United States General Accounting Office

GAO

130023

May 1986

Guide for Conducting Report Conferences

PREFACE

Report conferences are intended to bring together key staff after the job results are known but before extensive writing is done. Through a report conference, staff and management can examine the job's findings and agree on the message of the assignment and how the message will be conveyed (chapter report, letter report, briefing report, fact sheet, etc.). The Task Force on GAO Reports called for the adoption of such conferences as a means of enhancing job and product quality and improving the timeliness of products. It was anticipated that the technique would improve coordination and communication during the product drafting and processing stages.

The Office of Quality Assurance (OQA) issued a draft concept paper on conducting report conferences in September 1983. In 1985 and in 1986, OQA and the Denver Regional Office conducted three separate studies to learn how to best make use of conferences. The studies have shown that conferences can be beneficial to completing the job and preparing the product. They also provide a means for focusing on the job's objectives and findings in order to reach agreement on the product's main message, format, and organization. These guidelines reflect what has been learned about the factors which contribute to successful conferences.

Donald J. Horan

Donald J. Horan
Assistant Comptroller General
Planning and Reporting

May 30, 1986

CONTENTS

	Page
INTRODUCTION	1
WHAT IS A REPORT CONFERENCE?	1
WHAT ARE THE BENEFITS OF CONFERENCES?	3
IS A REPORT CONFERENCE REQUIRED?	4
WHAT ARE THE ELEMENTS OF A SUCCESSFUL REPORT CONFERENCE?	4
Adequate preparation	5
Informal atmosphere	7
Assign a facilitator and recorder	8
Allow sufficient time	9
Right people present	9
Approprate timing	12
In-depth discussion of findings and support	12
Written memorandum of conference decisions	14
HOW IS A REPORT CONFERENCE CONDUCTED?	14
Executive Summary	15
Outline	16
AAPENDICES	
I Example of a Report Conference Agenda	17
II Questions Helpful to Assessing Job Results and Developing the Products Message	18
III Example of a Message Design Form	23

CONDUCTING REPORT CONFERENCES

Report conferences are used by the General Accounting Office as a technique to improve product quality and timeliness. This policy encourages agreement on a product's message before substantial resources are expended on writing a first draft.

When report conferences are not held, line managers have traditionally waited until they review the draft report to make decisions about the its content. Such practices often result in inefficient use of drafting time and effort. Via a report conference, decisions on a product's content can be made before the audit team has invested time and effort in producing a first draft. While agreement on what we want to report--our message--can be reached in various ways, bringing appropriate division and region managers together with the project staff in a carefully structured report conference is an effective approach.

The guidance presented here is intended to help GAO staff conduct successful report conferences. It is based on collective research and GAO's experiences with report conferences.

WHAT IS A REPORT CONFERENCE?

A report conference is a meeting of the key staff, line managers associated with a particular job, and other staff (writer-editors, technical advisory group, legal staff, etc.) as appropriate. The staff come together to review and agree on the following:

--The job's message; the reporting objectives, the conclusions or answers to the objectives, the findings, the evidence that will be used and how it will be marshalled to support the conclusions, and the proposed recommendations.

--How the message will be conveyed; the product type--chapter report, letter report, briefing report, fact sheet, testimony, etc.-- the audience for the product, the tone, the significance of the issues, the sequence of the findings, and who should comment on the draft.

The result of these agreements should be an outline with charge paragraphs and/or an executive summary and the assignment of writing responsibilities.

A report conference is usually one of the last meetings for discussing and deciding major issues which affect a product's content. During scoping and planning, the job team and managers agree on the review objectives and decide what evidence to gather to accomplish them. During implementation, team members hold meetings to review progress, assure objectives are being met, assess the evidence being gathered, and preview the message. At each succeeding step, audit staff should become more certain about what they will be able to report.

WHAT ARE THE BENEFITS OF CONFERENCES?

Report conferences improve coordination and communication during product writing and processing by

--informing those who will review and approve the product of the subject and related issues,

--insuring that the requestor's questions have been adequately addressed,

--surfacing and resolving differences or weaknesses in the message,

--clarifying for the team (especially the writers) the decisions staff and managers have made,

--providing staff with the opportunity to participate in discussions about how their work will be used in preparing the product,

--allowing line managers to assess the level of difficulty of the product writing task and to assign additional staff (e.g., writer-editor) to help draft the product, if appropriate, and

--ensuring that people at different sites and regions are working with the same message and presentation in mind.

Resolving these issues makes writing easier and faster. The savings in time and effort in the writing and reviewing process are expected to more than compensate for the investment of resources in a report conference.

IS A REPORT CONFERENCE REQUIRED?

A report conference is a valuable tool for reaching agreement on a product's message, and should be used for most assignments. However, for some jobs, an informal conference call and/or use of electronic communications between the Evaluator-In-Charge (EIC), Group Director, Assistant Regional Manager, (ARM), and Associate Director may be all that is necessary to reach agreement on a short report with a clear-cut, undisputed message.

Divisions may specify particular types of jobs that require conferences, or they may require them for all jobs but tailored to each job's needs. It is important that an early agreement on the product's message is reached whether or not a formal conference is held. Experience has shown that failure to do so usually leads to problems in report processing.

WHAT ARE THE ELEMENTS OF A SUCCESSFUL REPORT CONFERENCE?

While there are differing views on how to conduct a report conference, there is general agreement that the following elements contribute to a successful conference:

--Prepare for the conference by developing and circulating material in advance.

--Provide an atmosphere that is conducive to interactive discussions and questions.

--Designate a leader to maintain focus of the discussion and a recorder to take notes on the conference's discussions and summarize the agreements at the end of the session.

--Allow sufficient uninterrupted time to discuss the product's message and presentational issues.

--Ensure key staff attend the conference--those who are familiar with the work, those responsible for preparing the product, technical advisors, and line managers who will approve the product.

--Hold the conference when the staff has completed sufficient work to enable them to discuss the assignment results and decide on the product's message. Allow time after the report conference to complete gaps identified in the work and follow-up questions.

--Discuss the findings in sufficient detail to assess the sufficiency and relevance of their support and the link between the findings and the assignment's objectives, conclusions and recommendations.

--Prepare a written memorandum describing the agreements reached during the conference and distribute it to each participant.

Adequate preparation

Thorough preparation is essential for a successful report conference. Preparation includes analyzing the work results, developing and circulating preconference materials, and making meeting arrangements.

A few steps taken before the conference can help to ensure the participants are prepared. Under most circumstances, analyzing the job results necessitates an early discussion between key job staff (including regional staff, the Group Director, and the EIC) to confirm the report objectives, decide what can/should be discussed in the product, test the evidence to make sure it supports these findings, and develop conclusions and recommendations. Discussions might include:

--Have findings discussed at earlier meetings changed?

--Do they respond to the objectives?

--If applicable, have the objectives changed from the request letter? If changes have occurred, have they been discussed with the requestor and documented in contact memos?

--Have recent legislative or political changes affected the job results?

--Should some information be dropped from the proposed product/outline?

--Does the evidence for each finding's elements meet the standards of sufficiency, competence, and relevance?

--Is there a theme that ties the findings together?

--Is there a need to perform additional work in order to better support/link the findings?

If confident of the results of these discussions, the team may begin discussing presentational issues, such as the type of final product, and to prepare a preliminary outline, with charge paragraphs.

The conference preparation should result in a statement of the assignment and reporting objectives, a summary of the scope and methodology, a summary of the answers to or conclusions on the objectives, and a list of findings with key supporting evidence. These summaries, along with pertinent congressional correspondence and a conference agenda (see Appendix I, page 17, for an example of a conference agenda), should be circulated to all conference participants to review in advance of the meeting. These materials help inform attendees of the job's subject matter and serve as the starting point for discussion at the conference.

During the preparation stage, arrangements should be made so that essential participants can attend the conference. Regional and headquarters staff should work together to determine who will attend and to ensure that the conference objectives are met. When making arrangements, consideration should be given to holding the conference away from the regular worksite, since it minimizes interruptions caused by competing priorities. All key attendees should be able to give their full attention to the issues.

Informal atmosphere

The EIC and Group Director should take steps to make sure that staff is aware that the report conference is a working session where people are expected to discuss the results of the work and decide on the best way to communicate it in a GAO product. For the conference to work best, participants must understand that everyone is a team member trying to assess the job results and work out the best way of presenting them.

Since the purpose of the conference is to surface and resolve as many issues as possible, it should not be viewed as merely a review of prepared materials. It is important that all those attending the meeting systematically discuss issues and facts in a congenial fashion. The discussion should be a team exercise in a nonthreatening environment with everyone present participating.

