief that the philosophy of science school is dominant. The following is more or less the typical presentation of the procedures for hypothesis testing in statistical texts:

1. Set the tolerable level of alpha risk, or alternatively, the desired confidence level.
2. Determine the acceptance and the critical or rejection regions under the relevant sample statistic distribution.
3. Compute the relevant test statistic.
4. Conclude the test as follows: accept or fail to reject the null hypothesis if the test statistic falls within the acceptance region, and reject the null hypothesis if the test statistic falls within the critical region.

Beta risk is not typically included in the procedures. Usually it is discussed in a separate discussion, where its magnitude is computed as a consequence of setting alpha at a certain level. Some old texts did not even show how to compute beta. Thus, statistical texts in general do not treat alpha and beta risks on an equal footing.

## ALPHA AND BETA RISKS IN AUDITING LITERATURE

The dominance of the philosophy of science school indicates the bias of statistical texts toward the need of scientific research. Popperian philosophy has been the generally accepted, fundamental philosophy underlying natural science research. In the social science branch, it continually earns more and more acceptance, and it is currently the methodological foundation of the mainstream social

---

***[8] Discussions about these two different interpretations can be found in the following articles, all in Bernhardt Lieberman (ed.) Contemporary Problems In Statistics (New York: Oxford University Press, 1971): 1) David Balkan, "The Test of Significance in Psychological Research" pp. 147-162; 2) Arnold Binder, "Further Consideration on Testing the Null Hypothesis and the Strategy and Tactics of Investigating Theoretical Models." pp. 135-142; 3) David A Grant, "Testing the Null Hypothesis and the Strategy and Tactics of Investigating Theoretical Models," pp. 127-134; 4) William W. Rozeboom, "The Fallacy of the Null Hypothesis Significance Test," pp. 116-126; 5) Warner Wilson et.al., "Much Ado about the Null Hypothesis," pp. 163-172.***

sciences (Taylor, 1977). However, scientific research is not the only area of application of statistical techniques. As previously noted, auditing is an area where statistical techniques are becoming more and more popular.

The difference between the context of scientific research and that of auditing suggests that the kinds of statistical techniques suitable for these two areas are different. Popperian philosophy, of course, is not on the minds of auditors conducting statistical tests on accounting populations. Their concern is that their tests conform to the Generally Accepted Auditing Standards (GAAS). The statistical tests are intended to help auditors make decisions about the fairness of accounting values. Their tests are then determinative in nature. In this determinative context, acceptance and rejection of a null hypothesis are of equal relevance, as is implied by the general standard of independent mental attitude. Therefore, audit tests would be better served if statistical hypothesis testing in auditing literature were developed within the framework of the mathematical school. Within this framework, as previously noted, acceptance and rejection of a null hypothesis, and therefore the related beta and alpha risk, all represent important aspects of the tests. Still, because of the potential severity of the consequences of a type II error, auditors should normally be more concerned with beta than with alpha risk.

However, it was not until the 80s that the awareness of the need to control beta risk appeared in auditing literature. Boatsman and Crooch (1975) observed that "no undergraduate texts contain a discussion of how the risk might be controlled". The following year, according to Pushkin (1980) "at least two auditing texts contain(ed) a discussion of how the risk might be controlled." Until the issuance of Statement on Auditing Standards #39, in 1981, there was no official pronouncement of the AICPA that addressed the issue of beta risk.

Auditing literature of the 80s and after, however, has begun to treat beta risk more appropriately. Most texts discuss the subject of statistical test within the framework of the classical tradition, and some of them appropriately balance the importance of considering alpha and beta risks.[9] Besides, there are many articles dealing with the issue of audit sampling risk.

## STATEMENT OF THE PROBLEM

In an audit test, the consequences of committing a type II error are generally more serious than those of committing a type I error. However, there are some reasons to suspect that the majority of CPAs may still not realize this. The following observations support this suspicion:

1. statistical literature typically available and read by accounting students historically has put more emphasis on alpha than beta risk;
2. not until the 80s did the auditing literature in general exhibit a proper perspective with respect to alpha and beta risk; and
3. by its very nature beta risk is more difficult to comprehend and to quantify.

If this suspicion is true, then most significance tests conducted by CPAs in their audits may not be as effective as desired. The increasing use of statistical sampling as an audit tool suggests that more empirical evidence on this matter is needed.