The key to creating this informal atmosphere is for each participant to come to the conference with an open mind and a willingness to reach consensus on the product's message. This is not to say that participants should not enter the conference with a clear idea of what they hope to accomplish--this is what gives the conference much of its focus and direction. Instruments such as message design forms and outlines give the assignment's staff an idea of what the product will look like on paper. However, the key is to treat these ideas as starting points which will be refined to more effectively and adequately present the product's message.

Assign a facilitator and recorder

The facilitator ensures the roles of all conference participants are understood and, throughout the conference, maintains the focus of the discussion, elicits comments, and assures that important points are thoroughly discussed and carried to their logical conclusion. The facilitator is usually the Group Director, Associate Director, or a senior person who participated in advance preparation.

Someone should also be assigned the tasks of a recorder. The recorder should write agreements reached during the conference on a board or flip chart for all participants to see. Before the meeting ends, the recorder should review the notes with the group to be sure important points have not been omitted or misunderstood. It may also be helpful to have someone to type agreements reached or other items of importance prior to the conference conclusion.

Allow sufficient time

Sufficient time should be scheduled to permit discussion of all pertinent issues. Depending upon the job, the conference may take several hours or a couple of days. Variables that may affect the length of a conference include complexity of the job, prior involvement of management, comfort with the issues, and degree of multi-regional staffing.

In some cases, the conference may surface gaps that require further work. In such cases, a second conference may be needed when the work is completed. Depending on the extent of work required, other arrangements, such as a follow-up telephone call to discuss the results of the additional work, may be used.

Right people present

Staff who are most familiar with the work, staff responsible for preparing the product, and managers who will approve the product's contents should attend. Essential report conference participants are:

--EIC or project manager,

--key audit team and technical assistance staff involved in data collection and analysis, and

--Group Director.

Whenever possible, the Associate Director should attend the conference. It is critical that the Associate Director agree with decisions made at the report conference concerning the product's message and presentation in order for the conference to be effective in reducing the drafting and review timeframe. If the Associate Director cannot attend, the Deputy Associate Director, Group Director, or other senior person in attendance is expected to represent management. The Associate Director should then be briefed on the agreements reached in order to obtain his/her concurrence on the planned message before substantial time is spent drafting the product. Another option is to use electronic communications to transmit the agreements reached to the Associate Director. This may aid in the Associate Director's review and concurrence before the conference reaches closure.

Attendance of staff responsible for preparing the product will help ensure that agreements reached are reflected in the product and can speed the writing process. Therefore, the individuals assigned to do the writing (preferably experienced GAO writers) should participate in the conference.

In many cases it will be important that other people who have had a key role in the job or who are expected to have a key role in reviewing the product also participate in part or all of the conference. These include:

--regional manager, ARM, or Regional Management Representative from lead region,

--division report reviewer,

--division or regional writer-editor,

--subject area experts (inside or outside of GAO), and

--representatives from other divisions and offices with whom the report must be coordinated.

The conference process can be aided by having a "cold reader" attend to provide an independent point of view by asking probing questions about the evidence, findings, conclusions, and recommendations. One of the previously identified conference participants or someone completely independent from the job could serve as the cold reader.

The circumstances under which the Division Director or the Deputy Director for Planning and Reporting should attend a conference should be determined on a case-by-case basis, e.g., major GAO jobs or groups of large-scale, multi-region audits concerning controversial issues.

Since a report conference is a working, decision-oriented meeting, the objective should be to keep it small; only key people should attend. Usually, 10 to 12 participants is the largest group that can work effectively.

Appropriate timing

Generally, the appropriate time to hold the conference is near the end of field work and before staff have been released from the assignment. It is important that there is time allowed and staff available after the report conference to follow up on points raised at the meeting. In any event, a report conference should not be scheduled until the team has analyzed sufficient data to formulate, with some confidence, the job's findings, supporting evidence, conclusions, and recommendations, and is ready to have them approved by the Group Director and Associate Director (and other top managers as appropriate) before substantial time is invested in writing.

The tendency to hold this meeting prematurely and base the discussion on expectations rather than on what has actually been found misses the point of the conference and represents a potential threat to its utility. On the other hand, experience indicates that if extensive, detailed report writing has not been done, staff are more open to new ideas and the conference is more valuable to them if they are not committed to one way of presenting the product message.

Indepth discussion of findings and support

A detailed discussion of evidence obtained to support the findings is important to assess the job's issues, prioritize the findings, and determine the product's main message. In-depth analysis of each major finding should be performed. Depending on the job's objectives, this may include questioning and identifying the sources of criteria, condition, cause, and effect. Discussion of the job's objectives and the link between these objectives and the elements of the finding(s) is **critical.** The discussion could identify gaps which require additional work or, alternatively, findings or evidence not pertinent to the report's message which could be eliminated.

It also helps the writer(s) prioritize the issues and organize the presentation of facts.

Conference participants should review the Report Manual chapters on evidence, findings, conclusions, and recommendations for guidance on types of questions to be asked and examples of common problems to avoid. (Appendix II, pages 18 through 22, provides some questions that can be helpful in assessing the job results and developing the message.) There are different processes that will help facilitate the discussion of findings, such as completing a "message design form" or a "logic chart." Both of these documents include basically the same information but in different formats. (An example of a "message design form" is included as Appendix III.) An in-depth discussion of findings and support may also be facilitated by other processes, such as developing an executive summary and/or preparing an outline (see pages 15 through 16).

If several findings or the results of the work from several regions are to be discussed, it may be advisable to complete the forms before the conference so that sufficient time can be spent discussing their content during the conference.

Written memorandum on conference decisions

A memorandum or document describing the agreements reached during the report conference should be prepared and distributed to each participant. This is important because conference participants may interpret discussions differently. A product outline that includes charge paragraphs, or an executive summary, may be circulated to show how the agreements will be implemented. A timetable for drafting deadlines should also be prepared and circulated.

The Associate Director can refer to the outline when reviewing the first draft. If, as the writers begin drafting, some of the conference agreements are modified, all participants should be informed to assure that expectations and understandings are in accord.

HOW IS A REPORT CONFERENCE CONDUCTED?

No prescription exists for conducting a report conference. Focusing discussions on developing an executive summary or an outline are two techniques that have been found to be useful in GAO to facilitate agreement on the product's message.

Executive Summary

Developing the standard modules of an executive summary is an effective tool for reaching agreement on a report's message. It has been found to be a successful approach even if the final product is a letter report.

When an executive summary is to be developed during the meeting, participants would begin by discussing the purpose of the review, and specifically the report's objectives. The audit team might, as part of advance preparation, have written what it believes the reporting objectives are on a flip chart or on transparent vu-graphs that can be changed to reflect what is agreed upon. If the reporting objectives differ from the original objectives, this should be discussed.

After reaching agreement on objectives, the participants would discuss and agree on the report's main message (Results in Brief)--their bottom-line response to the objectives--and the evidence supporting this response (Principal Findings or GAO's Analysis). This discussion should encompass all findings and conclusions the team wishes to report. The findings can then be grouped according to major issues, and the participants can decide which are the most significant. Once these agreements have been made, the participants can begin to critique and test the recommendations. Participants should closely examine the link between the recommendations and the causes of the identified problems.

Finally the conference participants should decide what background information is needed to understand the reporting objectives and message. The background should include program-specific information to help put the message in its proper context.

Outline

Another useful technique for focusing the report conference discussions and reaching agreement on the product's message is to prepare/revise an outline. The length and detail of the outline will vary depending on the complexity of the product's message. The more detailed the outline is, including charge paragraphs and topic sentences, the less chance there is of misinterpreting the report's main message and its presentation.

Detailed outlines provide format and serve as a "skeleton" of the product, clearly showing its structure. The charge paragraphs will embody the product's message, its scope, order, and tone. Branching from the charge paragraphs are the headings, subheadings, and topic sentences supporting the message and giving it organization. If the outline is developed to this extent in the conference, the writer essentially just needs to fill in the supporting details to complete the draft.

EXAMPLE OF A REPORT CONFERENCE AGENDA

I. Introduction

   A. Administrative matters
   B. Schedule for conference

II. Agree on the product message

   A. Purpose of report
   B. Background
   C. Conclusion/results in brief
   D. Findings and evidence or GAO's analysis

      1. Major
      2. Secondary

   E. Recommendations

III. Agree on how to present the message

   A. Issues

      1. Product type
      2. Audience
      3. Tone
      4. People to comment on draft
      5. Significance and order of information

   B. Results

      1. Outline/Charge paragraphs/Executive Summary
      2. Writing responsibilities
      3. Timeframes to complete the writing and to process the report.

QUESTIONS HELPFUL TO ASSESSING JOB RESULTS AND DEVELOPING THE PRODUCT'S MESSAGE

REPORTING AND ASSIGNMENT OBJECTIVES

- Are there differences in the reporting and assignment objectives?
- Are they clear, do I understand them, and do they make sense?
- Are they measurable; are they too general to be meaningful, do they allow reader focus?
- Are reporting objectives doable?
- Are they biased, are unstated assumptions implicit in them?
- Are they linked to the request or, if not, are they reconciled?
- What expectations do they set; what elements of a finding would a reader expect to see?

SCOPE AND METHODOLOGY

- Is there a clear/understandable statement of what work was done to respond to the objectives?
- Are there limitations to the work that should be recognized in reporting the findings?
- If the methodology is complex, should review by a technical specialist be obtained?

FINDINGS

o Are all the elements [1] necessary to meet job objectives adequately developed?

o CRITERIA: what should be the state of affairs that is desired or required? (What condition is compared to in order to demonstrate a deficiency or achievement.)

-Is it stated fairly, explicitly, and completely?

-Is the source identified (laws, regulations, procedures, etc. or assertions)?

-Do we need OGC agreement on laws, regulations, or congressional intent?

-Is the applicability self-evident or explained?

-Are alternatives dealt with and explained?

-Does it make sense; is it consistent with sound management principles?

o CONDITION: what is the actual state of affairs?

-Is it clearly stated and shown to exist through convincing evidence?

-Is an accurate perspective given on extent or scope?

-If gaps exist, are they so significant as to give a seriously incomplete picture? (This could affect conclusions and recommendations.)

1 Depending on the particular job's scope and objectives and on the evidence gathered, a traditional audit finding may include one or more of the elements criteria, condition, cause, or effect. Other types of jobs--e.g., evaluations or economic forecasts--could have different elements.

- <u>CAUSE</u>: What factors are responsible for differences between condition and criteria?

  - -Are they clearly stated?
  - -Is there a reasonable and persuasive argument for why they contributed to the difference as opposed to other possible causes?
  - -How strong/convincing is the evidence supporting them?
  - -Do they recognize the role of judgment?
  - -Do they provide the basis for recommendations?
  - -Are they attributable to internal management control weaknesses?

- <u>EFFECT</u>: what is the significance of a difference between a condition and criteria, the consequences of the difference, or the impact of an intervention (policy, program, procedure or action)?

  - -Is it stated clearly, concisely, and, where possible, in concrete terms?
  - -Is it distinguished from condition?
  - -Is there a clear, logical link to condition, criteria, and cause?
  - -Is the significance demonstrated through credible evidence?

- EVIDENCE:

  -What type of evidence has been gathered?

    - Is it testimonial, observation, documentary, or analytical?

    - How good/convincing is it? Will it withstand the most likely opposing views?

    - Is corroborating evidence needed?

  -Have the sources of the evidence obtained been identified?

  -Was the information used to develop the evidence based on the most current data available?

  -Is the evidence valid, complete, and relevant (is there a logical, sensible relationship between the evidence and the audit issues)?

CONCLUSIONS

- Do they clearly and logically flow from the findings?
- Do they overstate/understate findings?
- Do they recognize limits of findings and scope of work?
- Are they placed in context of real-world constraints?
- Do they provide balance/perspective?

RECOMMENDATIONS

- Is what needs to be done clear?
- Are they directed to the appropriate person/level?

- o Is there a clear logical link to identified problem and cause?
- o Are they feasible, practical, and workable?
- o Are there any adverse effects?
- o Are alternatives considered and dealt with?

AGENCY COMMENTS

- o Have the findings been discussed with agency officials?
- o What is the most likely reaction to our message? If sharp disagreements are expected, will we be able to overcome opposing views?
- o How will agency's agreements or disagreements be handled in the report?

EXAMPLE OF A MESSAGE DESIGN FORM [1]

Job Objective: Determine if Farmer Home Administration's (FmHA) procedure for testing farm well water for contamination was followed. If not, determine why and if citizens have been exposed to contaminated water from untested wells.

Condition: In 15 of the 34 counties GAO examined, individual water systems of a FmHA unit were not tested for contamination.

Criteria: FmHA agency bulletin requires that farm wells be periodically tested for chemical bacteria and be free of contaminants. Department of Health and Human Services water standards for human use are applied.

Discrepancy: Tests of water systems were not performed by all counties, as required. GAO noted 44 percent (15 of 34 counties) with no inspection.

Effect: Audit tests revealed contamination in 20 percent of untested wells. Therefore, the health of residents using water from bacterially contaminated wells was endangered as a consequence of the discrepancy.

Cause: FmHA personnel were unaware of the requirement and did not request health department well tests. FmHA management was not monitoring compliance with its testing requirement.

Conclusion: FmHA procedure for testing well water was not followed in 44 percent of counties examined. We found that 20 percent of these untested wells were contaminated. As a result, residents are using water from bacterially contaminated wells.

Recommendation: The FmHA Administrator should promptly notify all county offices of its well water testing requirement and establish a system to monitor compliance.

[1]This example illustrates a traditional audit finding that may include the elements criteria, condition, cause, and effect. Depending on the particular job's objectives, scope, and methodology, and on the evidence gathered, the elements of a finding may differ.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

Address Correction Requested

Special Fourth Class Rate
Postage & Fees Paid
GAO
Permit No. G100

United States General Accounting Office 34348

GAO

Program Evaluation and Methodology Division

April 1986

Using Statistical Sampling

Transfer Paper 6 035334/129810

Using Statistical Sampling

Transfer Paper 6

Program Evaluation and Methodology Division

April 1986

PREFACE

Sampling is a very important element in the design of an audit or evaluation. The purpose of this paper is to help GAO managers and evaluators learn more about statistical sampling and the role it plays in project design and execution. We have attempted to take the mystery out of what is often thought of as an esoteric subject by "walking" the reader through the various sampling procedures.

Using Statistical Sampling describes sample design, selection and estimation procedures, and the concepts of confidence and sampling precision. Two additional topics, treated in a briefer fashion, include special applications of sampling to auditing and evaluation and some relationships between sampling and data collection problems. Last, but not least, the strengths and limitations of statistical sampling are summarized. Some topics of a more technical nature appear in the appendixes, as follows:

- Appendix I discusses some of the theory behind sampling.

- Appendix II presents a comprehensive description of sampling procedures.

- Appendix III discusses the computations used for stratified and cluster sampling.

- Appendix IV is an annotated bibliography of books on statistical sampling.

- Appendix V lists various packaged, or "canned," computer programs that can do sampling computations.

GAO policy on sampling is set forth in the General Policy Manual, pages 7-7 and 7-19.

In this document, we have chosen to describe computations and sample selection procedures as if they were done manually, even though computer programs are typically used to select samples, determine sample sizes, and calculate estimates. We have done this because we think it is important that the persons using those programs understand what is going on inside the machine; otherwise, they would just be feeding the computer the required data and accepting uncritically whatever came out.

This paper makes the assumption that the reader has had a one-semester college course in statistics. However, those who have not had such a course (or who think they may have forgotten the basics) can refer to appendix I. We do not expect that every GAO evaluator, after reading this paper, will be able to design and carry out a statistical sampling plan without assistance. Rather, we hope to provide enough background on sampling concepts and methods to enable evaluators to (1) identify jobs that can benefit from statistical sampling, (2) know when to seek assistance from a statistical sampling specialist, and (3) work with the specialist to design and execute a sampling plan.

Using Statistical Sampling is one of a series of papers issued by the Program Evaluation and Methodology Division (PEMD). The purpose of the series is to provide GAO evaluators with handy, clear, and comprehensive guides to various aspects of audit and evaluation methodology, to explain specific applications and procedures, and to indicate where more detailed information is available. Other papers in the series include Designing Evaluations, Causal Analysis, Content Analysis, and Using Structured Interviewing Techniques. We welcome the comments of all our readers, who are encouraged to let us know of any questions, suggestions, or reactions they may have. These should be addressed either to Carl Wisler, Associate Director, or to me.