---

***[9] These include: 1) Roger H. Hermanson, et al., Auditing Theory and Practice, (Homewood Illinois: Richard D. Irwin, 1976); 2) Donald M. Roberts, Statistical Auditing (New York: AICPA, 1978); 3) Jack C. Robertson, Auditing, 2nd edition (Dallas: Business Publication Inc., 1977); 4) Walter G. Kell and Richard E. Ziegler, Modern Auditing (Boston: Jovanovich, Inc., 1981); 6) Dan M. Guy, op. cit.; 7) Alvin A. Arens and James K. Loebbecke, Applications of Statistical Sampling to Auditing (Englewood Cliffs: Prentice Hall, Inc., 1981); 8) Howard F. Stettler, Auditing Principles (Englewood Cliffs: Prentice Hall, Inc., 1982).***

## STATEMENT OF HYPOTHESES

Specifically, the study was focused on two things. First, it was designed to determine whether CPAs' responses, in general indicate an adequate understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. Second, it was also designed to determine whether the responses differ among various categories of CPAs. For the purposes of this research, three categories were identified, as follows:

1. CPAs who work in CPA firms and those who do not.
2. CPAs who work in big-eight CPA firms and those who work in non-big-eight CPA firms.
3. CPAs who work as academicians and those who do not.[10]

These concerns were formalized in the following hypotheses:

*Hypothesis #1*: The majority of CPAs do not have an adequate understanding of the relative seriousness of alpha and beta risks in statistical audit sampling.

*Hypothesis #2*: The level of understanding of CPAs who work in CPA firms is not different from that of those who do not.

*Hypothesis #3*: The level of understanding of CPAs who work for big-eight CPA firms is not different from that of those who work for non-big-eight CPA firms.

*Hypothesis #4*: The level of understanding of CPAs who work as academicians is not different from that of those who do not.

Hypotheses two, three and four are intended to identify the more specific sub-populations, in which the problem, if it-exists, is particularly serious.

[10] ***For the purposes of this study, CPAs who work as academicians were defined as those who were on accounting faculties at colleges and universities.***

## OVERVIEW OF METHODOLOGY

The target population were members of the AICPA. Eight hundred sixty-five CPAs were randomly sampled from this population to participate in the study. This number is a sum of four stratum samples: big-eight practitioners, non-big-eight practitioners, non-practitioners, and academic CPAs.

Survey data were collected using the method of mail survey. From the 865 questionnaires sent to potential respondents, 97 of them were returned by the Postal Service because, for various reasons, they could not be forwarded to the addressees. The number of potential respondents who received questionnaires was, then, 768. From this number of potential respondents, 402 of them returned the questionnaires. Therefore the response rate was 52.3%. These returns produced 377 usable questionnaires, which number represents the effective sample size for this study.

The focal variable being measured in this study was the adequacy of CPAs' understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. This variable was measured by using an eight-item questionnaire. Besides this variable, 24 supplemental variables were also measured, in the hope that they would allow this author to explain the results of the survey which pertained to the focal variable. These supplemental variables were measured by using one-item questions. The questionnaire uses a closed, multiple-choice response form.

Before the questionnaire was sent to potential respondents, it was tested four times to see if it was workable and to identify the necessary revisions to be made. Three of the four tests were "in-house" tests. In the other test the questionnaire was given to CPAs who practiced in Lexington, Kentucky. After the mail survey was completed, the questionnaire was checked with respect to its reliability. The check indicated that the questionnaire has a reliability coefficient of at least 0.85. The data produced by the questionnaire were checked to see if non-response bias was present in them. Two methods of checking were employed, and both of them indicated that non-response bias was not present in the data. The data were also tested against the possibility that a significant number of respondents answered the questionnaire by random guessing. The test results indicated that it is very unlikely that a significant number of respondents answered the questionnaire by random guessing. The data were then statistically analyzed.