Eleanor Chelimsky
Director

Contents

		Page
PREFACE		iii
CHAPTER		
1	INTRODUCTION	1
	Project design	1
	Sample design as an element of project design	2
	Representativeness: The goal of statistical sampling	3
	Random selection	4
	Distinguishing between samples and examples and between sampling operations	4
	The organization of this paper	5
2	AN OVERVIEW OF SAMPLE DESIGN AND SELECTION AND ESTIMATION PROCEDURES	7
	Sample design	7
	Selection procedures	12
	Estimation procedures	13
	Simple random sampling	14
	Stratified sampling	14
	Cluster sampling	15
3	BASIC ESTIMATION PROCEDURES AND FURTHER SAMPLE DESIGN CONSIDERATIONS	17
	The concepts of precision and confidence	17
	Sampling for variables	20
	Sampling for attributes	29
	A distinction between precision and accuracy	33
4	ADVANCED ESTIMATION PROCEDURES	34
	Ratio estimation	34
	Regression estimation	40
	Difference estimation	44
	Other advantages of ratio, regression, and difference estimation	48
5	SAMPLING IN THE AUDIT ENVIRONMENT	49
	Discovery sampling	49
	Acceptance sampling	52
	Audit judgment and statistical sampling	54
	The characteristics of a good sample	55

		Page
CHAPTER		
6	SAMPLE SELECTION PROCEDURES	58
	The selection of sampling units	58
	The application of selection procedures	64
	Selection with probability proportional to size	64
	A final check	66
7	DATA COLLECTION AND ANALYSIS CONSIDERATIONS RELATED TO SAMPLING	67
	Checking the reliability of data collection	68
	The problems of missing data and nonresponses	69
8	THE STRENGTHS AND LIMITATIONS OF STATISTICAL SAMPLING	72
APPENDIX		
I	Theoretical background	75
II	A comprehensive description of sampling procedures	89
III	Details on stratified and cluster sampling	116
IV	A selected annotated bibliography	141
V	Four computer software packages	143
VI	References	146
VII	Glossary	147
FIGURE		
3.1	Final sample size as universe size increases: First estimate of sample size is 100	28
I.1	Number of samples classified by percentage of red beads	77
I.2	The relative frequency of a large population of single digits between 0 and 9	78
I.3	The relative frequency of means of samples of two from a population of digits between 0 and 9	79

		Page
FIGURE		
I.4	The relative frequency of means of samples of three from a population of digits between 0 and 9	80
I 5	The smoothed curve of the relative frequency of means of samples of three from a population of digits between 0 and 9	81
I.6	Frequency distributions of the means of samples drawn from a universe of 228 U.S. cities with populations greater than 50,000 in 1950	83
II.1	Random number selection work sheet	95
II.2	Random number selection work sheet used for selecting 50 random numbers	97
TABLE		
3.1	An evaluation of sample sizes used in GAO studies	19
3.2	Work sheet for computing the mean and standard deviation of sample data	21
4.1	Work sheet for computing the sampling error of ratio for 50 small purchases	36
4.2	Audited results of the preliminary sample of 50 small purchases	45
4.3	Differences and squares of differences	46
5.1	The probability of finding at least one error in a universe of 500 for various numbers of errors and sample sizes	50
5.2	The probability of finding at least one error in a universe of 600 for various numbers of errors and sample sizes	51
5.3	An acceptance sampling table	53
6.1	Selection with probability proportional to size	65
I.1	Dice-throwing probabilities	76

Page

TABLE

TABLE		Page
I.2	Results of bead sampling experiment	76
I.3	Work sheet for computing the mean and standard deviation of universe data	85
I.4	Table of t factors	87
II.1	Random decimal digits	90
II.2	Selecting random numbers and random letters in compound numbering systems with the same quantity of items for each letter	101
II.3	Matching random numbers and random letters in compound numbering systems with the same quantity of items for each letter	102
II.4	Sorting random number-letter combinations in compound numbering systems with the same quantity of items for each letter	102
II.5	Selecting random numbers and random letters in compound numbering systems with a different quantity of numbers for each letter	104
II.6	Matching random numbers and random letters in compound numbering systems with a different quantity of items for each letter	104
II.7	Sorting random number-letter combinations in compound numbering systems with a different quantity of items for each letter	105
II.8	Selecting a single systematic sample from four groups of records	108
III.1	Savings from using direct air shipment, instead of air freight forwarder, classified by air freight forwarder shipping costs of less than $100	118
III.2	Savings from using direct air shipment, instead of air freight forwarder, classified by air freight forwarder shipping costs of more than $100	119
III.3	The proportional allocation of samples to strata	120

TABLE		Page
III.4	Proportional allocation: Sample shipments and the computation of totals	120
III.5	Work sheet for computing sampling error of total with proportional allocation	122
III.6	An example of the judgmental allocation of shipping costs	122
III.7	Work sheet for computing estimated total savings with judgmental allocation	123
III.8	Judgmental allocation work sheet for computing sampling error of total	123
III.9	Work sheet for allocating samples to strata with Neyman allocation in sampling for variables	124
III.10	Neyman allocation: Sample shipments and the computation of means	125
III.11	Neyman allocation: Computation of estimated total savings	125
III.12	Neyman allocation work sheet for computing sampling error of total	125
III.13	Work sheet for computing sample size	127
III.14	The results of a review of civilian payroll records at three military bases	129
III.15	The computation of the estimated number and percentage of payroll records with errors	129
III.16	Work sheet for computing payroll sampling error	129
III.17	Work sheet for computing payroll sample size	131
III.18	Work sheet for allocating samples to strata with Neyman allocation in sampling for attributes	132
III.19	Work sheet for allocating samples to strata with assumed stratum mean	134
III.20	Number and dollar amounts of transactions	135
III.21	The results of a review of 40 bundles of requisitions	137

ABBREVIATIONS

LTPD	Lot tolerance percent defective
PPS	Probability proportional to size

CHAPTER 1

INTRODUCTION

Sampling is nothing new or unusual. For thousands of years, people have been basing judgments about a large group of objects on their observations of a few of them. Prehistoric humans probably decided whether the berries on a bush were edible by tasting a few of them (with possibly fatal results). At harvest time, farmers judged the quality and expected yield of a wheat field by rubbing the husks off a few ears of grain pulled from various parts of the field. People have used sampling techniques such as spot checking for many years. The great improvement in the last 50 years or so has been the development of statistical sampling. We now have ways of drawing and analyzing samples to produce more objective information of better quality and of being explicit about its limitations.

Sampling is one aspect of GAO projects and, consequently, the design of a sample is one part of an overall project design. The time to start consideration of sampling is during project design.

PROJECT DESIGN

The design of any project starts with the question being asked. Designing Evaluations, a PEMD transfer paper issued in July 1984, classifies audit or evaluation questions as descriptive, normative, or cause and effect.[1] The answers to descriptive questions provide information on existing conditions. The answers to normative questions compare an observed outcome with a given level of performance. The answers to cause-and-effect questions indicate whether observed conditions, events, or outcomes can be attributed to program operations. The methods used to answer evaluation questions, known as "audit" or "evaluation" strategies, can also be classified. As discussed in Designing Evaluations, the strategies and the types of questions most commonly addressed by each strategy are

Audit or evaluation strategy	Type of evaluation question most commonly addressed
Sample survey	Descriptive and normative
Case study	Descriptive and normative
Field experiment	Cause and effect
Use of available data	Descriptive, normative, and cause and effect

---

[1]Here we refer to the few broad questions that dictate an evaluation's objective; later we will be concerned with the much narrower issues that must be addressed in designing a sample.

In a sample survey, data are collected from a sample of a universe to determine the prevalence, distribution, or interrelationship of events and conditions. The case study analytically describes an event, a process, an institution, or a program; this strategy can use either a single case or multiple cases. The field experiment compares outcomes of program operations with estimates of what the outcomes would have been in the absence of the program. The use of available data refers to the use of previous reviews or data bases previously collected and still available.

No matter which strategy is used, evaluators need to consider several elements in designing a job. Designing Evaluations lists seven design elements:

1. kind of data to be acquired,
2. sources of information (for example, types of respondent),
3. methods for sampling information sources (such as statistical sampling),
4. methods of collecting data (such as self-administered questionnaires),
5. timing and frequency of data collection,
6. basis for comparing outcomes with and without a program (for cause-and-effect questions), and
7. analysis plan.

In this paper, we are concerned primarily with the methods used for sampling information sources. Although we briefly discuss data collection methods, two other PEMD transfer papers, Developing and Using Questionnaires (publication pending) and Using Structured Interviewing Techniques, describe these two methods in much greater detail. In the future, we will issue documents detailing the four evaluation strategies as well as data analysis methods.

SAMPLE DESIGN AS AN ELEMENT OF PROJECT DESIGN

In the context of auditing and evaluation, a sample is a portion of a universe of possible information sources and sampling refers to the methods for selecting those sources. Sampling is an element of project design, which, along with such other elements as data collection and analysis methods, determines the soundness of the answers to our evaluation questions.

The broad sampling options available may be understood by an example. Suppose we want to know how federal centers for runaway youths are operated. For certain kinds of information (e.g., project cost and staff size), the directors of runaway-youth centers are probably the best source of information. We may then regard the center directors as our universe of possible information sources. An important project design issue is how to select the directors from whom we will seek the information we need.

Three options are available for choosing the directors. The first possibility is to gather information from all center directors. This is called a "census," and it may be thought of as a special case of sample--a sample of all possible information sources. Sometimes, conducting a census is a desirable course, but this strategy is not the main subject of this paper.