## STATISTICAL ANALYSES

To test research hypothesis #1, a criterion should first be established regarding what is meant by adequately understanding the relative seriousness of alpha and beta risks in statistical audit sampling. A respondent is considered to have an adequate understanding of the focal issue if s/he scored at least 6. If s/he scored less than 6, s/he is not considered to adequately understand the issue. The score of 6 was used as the cut-off point for the following reasons. First, for a respondent to be considered as adequately understanding the issue, s/he must provide more correct answers than incorrect answers. Thus, s/he must score at least 5. Second, however, a score of 5 should not be used as a cut-off point, because the probability of getting at least 5 correct responses by random guessing is 0.3633. This probability is considered too high for the purposes of this study. The probability of getting at least 6 correct responses by random guessing is only 0.1445. This probability is considered sufficiently low, and therefore the score of six was chosen as the minimum score that should be achieved by a respondent in order for him or her to be considered as adequately understanding the issue.

The majority of CPAs will be considered as adequately understanding the issue if the majority scored 6 or more. In this case, the proportion of those who scored 6 or more should be significantly greater than 0.50. Thus, the statistical test needed to test the above research hypothesis is a test of proportion.

Research hypothesis #2 was tested using a test of homogeneity. The test determined whether the random samples of scores of the two subpopulations were drawn from the same population or from different populations. If they were drawn from the same population, the distributions of scores of these two subpopulations should be very similar, and so their levels or understanding of the focal issue are not significantly different. If, on the other hand, the two samples were drawn from different populations, the distributions of scores of the two subpopulations should follow different patterns. In this case, it can be inferred that the two subpopulations have different levels of understanding of the focal issue. The test of homogeneity was made using a chi-square test. Hypotheses 3 and 4 were tested using tests of homogeneity similar to the one used for testing research hypothesis no. 2

## OVERVIEW OF THE SURVEY DATA

**Table 1** shows simple, descriptive statistics derived from the survey sample and other relevant information.

**Table 2** shows the distribution of scores- over all subjects- across strata. The stratum percentages of respondents who scored 6 or more are (ordered from the highest to the lowest):

| | |
|---|---|
| big-eight CPAs | 61.41% |
| academic CPAs | 53.13% |
| non-practitioners | 41.18% |
| non-big-eight practitioners | 37.77% |
| overall | 45.60% |

The rank ordering of these percentages intuitively indicates that the big-eight CPAs provided the best responses; the academic CPAs the second best responses; and the non-practitioner and non-big-eight practitioner CPAs the worst responses.

**Table 1.** Sample Statistics and Other Relevant Data

| | Big-eight CPAs | Non-big-eight practitioners | Non-Practitioners | Academic CPAs | Overall |
|---|---|---|---|---|---|
| Population Size (000) | 20 | 85 | 90 | 5 | 200 |
| Effective Sample Size | 55 | 98 | 158 | 66 | 377 |
| Sample Mean Score | 5.053 | 4.144 | 4.072 | 5.125 | 4.227 |
| Sample Standard Deviation | 2.482 | 2.683 | 2.817 | 2.705 | 2.735 |
| Standard Error | 0.324 | 0.283 | 0.228 | 0.338 | 0.143 |
| Sample Coefficient of Variation | 49.11 | 64.76 | 69.18 | 52.78 | 64.71 |
| Sample Median Score | 6 | 5 | 5 | 6 | 5 |
| Sample Modal Score | 6 | 5 | 7 | 8 | 6 |

**Table 2.** Distribution of Scores Across Strata (top entries indicate frequencies and lower entries stratum percentages)

| **Strata** | **Scores** | | | | | |
|---|---|---|---|---|---|---|
| | 0-2 | 3-5 | 6 | 7 | 8 | Total |
| Big-eight CPAs | 9<br>15.79 | 13<br>22.80 | 19<br>3.33 | 8<br>14.04 | 8<br>14.04 | 57<br>100.0 |
| Non-big-eight practitioners CPAs | 29<br>32.22 | 27<br>30.00 | 12<br>13.33 | 13<br>14.44 | 9<br>10.00 | 90<br>100.0 |
| Non-practitioners | 51<br>33.33 | 39<br>25.49 | 20<br>13.07 | 32<br>20.92 | 11<br>7.19 | 153<br>100.0 |
| Academic CPAs | 12<br>18.75 | 18<br>28.13 | 11<br>17.19 | 4<br>6.25 | 19<br>29.69 | 64<br>100.0 |
| Overall | 101<br>27.75 | 97<br>26.64 | 62<br>17.03 | 57<br>15.66 | 47<br>12.91 | 364<br>100.0 |

## THE ADEQUACY OF CPAs' UNDER-STANDING OF THE RELATIVE SERIOUSNESS OF ALPHA AND BETA RISKS IN STATISTICAL AUDIT SAMPLING

As indicated previously hypothesis #1 was tested by testing whether the proportion of CPAs who scored 6 or more is greater than 0.50. Table 2 indicates that the sample proportion of those who scored 6 or more is 45.60%. The test failed to reject the null hypothesis that the population proportion is equal to or less than 0.50. Thus, CPAs in general did not appear to have an adequate understanding of the relative seriousness of alpha and beta risks in statistical audit sampling.

The question is now whether the above result is also true in the case of each of the strata. Six separate tests similar to the one above were made to answer these questions. The results of the tests are documented in **Table 3**.

**Table 3.** Results of Tests to See if Subpopulation Proportions of CPAs Who Scored 6 or More Are Greater than 0.50

| Subpopulation | Sample proportion | Z Value | Significance |
|---|---|---|---|
| Big-eight CPAs | 61.41% | 1.724 | Significant |
| Non-big-eight practitioners | 37.71% | -2.321 | Not significant |
| Non-practitioners (excluding academic CPAs) | 41.18% | -2.183 | Not significant |
| Academic CPAs | 53.13% | 0.501 | Not significant |
| Practitioners CPAs | 46.94% | -0.743 | Not significant |
| Non-practitioners (including academic CPAs) | 44.70% | -1.563 | Not significant |

**Table 3** shows that the stratum of big-eight CPAs had a proportion significantly greater than 0.50, while the rest of the CPAs did not. Thus, one can appropriately say that the survey data supported a proportion that the majority of big-eight CPAs appeared to have an adequate understanding of the relative seriousness of alpha and beta risks in statistical audit sampling, while the majority of CPAs of the other strata did not.[11] The table also indicates that neither practitioners nor non-practitioners (including academicians), as subpopulations of CPAs, had an adequate understanding of the issue. Apparently, the relatively high proportion of big-eight practitioners who scored 6 or more is not sufficient to compensate for the relatively low proportion of non-big-eight practitioners who scored in that range.

Note that the majority of academic CPAs did not appear to adequately understand the issue, an unexpected result. Being educators who play significant roles in the required training for CPA candidates, they are expected to adequately understand the issue. A possible explanation for this particular result is that a significant portion of academic CPAs are those who do not teach or have never taught auditing. Because their jobs do not require that they always be up-to-date on developments of statistical audit sampling literature, it is not surprising that they did not score well. If the academic CPAs included in the sample were limited to those who teach or ever taught auditing, maybe the result would be different.

## RELATIONSHIP BETWEEN STRATA AND INDIVIDUAL SCORES

Research hypotheses 2, 3 and 4 involve comparisons between subpopulations of CPAs on their understanding of the focal issue. Before these specific comparisons were made, a test was made to see if a dependency relationship exists between "strata" and "score." If the dependency relationship does not exist, then there is no point to making comparisons between strata. If it does, then it can be said that "strata" has some bearing on the pattern of individual scores.[12] The test is a test of Independence using a chi-square distribution. It determines whether the patterns of distributions of scores over all strata are different from the hypothesized patterns under the assumption of independence. The test produced a square-square value of 46.937 which

[11] ***In the questionnaire, the questions on the relative seriousness of alpha and beta risks in statistical audit sampling are preceded by instructions which provide definitions of alpha and beta risks in audit sampling. These definitions, to some degree, provide clues of the correct answers to the questions, and so respondents' scores may have been inflated. Thus, the conclusion that the majority of big-eight CPAs had an adequate understanding of the relative seriousness of alpha and beta risks in statistical audit sampling may not be correct.***

[12] ***The dependency relationship cannot go in the other direction. That is, it is inconceivable that "score" has some bearing on "strata".***

is greater than the critical value for 0.05 significance level (critical value = 25; p = 0.0001). Thus, there is a dependency relationship between "strata" and "score." In other words, "strata" has some Influence on the pattern of individual scores.

## COMPARISON BETWEEN PRACTITIONERS AND NON-PRACTITIONERS

Research hypothesis #2 was tested by using a chi-square test applicable for testing the homogeneity of two populations. The test results indicated that the sample data failed to reject the null hypothesis that the pattern of distribution of scores of the subpopulation of practitioners is similar to that of the subpopulation of non-practitioners (p = 0.7652). Thus, practitioners and non-practitioners did not appear to have significantly different levels of understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. Research hypothesis no. 2 was, therefore, supported by the research data.