A second possibility is to apply a judgmental process to the selection of center directors. We might look at the locations of the centers and choose directors so that cities, suburbs, and rural areas are each represented to some degree. Or, bearing in mind the cost of travel to the center sites, we might choose the directors who are located closest to our office. Judgment sampling, which can be used to select information sources in many different ways, is largely outside the scope of this paper.

The third possibility is to select center directors by statistical sampling. Here, chance determines which directors are selected. Most of this paper is devoted to statistical sampling and to describing the variety of ways in which chance can be invoked (sample design), the processes for choosing information sources (selection procedures), and the methods for drawing conclusions about a universe based on information about a sample (estimation procedures).

REPRESENTATIVENESS: THE GOAL OF STATISTICAL SAMPLING

In many GAO projects, the objective is to answer questions about a universe of people or things. In the example of centers for runaway youths, we wanted to know how the projects in the universe of projects for runaway youths were operated. This objective can be achieved by looking at a sample of projects, if the sample is representative of the universe.

A representative sample has approximately the same distribution of characteristics as the universe from which it was drawn. A detailed discussion of the concept of a representative sample is outside the scope of this paper but most people have an intuitive understanding of representativeness (see Kruskal and Mosteller, 1979 and 1980, for an extensive treatment).[2] A representative sample of runaway-youth centers is like the

---

[2]Appendix VI contains complete bibliographic data.

universe in terms of characteristics such as number of center staff, types of youths who come to the center, average duration of stay, and so on. With such a sample, we infer that the characteristics of the universe, which we do not know, are like the characteristics of the sample, which we do know.

Statistical sampling produces a sample that, it can be persuasively argued, is representative of a universe. However, samples of a universe differ from one another as well as from the universe itself. Hence, it is desirable to have an objective measure of the possible variation between samples and of the sample's relationship to the universe. With this information, it is possible to determine the amount of error that arises because our sample does not correspond exactly to the universe. This is an important feature of statistical sampling. It allows us to be precise about the error introduced by the sampling process. We can then decide whether the amount of error is tolerable when weighed against trade-off factors, such as the cost of obtaining a larger sample that will have less error.

RANDOM SELECTION

The essence of statistical sampling is selecting a sample by some random (or chance) process. By randomizing the sample selection, we make sure that the sample represents the universe, within the limits of sampling error, and we can measure the precision of the information yielded by the sample.

The term "random selection" does not mean a haphazard or "catch as catch can" sample, such as inspecting poison gas shells that are stored closest to the entrance to an ammunition bunker or interviewing "average-looking" people on street corners. Rather, to select randomly is to eliminate personal bias or subjective considerations from the choice of the sample items. Every item in the universe has an equal or known probability of being selected, and items are selected independently. Although the results obtained from different random samples drawn from the same universe differ (as we will see in chapter 3), the differences stem from chance, not personal bias or other systematic factors.

The selection of a sample by some random method in order to obtain information or draw conclusions about a universe is referred to as "probability" or "statistical" or "scientific" sampling. Regardless of the name used to describe the method, the key elements are that (1) each possible sample from the universe has a known (nonzero) probability of being selected and (2) the actual selection technique truly executes the random method.

DISTINGUISHING BETWEEN SAMPLES AND EXAMPLES AND BETWEEN SAMPLING OPERATIONS

It is important to distinguish between samples and examples. A statistical sample, as we have stated, is selected in a way such that the information obtained represents the characteristics of the universe from which it was selected. An example, however, assists in describing findings and recommendations or demonstrating a particular point. Usually, the characteristics of an example are known before it is selected. The example may be selected as a typical case, or it may be selected to represent an unusual or problem situation. Examples can be chosen from the items that were already selected in a random sample. There is no objection to using carefully selected items as examples, provided that we describe them as examples and do not imply that they are representative of the universe.

It is also important to distinguish between three interdependent sampling operations: sample design, selection procedures, and estimation procedures. Sample design refers to the plans made for the overall way in which a sample will be related to a universe. Selection procedures are the methods used to select units from a universe. Estimation procedures are the ways of estimating the characteristics of a universe from information acquired about a sample.

The sample design will affect the estimation procedures to be used, and it may also affect the selection procedures. Conversely, the sample design is often affected by the estimation procedures the evaluators want to use. Further, selection procedures can have a major effect on how precision is estimated, and the types of estimates to be developed can have a bearing on the selection procedures to be used.

THE ORGANIZATION OF THIS PAPER

Three sampling strategies are available: census, judgment sampling, and statistical sampling. This paper focuses on statistical sampling. If statistical sampling is part of a project design, the further choice of a particular sampling procedure, such as simple random sampling or cluster sampling, is necessary. And to implement the project design, two other major steps are required: sample selection and estimation of the universe characteristics.

Chapter 2 provides an overview of all three components: sample design, sample selection, and estimation. Chapter 2 also provides more detail on statistical sample designs by covering simple random sampling, stratified sampling, and cluster sampling. Chapter 3 takes up matters of basic estimation, using the concepts of confidence and precision. Chapter 4 contains more-advanced estimation procedures. Some special sampling issues that apply more to auditing than to evaluation are discussed in chapter 5. Chapter 6 discusses sample selection--the considerations in randomly selecting the sample units. Chapter 7 provides a bridge between sampling and topics on data collection and analysis, such as missing data and nonresponses. Chapter 8 briefly summarizes the strengths and limitations of statistical sampling.

For readers who feel they need a refresher on the theoretical concepts underlying sampling, appendix I provides the level of introduction that might appear in a first course in statistics. Appendixes II-V contain material that extends the information of the main text, and appendix VI contains the bibliographical data for references in the text.

CHAPTER 2

AN OVERVIEW OF SAMPLE DESIGN

AND SELECTION AND ESTIMATION PROCEDURES

SAMPLE DESIGN

Sample design is a part of the overall project design composed of the seven elements listed in chapter 1. Although designing a project is an iterative process involving the several elements, we must tentatively formulate the audit or evaluation questions and make preliminary decisions about data collection methods before trying to settle sampling issues. It is also advisable to have a preliminary data analysis plan in mind before working on the sampling design.

Sample design involves the following steps:

1. defining the universe and the sampling units,

2. choosing the sampling strategy (census, judgment, statistical) and the type of sampling (simple random sampling, stratified sampling, etc.) to be used, and

3. determining the size of the sample.

Defining the universe and the sampling units

It is necessary to define the universe very carefully, because this is the entire collection or group of items to which our estimates and inferences apply. In many projects, more than one universe of information sources will be of interest. In the example of the runaway-youth centers, it may be desirable to obtain information not only from center directors but also from staff members, youths staying at the centers, and the parents of the youths. In principle, the sampling considerations in choosing directors are simply extended to other universes but, in practice, some designs may be more advantageous than others. Sampling specialists should be consulted.

The logical starting place may be with either the universe or the sampling units. We might begin, for example, with the understanding that we want to draw conclusions about the universe of rail shipments of ammunition in 1984. We must then decide upon the sampling units. Do we want to define the sampling unit as the total shipments made by a depot, the total shipments received by a military unit, the government bill of lading for an individual shipment, or something else?

If we begin with a sampling unit defined as a government bill of lading involving rail shipments of ammunition, then we must be clear about just what universe of bills we want to draw conclusions about. Do we want to include shipments from all ammunition depots in the country or Navy depots only, and are they shipments for an entire year, a single month, or a quarter?

Once the universe has been defined, we must either obtain or develop a sampling frame. The sampling frame is a list of items contained in the universe. The list can be printed on paper, it can be a magnetic tape file or a file of punchcards, or it can be a file of accounts-receivable ledger cards or stock record cards. The frame should have several characteristics. First of all, the frame should permit the sampler to identify and locate the specific item that is to be drawn into the sample and to differentiate this item from all other items in the sampling frame. The frame should also contain all the items in the universe. For example, if the universe has been defined as the civilian work force at a naval shipyard, the list of workers from which the sample is drawn should include all civilian workers on the date of the audit or study, should contain no duplicate entries, and should contain no entries not in the universe.

In addition, we may want to define subdivisions of the universe. One type of subdivision is the stratum, a subpopulation obtained by dividing the universe into two or more mutually exclusive groups, or "strata," which we can do if we know in advance the number of sampling units in each stratum. Independent random samples are selected from each stratum in order to obtain more precise estimates or to emphasize certain portions of the universe, such as units with a high dollar value or a great potential for error. Often, the stratification system is based on the locations of the units of observation. Examples of strata are households classified as urban and rural, naval bases classified by geographic location, and taxpayers classified by income.