The result is contrary to this author's expectation. Being more directly involved with statistical audit sampling, the practitioner group was expected to perform better than the rest of the CPAs. Perhaps this unexpected result was because many non-practitioners were once practitioners, so that in terms of experience in audit practice, the populations of practitioners and non-practitioners are not really mutually exclusive populations. In this study, 84.68% of non-practitioner respondents (including academic CPAs) reported that they were once practitioners. Therefore, the more appropriate comparison is between those who have and those who have not worked as practitioners. Another test of homogeneity was made for this purpose. However, this test could only be made for the strata of academic CPAs and non-practitioners because all big-eight and non-big-eight practitioners have, of course, worked as practitioners. In the former two strata, the test failed to reject the null hypothesis that the pattern of distribution of scores of those who have ever worked as practitioners is similar to that of those who have never worked as practitioners (p values are 0.242 for the academic CPAs and 0.973 for the non-practitioners). Thus, apparently experience in working as practitioners did not have a significant bearing on individual scores, for the strata of academic CPAs and non-practitioners.

An alternate explanation is the fact that most of the practitioners are non-big-eight practitioners. Among the four strata investigated in this study, the non-big-eight practitioner group has the lowest proportion of those who scored 6 or more, and has the second highest proportion of those who scored 2 or less.

One may be interested in seeing whether an exclusion of academic CPAs from the non-practitioner group makes the result different. Academic CPAs, for the purposes of this study, could not be regarded as typical non-practitioners, because they only make up 5.26% of the non-practitioner group and their sample mean score is higher than that of the rest of non-practitioners. For this reason, another test of homogeneity was made comparing the distributions of scores of practitioners and non-practitioners (excluding academic CPAs). The test failed to reject the null hypothesis that the pattern of distribution of scores of the subpopulation of practitioners is similar to that of the subpopulation of non-practitioners (excluding academic CPAs) (p = 0.1708). Thus, whether or not the academic CPAs were included in the non-practitioner group, the scores of practitioners did not appear to be significantly different from that of the non-practitioners.

## COMPARISON BETWEEN BIG-EIGHT AND NON-BIG-EIGHT PRACTITIONERS

Hypothesis #3 was tested using a chi-square test applicable for testing the homogeneity of two populations. The result of the test indicated that the sample data supported a rejection of the null hypothesis that the pattern of distribution of scores of the big-eight practitioners is similar to that of the non-big-eight practitioners (p = 0.025). The understanding levels of the two strata appear to be significantly different. Research hypothesis no. 3 was, then, not supported by the research data. Table 4 summarizes the two distributions in terms of percentages.

**Table 4.** Percentage Distributions of Scores of the Big-eight and Non-big-eight Practitioners

| Strata | Scores | | | Total |
|---|---|---|---|---|
| | 0-2 | 3-5 | 6-8 | |
| Big-eight practitioners | 15,78% | 22.81% | 61.41% | 100.00% |
| Non-big-eight practitioners | 32.22% | 30.01% | 37.77% | 100.00% |

The stratum of big-eight practitioners has proportionally more CPAs who scored 6 or more than does the stratum of non-big-eight practitioners. On the other hand, the stratum of big-eight practitioners has proportionally fewer CPAs who scored 2 or less than does the stratum of non-big-eight practitioners. Evidently, the big-eight practitioners understood the focal issue better than did the non-big-eight practitioners. This result is, of course, not surprising at all. The big-eight practitioners, being associated with firms that have more resources, are expected to score higher, on the average, than practitioners who are associated with firms with fewer resources.

The stratum of big-eight practitioners has proportionally more CPAs who scored 6 or more than does the stratum of non-big-eight practitioners. On the other hand, the stratum of big-eight practitioners has proportionally fewer CPAs who scored 2 of less than does the stratum of non-big-eight practitioners. Evidently, the big-eight practitioners understood the focal issue better than did the non-big-eight practitioners. This result is, of course, not surprising at all. The big-eight practitioners, being associated with firms that have more resources, are expected to score higher, on the average, than practitioners who are associated with firms with fewer, resources.

## COMPARISON BETWEEN ACADEMIC AND NON-ACADEMIC CPAs

Hypothesis was tested by using a chi-square test applicable for determining the homogeneity of two populations. The test rejected the null hypothesis that the pattern of distribution of scores of the academic CPAs is similar to that of the rest of the CPAs (p = 0.0016). This means that research hypothesis no. 4 was not supported by the research data. Table 5 summarizes the two sample distributions of scores in terms of percentages.

The subpopulation of academic CPAs has proportionally more CPAs who scored 6 or more than does the subpopulation of non-academic CPAs. On the other hand, the subpopulation of academic CPAs has proportionally fewer CPAs who scored 2 or less than does the subpopulation of non-academic CPAs. Clearly, the academic CPAs, in general, scored higher than did the rest of the CPAs. This result was expected because academic CPAs probably are more familiar with, or up to date with respect to, the development of statistical audit sampling literature.

**Table 5.** Percentage Distributions of Scores of the Academic and Non-academic CPAs

| Strata | Scores | | | Total |
|---|---|---|---|---|
| | 0-2 | 3-5 | 6-8 | |
| Academic CPAs | 18.76% | 28.13% | 53.13% | 100.00% |
| Non-academic CPAs | 29.67% | 26.35% | 44.00% | 100.00% |

Even though the academic CPAs scored better than the rest of the CPAs, as previously indicated academic CPAs generally did not adequately understand the focal issue. The big-eight-CPAs, on the other hand, did. Are the scores of this specific class of non-academic CPAs higher than those of academic CPAs? A test of homogeneity using a chi-square distribution was made to answer this question. The test failed to reject the null hypothesis that the patterns of distribution of scores of these two strata are similar (p = 0.6546). Evidently, the big-eight CPAs did not score significantly better than academic CPAs.

## OTHER FINDINGS

In this section, the results of the investigation of variables other than "strata" variables that may have some bearings on the patterns of individual scores were discussed. Twenty-four variables were investigated, the measurements of which were provided by parts I and II of the questionnaire.

The author found that most of these variables do not have significant relationships with "score" at 0.05 level. Only four variables were found to have significant relationships. The first variable is "number of actual clock hours of on-the-job training in statistical sampling." The values of this variable were classified into two ordinal values: 15 hours or less and 16 hours or more. For the stratum of academic CPAs, the author found that the coefficient of Somer's D C/R for the association between this variable and the variable "score" is 0.420. This coefficient is significantly greater than zero. One may then say that it is 42% more probable that more hours of on-the-job training would lead to higher scores than otherwise. For the other strata none of the coefficients of Somer's D C/R are significantly different from zero.

Thus, only in the stratum of academic CPAs, did on-the-job training in statistical sampling seem to be effective in advancing CPAs' understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. This can mean three things. First, maybe the quality of on-the-job training programs taken by academic CPAs is relatively better than that of programs taken by other CPAs. Second, holding the quality of training constant, maybe academic CPAs learned better than the rest of CPAs because they had better educational backgrounds. All of the academic CPAs included in the sample have master and/or doctoral degrees, while 68.6% of non-academic CPAs have only bachelor or no degrees. The relatively better educational background or academic CPAs could make their ability to comprehend materials given in training better than the ability of the other CPAs. Finally, it is also possible that academic CPAs both learned better and received better training.

The second variable that was found to have a significant relationship with "score" is "number of actual clock hours of continuing professional education program (cpe) in statistical sampling." The values of this variable were categorized into two ordinal values: 15 hours or less and 16 hours or more. For the stratum of academic CPAs the coefficient of Somer's D C/R is 0.368. Even though small, this coefficient is significantly greater than zero. The coefficient indicates that it is 36.8% more probable that more hours of cpe program would lead to higher scores than otherwise.

For the other strata none of the coefficients are significantly different from zero.

Clearly, it is only in the stratum of academic CPAs that cpe programs in statistical sampling seemed to be effective in promoting CPAs understanding of the relative seriousness of alpha and beta risks in statistical audit sampling. Just as in the case of "number of actual clock hours of on-the-job training in statistical sampling," this observation can mean three things. First, maybe the cpe programs attended by academic CPAs were relatively better than those attended by non-academic CPAs. Second, may be the quality of cpe programs was relatively the same across strata, but academic CPAs learned better than non-academic CPAs. Finally, it is also possible that academic CPAs both attended better cpe programs and learned better in the programs.