Another type of universe subdivision is the domain of interest. This type of subdivision is necessary when separate estimates are needed for each of a number of classes into which a universe may be divided but we do not know in advance the number of sampling units in each class. Thus, we must depend on the sample if we are to develop this information. Examples of domains of interest are students at a university who intend to major in education, travel vouchers involving the use of a personally owned vehicle, and farms worked by tenants.

The sampling units are often defined to be persons or things we want to study--the units of the universe about which we need information. But sometimes, because of the arrangement of the universe, the lack of a list of items we want to observe, and practical considerations, we may have to select a sampling unit that is larger than the item about which we want to obtain data. An example is selecting a household in order to determine the employment or health status of its members. In this example, the item of interest, the household member, is called the "secondary sampling unit," and the larger unit, the household, is called the "cluster" or "primary sampling unit."

The primary sampling units must (1) be mutually exclusive and (2) include the entire universe. This means that each unit being observed, the secondary sampling unit, must belong to one and only one primary sampling unit and that the primary sampling units must cover the entire universe.

Sometimes, the cluster or primary sampling unit consists of so many items that we must select a sample of items within each primary sampling unit. (This is called "two-stage sampling.") Occasionally, it is necessary to select a sample of primary units, a sample of secondary units from within each primary unit, and a sample of items from within each secondary unit. (This is called "multistage sampling.")

Sometimes, samples are taken in two or more phases, or "waves." This technique may be used to take a large preliminary sample, classify the sample into two or more domains of interest, and then draw smaller subsamples from the domains of interest. This type of sampling is known as "double" or "two-phase" sampling. An excellent example of double sampling cited by Cochran (1977) involved surveys of the German civilian population in 1945, when the sample from each town was usually drawn from rationing lists. It was proposed that the population be stratified by sex and age. Because the sample had to be drawn in a hurry and the rationing lists were in constant use, it was not possible to tabulate the population by sex and age. However, a moderately large sample of names could be selected quickly. Each person selected was classified into the appropriate sex-age class. From these classifications, smaller samples of persons to be interviewed were selected.

Another type of two-phase sampling is drawing repeated samples from the same universe. The usual purpose of these samples is to measure change from a preceding time period or periods and to obtain current estimates on statistics of interest. The general procedure is to replace part of the sample (or select new sample units) and retain part of the sample every time the data are collected. An example is the current population survey conducted by the Bureau of the Census to measure employment and unemployment. In this survey, one fourth of the households are replaced by new sample households each month, so that a household is in the survey for 4 months. The household is omitted from the survey for the following 8 months, brought back into the survey for 4 more months, and then dropped. The current population survey uses this procedure because (1) more accurate measures of change are obtained by looking at differences in the same units over time and (2) concern about burdening respondents limits the number of periods any one household can be included.

Choosing a sampling strategy

Three broad sampling strategies were outlined in chapter 1. The choice of a census, a judgment sample, or a statistical sample is a project-design decision of great importance. Besides the project objective, factors such as cost, precision, and the feasibility of drawing certain kinds of samples must be considered. Although this paper is primarily about statistical sampling, a brief outline of the pros and cons of different sampling strategies is appropriate. We recommend that before you decide on a sampling strategy, the project staff seek assistance from sampling statisticians in PEMD.

For some GAO projects, a census is appropriate, as when the individual items in the universe are very important in themselves or when the information to be obtained is critical and the universe is small enough to allow 100-percent sampling. On other occasions, the universe may be so small that sampling is not needed. Also, when all the data are already on a computer or in machine-readable form, it is often no less efficient to analyze every item. This is because most of the work is in setting up the programs, not in processing the items, and because the computer must read every record to select it for inclusion or exclusion. Aside from special cases, the main disadvantage of a census is usually the high cost relative to other options.

Judgment sampling is not statistical or scientific sampling: it is discretionary. In this type of sampling, the evaluator bases the selection of a sample on knowledge or judgment about the characteristics of the universe. Haphazard or "catch as catch can" samples--for example, grabbing a few items "at random"--are usually included in the category of judgment sampling.

Judgment samples have valid uses. When one need not generalize to a universe, a census or a statistical sample is not necessary and a small judgment sample might be cost effective. For example, if the objective of an audit is to show vulnerability to fraud (without regard for the probable incidence of fraud), a judgment sample may be satisfactory.

The case study approach uses judgment sampling. By definition, one of the features of the case study strategy is that it is not a census and does not involve statistical sampling of cases (GAO, 1984). There are a variety of situations in which case studies, and thus judgment sampling, would be appropriate.

Sometimes the job objective is to generalize, but it is not possible to obtain a suitable list of the universe. Statistical sampling is then not possible, and we may be forced to use a judgment sample. Although not necessarily less accurate than probability samples in describing a universe, judgment samples lack three important characteristics of statistical samples: (1) random selection of the cases to be examined, (2) scientific determination of the sample size, and (3) objective evaluation of the sample results.

The key problem with using a judgment sample when we want to generalize is that we have no way of knowing how near the results obtained are to the universe characteristic we are attempting to measure. A statistical sample's results, in contrast, can be computed and expressed in quantitative terms. That is, evaluators can numerically measure the precision of the sample results and the probability that the sample estimate is within the calculated precision of the universe characteristic.

When the objective of a project is to draw conclusions about a universe of people or things and when we can list the universe, statistical sampling is the method of choice. (It is not necessary to literally "list" the universe. For example, it is possible to randomly select from the list of all possible phone numbers without possessing a physical list of such numbers. Sometimes the list exists only in a conceptual sense.)

Because no individual's judgment is infallible and because the ability to make effective judgments varies widely from individual to individual and even in the same individual from time to time, the evaluators' judgment and objectivity can always be questioned in judgment sampling. This is not so in statistical sampling, which is based on the widely accepted theory of probability, because the sample is scientifically selected and evaluated. Certainly, the complaint that evaluators looked at only the worst cases would have no merit.

Using statistical sampling, a third party can repeat a study and expect to come to comparable conclusions about the characteristics of the universe being measured. Although the study results may be interpreted differently, there can be no question about the facts. Likewise, statistical samples can be combined and evaluated even if they were taken by different persons. Evaluators working at different locations can participate independently in the same study, and the results from the several locations can be combined to develop one estimate. Also, a study started by one evaluator can be continued by another without difficulty. Further, if evaluators decide to extend the sampling, they can do so easily and combine the results.

Statistical sampling provides a means of objectively determining the sample size in order to provide results having the precision required for the universe being examined and the question being answered. This approach usually results in a smaller sample, with resultant savings in time and money, than that found in judgment sampling. Because of the intuitive but incorrect belief that an adequate sample must always be a fixed percentage, say 5 or 10 percent, of the universe, oversampling occurs frequently in judgment sampling. However, if the universe is small, using the intuitive approach of selecting a sample that is equal to a fixed percentage of the universe could yield a sample too small to produce reliable results. For example, if the universe consisted of 200 items and a 10-percent sample were drawn, the sample size would be only 20 items.

Statistical sampling may sometimes be a more powerful method of discovering fraud or misuse of resources. After several reviews, an agency employee might be able to figure out the auditors' selection pattern if they used judgment sampling. The employee could then arrange files so that the auditors would not select documents containing evidence of fraud. However, in probability sampling, all documents have a certain probability of selection, and manipulating their location in the files will not affect this probability. Also, because statistical sampling results in selecting items from more files, agency employees may feel that the evaluators are making more thorough examinations and therefore may be less likely to continue the fraud or other abuse.

A particular GAO project may use a combination of samples. For example, runaway-youth centers might be chosen judgmentally, but within each center information could be sought from a random sample of youths who used a center's services. The most appropriate combination depends upon the project's objectives and constraints.

Determining the type of statistical sampling

If a statistical sample is the choice, a further decision must be made among the possible types of statistical sampling methods. Among the types that might be used, three common ones--simple random sampling, stratified sampling, and cluster sampling--are discussed later in this chapter. Two additional sampling types, discovery sampling and acceptance sampling, are relevant to some audits and are discussed in chapter 5.

Determining the sample size

The determination of an appropriate sample size is part of sample design. However, we do not treat sample size in the discussion on sample design for two reasons: (1) factors that must be considered in calculating sample size, confidence, and precision are not introduced until chapter 3 and (2) sample size also depends upon the estimation procedures, which are discussed in chapters 3 and 4.

To use this paper for guidance in determining sample size, evaluators should decide on the sampling method to be used, the estimation procedure, the precision required, and the confidence level desired. Reference to the appropriate section on calculating sample size in chapters 3 or 4 will then provide the necessary guidance.

SELECTION PROCEDURES

Selection procedures involve the method of actually picking the sampling units (sometimes called "drawing" the sample). All types of statistical samples use random selection procedures. The selection procedure may be dictated by the universe's arrangement, the evaluators' knowledge or "guesstimate" about how the sampling units are arranged within the universe, the proportion of the universe that will be drawn into the sample, or the method used to identify the sampling unit. Several selection procedures may be used for a single sample design.