Note that both on-the-job training and cpe programs in statistical sampling attended by practitioners, either big-eight or non-big-eight, did not appear to be effective in advancing practitioners' understanding of the focal issue. This finding should, of course, be of concern to CPA firms. The understanding of the focal issue is an indication of the appropriateness of applications of statistical sampling in auditing. Therefore, adequate understanding of the issue should be an objective of on-the-job training and cpe programs in statistical sampling held by CPA firms. The findings above indicate that these programs have failed to achieve this objective.

The third variable that has a significant relationship with "score" is "number of credit hours of undergraduate auditing." The values of this variable were classified into two ordinal values: 3 credit hours or less and 4 credit hours or more. For the stratum of academic CPAs the coefficient of Somer's D C/R is 0.308, which is significantly greater than zero. This means that it is 30.8% more probable that more credit hours of undergraduate auditing would cause higher scores than fewer credit hours. This is not true for the other strata, because their coefficients of Somer's D C/R are all not significantly different from zero.

Apparently, "number of credit hours of undergraduate auditing" significantly affects "score" only in the stratum of academic CPAs, and not in the other strata. One possible explanation for this is that academic CPAs may have attended undergraduate auditing courses that were, in general, relatively better than those attended by other CPAs. This could happen simply because academic CPAs received their undergraduate education from schools that are qualitatively better than those attended by other CPAs.

Another possible reason is that perhaps academic CPAs are, in general, relatively more intelligent than non-academic CPAs. Being more intelligent, they were more able to take benefits from their undergraduate auditing training in promoting their understanding of the focal issue.

The fourth variable that has a significant relationship with "score" to "accounting firm-category in which a CPA ever worked." This variable has four categorical values: national firm, regional firm, local firm and never worked in an accounting firm. A national firm is defined as one whose offices and/or branches reside in more than 25 states. A regional firm is one whose offices and/or branches reside in more than one state but in fewer than or equal to 25 states. A local firm is one whose offices and/or branches reside in only one state. If a respondent had ever worked in a national firm, irrespective of whether s/he also had ever worked in other type(s) of firm(s), s/he was included in the national firm category. If s/he had ever worked in a regional firm but never in a national firm, irrespective of whether s/he had ever worked in a local firm, s/he was included in the regional firm category. Consequently, the local firm category is exclusively reserved for those who have only worked in a focal firm, but never in national and/ or regional firms. The categorization was made so, because the author is interested in seeing whether the largest size of firm a respondent ever worked in has a significant effect on individual scores.

The author found that for the stratum of academic CPAs, the results of the chi-square test indicate that there is a dependency relationship between this focal variable and "score" (p=0.027). Evidently, this focal variable has some bearing on the pattern of academic CPAs' scores. The uncertainty coefficient C/R for this stratum is 0.111, which, though small, is significantly greater than zero. This means that there is an 11.1% reduction in uncertainty in predicting an academic CPA's score that results from knowing his or her accounting firm category in which s/he ever worked. For the other strata, no significant relationship between the focal variable and the variable "score" was found.

## IMPLICATIONS

As previously noted, it was found that in general CPAs do not adequately understand the focal issue. This finding should be of concern to parties like AICPA, CPA firms, individual CPAs colleges and universities, and accounting faculties. The finding suggests that some improvement still has to be made. Such improvement is very important because the findings indicate that the effectiveness of statistical audit samplings that have been or will be done by CPAs is questionable.

To the AICPA the finding should suggest that the institute's pronouncements with respect to statistical audit sampling have to be reviewed. A consideration must be made as to whether the relative seriousness of alpha and beta risks in statistical audit sampling has been appropriately emphasized.

To individual CPAs and CPA firms, especially non-big-eight CPAs and firms, the finding reveals that their technical proficiency in statistical audit sampling should be improved. Special attention should be given to on-the-job training and continuing professional education (cpe) programs in statistical sampling. The author found for the strata of big-eight and non-big-eight practitioners, "number of hours of on-the-job training in statistical sampling" and "number of hours of cpe program, in statistical sampling" did not affect "score." This indicates that, in general, on-the-job training and cpe programs in statistical sampling do not place a proper emphasis on the relative seriousness of alpha and beta risks in statistical audit sampling.