Practical selection procedures are discussed in detail in chapter 6, but a short example will illustrate a procedure. Consider the runaway-youth program example again. Suppose we wish to use the simple random sampling design for selecting 50 center directors from a universe of 200. One procedure would be to write the name of one director on each ping-pong ball, one for each center in the universe, and put the 200 balls into an urn. The urn would be thoroughly shaken, and a person would draw 50 balls from the urn to form the sample. In this procedure, each ball, and therefore each director, would have an equally probable chance of being included in the sample. The procedure would be random, and the sample would conform to the requirements of the simple random sampling design.

ESTIMATION PROCEDURES

Estimation procedures, discussed in chapters 3 and 4, refer to the mathematical formulas used to calculate both the estimates of universe characteristics obtained from sampling and the precision of these estimates. The confidence level and sampling error estimates tell us how much reliance can be placed on the universe estimates and how precise they are, respectively. The various types of computation methods, such as manual calculations, with or without a calculator, or computer calculations, may be considered part of estimation procedures.

To briefly illustrate an estimation procedure, we can consider the runaway-youth example again. Suppose we want to use the information acquired from our sample of directors to estimate the total number of staff members employed by all the centers. If simple random sampling was used, the estimation procedures are easy.

Fifty of the center directors, or one fourth of the universe, were in our sample. If the 50 directors reported, collectively, that 287 staff members worked in their centers, then our best estimate of the total staff members for all centers would be 4 times 287, or 1,148. In this simple case, the universe estimate is just inversely proportional to the sampling fraction of one fourth.

The foregoing universe estimate will almost certainly be incorrect by some amount because of sampling error. However, by using the concepts of precision and confidence, we can also estimate the amount of error in the estimate. Procedures for doing so are discussed in chapters 3 and 4.

Data collection from a sample seldom proceeds exactly as planned. When we get nonresponses to questionnaires or when sample values are missing, special estimation techniques are required. Some of the interplay between sampling and data collection problems is discussed in chapter 7. In general, evaluators should consult with a specialist for advice on how to cope with data problems when making estimates.

SIMPLE RANDOM SAMPLING

Simple, or unrestricted, random sampling is the simplest method of drawing a statistical sample, and this design is basic to all others. The assumptions underlying the use of simple random sampling are that the population is homogeneous and is in one location, or it can be sampled from a single list of sampling units if it is in several locations, and that there is only moderate variation among the values of the items in the universe. Once the universe list has been developed, the sample can be drawn by using one of the selection procedures described in chapter 6 or appendix II. No attempt is made to segregate or separate any portion of the population into separate groups before the sample is selected. Thus each individual item in the universe has an equal probability of being included in the sample. This is the most common method of sampling but sometimes is less efficient than other methods.

An example of simple random sampling in GAO work is selecting a random sample of children participating in one school district's lunch program, in order to determine whether their family income meets that program's eligibility criteria.

STRATIFIED SAMPLING

Stratified sampling refers to the situation in which the universe is divided into two or more parts (strata) and a random sample is selected from each part (stratum). An estimate is determined separately for each stratum, and these are combined to form an estimate for the entire universe. A stratum is a subpopulation from the total population. The terms "high income," "middle income," and "low income" indicate three strata of a universe of people classified by the income they received. Invoices might be divided into two strata, one for those for \$1,000 or more and another stratum for those of less than \$1,000. A stratum in the universe of accounts receivable might be composed of all accounts with balances of \$1,000 or more. A stratified sample can be used to

--obtain equal precision with a smaller sample or greater reliability with a given sample. Stratification generally reduces the cost of a sample for a given precision;

--obtain separate estimates for the groups in the individual strata, if such estimates would be useful for comparison purposes; and

--give special emphasis to certain types of transactions, such as those of high dollar values or those with a great error potential.

Sometimes stratification is necessary because the universe is divided up among several locations and it is not possible to develop a single sampling frame. For example, the objective of the project may be to measure error in the pay of civilian employees of the Air Force working at several different air bases. If it is not possible to develop a single list of employees at all the air bases, a separate sample will have to be drawn at each air base, and estimates for each air base will have to be combined in order to obtain one overall estimate for the entire universe.

Stratification may be desirable if the costs of data collection differ from stratum to stratum. For example, in one stratum we may have to collect the data by personal interview but in another stratum we may be able to use mailed questionnaires.

When defining strata and setting stratum boundaries, evaluators should keep certain rules in mind. (1) Each sampling unit can be included in one, and only one, stratum. (2) The strata must not overlap. (3) The items in each stratum should be as much alike as possible in relation to the characteristic being measured.

Each stratum is considered a separate universe from which items are selected independently; that is, the sample selected in one stratum must not depend on, or be related to, the sample selected in another stratum. One of the acceptable random selection procedures is used to draw the sample in each stratum.

The total sample may be allocated to each stratum in proportion or in disproportion to the number of sampling units in that stratum. With proportional allocation, the sampling fraction (sample size divided by universe size) is the same in each stratum. With disproportional allocation, sampling fractions differ in two or more strata. Disproportional allocation may be based on professional judgment or on mathematical formulas, in order to minimize the overall sampling error or the overall cost of data collection. Appendix III discusses the allocation of sample size to strata as well as the calculation of estimates, sampling errors, and sample sizes with a stratified sample design.

CLUSTER SAMPLING

Another type of sampling is cluster sampling, which is the selection of groups of items (or clusters) rather than the selection of individual items directly. Examples of clusters are folders in filing cabinet drawers, baskets of produce, counties in a state, and the persons in a household. We are sometimes able to examine all the items within the sample cluster.

However, if the clusters are large, it is often preferable to select a random sample of items within the cluster. This is referred to as "two-stage" sampling.

Because of the size and complexity of some universes, cluster sampling must on occasion be done in three stages. We select first the primary sampling units, then the secondary sampling units within each primary unit, and finally the items within each secondary unit. An example of cluster sampling is given in appendix III.

CHAPTER 3

BASIC ESTIMATION PROCEDURES AND FURTHER SAMPLE DESIGN CONSIDERATIONS

This chapter introduces the concepts of precision, confidence, and sampling error. In the context of sample design, these concepts lead to a determination of the sample size appropriate for a particular evaluation or audit. Almost invariably, one of the first questions someone interested in sampling asks is, "How large a sample do I need to take?" Procedures for calculating sample size are presented in this chapter. Basic estimation procedures are presented for two situations: sampling for variables and sampling for attributes.

Sampling for variables is used when we are estimating something that can be quantified or measured in dollars, pounds, feet, and so on. This measurement is known as a variable. Some examples of variables are a person's weight, the tensile strength of wire, and the dollar error in an accounts-receivable balance.

When sampling for attributes, we want to determine how frequently items having a certain characteristic occur in a universe. The characteristic, or characteristics, that we are interested in is called the attribute of the item. Either the item has the characteristic or it does not, although a third, unknown value can be ascribed. Sometimes the characteristic is only one of several choices, as when people are classified by sex, race, educational level, or employment status. Examples of attributes for which we might sample are travel orders without proper approval, health insurance claims that were paid without supporting documentation, and farm loans that are unpaid.

We can use a single sample to develop estimates for both variables and attributes. For example, in examining purchase orders, we can take one sample to estimate both the rate of occurrence and the dollar amount of unjustified purchases. However, in general, the sample sizes required for estimating variables are larger than those required for estimating attributes. Therefore, when we calculate the sample size, we should base it on the precision we want to obtain for the variables estimate, not the attributes estimate.

THE CONCEPTS OF PRECISION AND CONFIDENCE

Specifying the precision needed for sample estimates is an important part of sample design. The precision is the amount of sampling error that can be tolerated but that will still permit the results to be useful. This is sometimes called "tolerable error" or the "bound on error."

Because precision is a way of expressing the amount of error that can be tolerated, it is related to the accounting concept of materiality or the evaluative concept of importance. The notion of materiality, according to a 1957 statement of the American Accounting Association, says that an item should be regarded as material if there is reason to believe that knowledge of it would influence an informed investor's decision. In policy or evaluation research, a result is considered important if there is reason to believe that knowledge of it would influence a decisionmaker's behavior or be important in public debate.

Importance and materiality are relative concepts rather than absolute. For example, a $100,000 overstatement of the assets of a company whose total assets are only $200,000 would be material. A $100,000 overstatement of the total assets of a multibillion dollar corporation would be immaterial. A 10-percent misstatement about the notes-receivable account of a small loan company would probably be material, but a 10-percent misstatement in the office supplies account balance of the same company would be immaterial. Since importance and materiality are relative, we need a basis for establishing whether a finding is important or material.