As previously noted, the author found that CPAs in general do not adequately understand the relative seriousness of alpha and beta risks in statistical audit sampling. To colleges, universities, and accounting faculties this finding indicates that accounting curricula have to be reviewed. Special attention should be given to auditing and statistics syllabi. The author found that "number of credit hours of auditing courses taken by a CPA when s/he was a student" and the "extent of auditing-courses coverage on statistical sampling" were not significant factors affecting Individual scores.[13] Clearly, auditing courses in general do not adequately cover the relative seriousness of alpha and beta risks in statistical audit sampling. The study also found that the "number of credit hours of statistics taken by a CPA when s/he was a student" was not a significant factor affecting individual scores. Apparently, statistics courses taken by accounting students in general do not appropriately discuss sampling risk.

This study was conducted in the US environment in 1984-1985. It is still, in this writer's judgment, worthwhile to replicate this study in the US now to see how understanding of audit sampling risk have been progressing there after time elapsed for 15 years. Assuming IAI's members competence in statistical audit sampling in the year 2000 is no better than that of CPA's of the US in the 80s,

***
[13] *Except for "number of credit hours of undergraduate auditing" for the stratum of academic CPAs.*

it is also then worthwhile to replicate this study in Indonesian setting now.

## REFERENCES

AICPA, "Statement on Auditing Standards No. 39: Audit Sampling", 1981.

Arens, Alvin A. and Loebbecke, James K., *Aplications of Statistical Sampling to Auditing*, (Englewood Cliffs: Prentice Hall, Inc., 1981).

Baily Jr., Andrew D., *Statistical Auditing*, (New York: Harcourt Brace Jovanovich, 1981).

Bakan, David, "The Test of Significance in Psychological Research," in Bernhardt Lieberman (ed.) *Contemporary Problems in Statistics* (New York: Oxford University Press, 1971).

Binder, Arnold, "Further Consideration on Testing the Null Hypothesis and the Strategy and Tactics of Investigating Theoretical Models," in Bernhardt Lieberman (ed.) *Contemporary Problems in Statistics* (New York: Oxford University Press., 1971).

Boatsman, James R. and Crooch, Michael G., "An Example of Controlling Type II Error for Substantive Test in Auditing," *The Accounting Review*, July 1975, pp. 610-614.

Grant, David A., "Testing the Null-Hypothesis and Strategy and Tactics in Investigating Theoretical Models," in Bernhardt Lieberman (ed.) *Contemporary Problems in Statistics* (New York: Oxford University Press, 1971).

Guy, Dan M., *An Introduction to Statistical Sampling in Auditing*, (New York: John Wiley and Sons, 1981).

Hermanson, Roger H. et. al., *Auditing Theory and Practice*, (Homewood, Illinois: Richard D. Irving, 1976)

Jaenicke, Henry R., *The Effect Litigation on Independent Auditors*, Research Study No. 1 of the Commission on Auditors' Responsibilities of the AICPA, 1977.

Kell, Walter G. and Ziegler, Richard E., *Modern Auditing*, (Bostom: Warren, Gorham and Lamont, 1980).

Pushkin, Ann B., "Presenting Beta Risk to Students, "*The Accounting Review*, January 1980, pp. 117-122).

Roberts, Donald M., "Controlling Audit Risk- A Method for Optimal Sampling Design, "*Journal of Accounting*, Auditing and Finance, Fall 1980, pp. 57-69.

Robertson, Jack C., *Auditing*, 2nd Ed., (Dallas: Business Publication Inc., 1977).

Rozeboom, William W., "The Fallacy of the Null Hypothesis Significance Test," in Bernhardt Lieberman (ed.) *Contemporary Problems in Statistics* (New York: Oxford University Press, 1971).

Stettler, Howard, *Auditing Principles*, (Englewood Cliffs: Prentice–Hall, Inc., 1982).

Taylor, Charles, "Interpretation, and the Science of Man," in Fred R. Dallmayr and Thomas A. McCartily (eds.) *Understanding and Social Inquiry* (Notre Dame: University of Notre Dame Press, 1977).

Wilson, Warner et al., "Much Ado about the Null Hypothesis," in Bernhardt Lieberman (ed.) *Contemporary Problems in Statistics* (New York: Oxford University Press. 1971).