Materiality, or importance, is linked to precision in the following way. To develop a reasonable specification of precision, project managers must gauge the materiality or importance of the variables to be measured and use this information to decide how much the statistical estimates can vary from the true universe value and yet provide useful information. Going back to the example above, if we were attempting to verify the notes-receivable balance, we would probably be very unwilling to allow the estimate to vary from the actual amount by as much as 10 percent. Thus, we would probably want to take a large enough sample of individual loans and confirm the balances to keep the estimate well within 10 percent of the actual figure. However, if we were evaluating the office supplies account balance, we probably could live with a misstatement of almost 100 percent, if we bothered to consider the account at all.

In addition to specifying the precision of the estimate, the project manager must specify the reliability of the estimate. As used here, reliability means the probability that the estimate obtained from the sample is within the precision limits of the actual universe characteristic being estimated. Referred to as "confidence level," this is expressed as a percentage. It is the complement of the risk that the project manager is willing to take that the estimate misses the mark. (The concept of confidence is developed more fully in appendix I.) The confidence level should be determined by the importance of the results of the project taken as a whole; precision relates to the materiality or importance of individual estimates.

Project managers should decide on the confidence level right after they define the problem. The decision should not be postponed until after a sample has been taken and evaluated, in order to get a confidence level that makes the sampling error look smaller. Some examples of precision specifications are

- o The sampling error of the estimated total should be $600 at the 95-percent confidence level.
- o The precision of the estimated total overstatement of the accounts receivable balance should be $450,000 at the 90-percent level.
- o The average weight of men employed by GAO should be estimated within 5 pounds at the 95-percent confidence level.

General examples of the sample sizes used in GAO studies and an evaluation of their appropriateness appear in table 3.1.

Other, related considerations are the costs and time required to obtain the sample data. If the precision is specified "too tight," without considering cost and time, the sample size may be larger than is practical, given these considerations. Usually, only limited resources (of money, staff, etc.) and time are available for data collection. This must always be remembered when the estimate's precision is being specified.

Various mathematical formulas can take data collection costs into account for computing sample sizes and specifying precision. However, such formulas are beyond the scope of this paper. Perhaps the most practical guidance that can be given here is that the evaluator should specify the precision required, calculate the

Table 3.1

An Evaluation of Sample Sizes Used in GAO Studies

Description of estimate	Sample size	Estimate	Sampling error	Remarks
Error rate in military pay records at an air base	354	49%	5%	Sample probably too large, considering error rate found; with a sample of 100, sampling error would be 10%
Error rate in military pay records at another air base	355	1.4%	1.1%	Sample possibly too small, especially if base officials were claiming that the error rate was less than 1% and GAO was trying to show it was higher
Amount of discounts for prompt payments lost at a disbursing office	200	$163,000	$191,000	Sample too small; sampling error larger than estimate; however, if criterion is to lose no discounts, this certainly indicates a problem
Total dollar error in Medicare payments	100	$281,000	$142,000	Sampling error slightly more than half the estimate, but estimate does point to a problem; that is, even if $139,000 error is too much, sample is adequate
Number of aliens living abroad and receiving Social Security benefits	313	206,000	25,000	Considering magnitude of estimate, sample size probably adequate
Amount of overpayments in Black Lung Program	286	$44.6 million	$31.1 million	Although the sampling error is 70% of the estimate, the results indicate a problem that could cost at least $13.5 million; sample size may be adequate

sample size needed to achieve this precision, and then estimate the cost or time required to collect the data for the computed sample size. If the cost is more than can be afforded, or the required time is more than can be allowed, the precision should be relaxed (sampling error should be allowed to increase) until an affordable sample size is found. As we noted above, an adjustment like this should be made by relaxing the specified precision, not by manipulating the confidence level.

SAMPLING FOR VARIABLES

As we noted above, a variable is something that can be quantified, or measured in dollars, pounds, and the like. When sampling for variables, we usually want to estimate the total value for the universe of interest--for example, the total amount of assessed taxes that were not collected. For some reviews, however, the arithmetic mean (or arithmetic average) may be of more interest.

The first step in estimating the universe total is to compute the arithmetic mean of the sample items. The mean is simply the sum of the sample values divided by the sample size. The mean, a very important measure of central tendency, can be manipulated arithmetically, which is not true of some other measures of central tendency such as the median and the mode.

For certain types of data, the median is a better measure of central tendency than the mean. The median, the middle value or measurement of a set of values, is selected in a way such that half the values are below it and half are above. Thus, the median is a locational measure of central tendency. An example of the type of data for which the median might be a better measure of central tendency is salaries, or wages, for which there are generally a few extremely high values but the majority of values tend to concentrate in a rather narrow band at the lower end of the distribution.

An extremely simplified instance of sampling for variables might occur if our objective were to estimate the dollar amount of small purchases made by an agency during a specific fiscal year. The universe would then be defined as all small purchases during the fiscal year. During the year, there were perhaps 100 such purchases. (It is somewhat unusual to sample from such a small universe, although doing so may be necessary on occasion; we use the small universe here for its convenience as an example.)

In this paper, we will generally use uppercase letters to represent universe parameters and lowercase letters to represent sample statistics. If the universe parameter is estimated from a sample, it will be designated by an uppercase letter with the caret (usually referred to as a "hat") above it. Also, note that the formulas given in this chapter can be used only with simple, or unrestricted, random samples. They cannot be used with

Table 3.2

Work Sheet for Computing the Mean and Standard Deviation of Sample Data

Sample item (i)	Amount (y_i)	y_i^2	Sample item (i)	Amount (y_i)	y_i^2
1	$ 147	$ 21,609	16	$ 241	$ 58,081
2	259	67,081	17	232	53,824
3	185	34,225	18	233	54,289
4	164	26,896	19	205	42,025
5	150	22,500	20	226	51,076
6	187	34,969	21	236	55,696
7	137	18,769	22	202	40,804
8	159	25,281	23	89	7,921
9	125	15,625	24	248	61,504
10	172	29,584	25	160	25,600
11	277	76,729	26	194	37,636
12	142	20,164	27	177	31,329
13	231	53,361	28	135	18,225
14	125	15,625	29	96	9,216
15	172	29,584	30	163	26,569
				$5,469	$1,065,797

stratified samples and cluster samples because such samples are not drawn from a single, undivided universe.

We will use the symbol N for the universe size, which equals 100 in our example. Assume that we take a simple random sample of 30 items from this universe. We will let the symbol n represent the sample size, which is 30. For each sample purchase, we will research the file and record the dollar amount of the purchase. (The sample purchases and their amounts are shown in table 3.2.) Let y_i represent the amount for sample purchase number i, where i varies from 1 to the total sample size, n. Also let

$$\sum^{n} y_i$$

indicate the sum of all the values in the sample and

$$\sum^{N} y_i$$

indicate summation over a universe, where it is understood that i will vary from 1 to N. The formula for calculating the sample mean ($\bar{y}$) is

$$\bar{y} = \frac{\sum^{n} y_i}{n}$$

For the example above,

$$\bar{y} = \frac{5{,}469}{30}$$

$$\bar{y} = 182.30$$

Thus, the arithmetic mean of the sample purchases is $182.30.

In most situations, as noted above, we are more interested in the universe total than the mean. To estimate the universe total ($\hat{Y}$), we simply assume that the sample arithmetic mean is an estimate of the universe mean and multiply the sample mean by the number of items in the universe. This is called "expansion estimation." That is, $\hat{Y} = N\bar{y}$. For the example above,

$$\hat{Y} = (100)(182.30)$$

$$\hat{Y} = 18{,}230$$

Thus, the estimated total amount of small purchases made by the agency during the fiscal year in question is $18,230.

Does this adequately estimate the true total for the universe? How can we be sure? This depends on how good our assumption was that the sample mean is an estimate of the universe mean. To measure how good our estimated total is, we have to determine the precision of the estimate and the confidence level at which the precision is stated.

Calculating the sampling error

To compute the precision, or sampling error, of the estimated total, we first compute the standard deviation of the purchase amounts. The standard deviation is a numerical measure of the dispersion of a group of values about their mean. Understanding this statistic is a key to understanding much of sampling. It is a measure of the average squared deviation from the mean. The first step is to get the deviation of each item from the mean ($y_i - \bar{y}$). These items are first squared and then summed (see table 3.2). This result is then divided by n - 1.[1] Finally, the

[1]Note that we divide by n - 1 rather than the full sample size n. Stated simply, the reason for doing this is that we have used the sample mean to estimate the universe mean, which we do not know. The effect of this is to "use up" one of the sample values, leaving only n - 1 values as a basis for estimating the standard deviation. We lose one value (technically, one "degree of freedom") for every universe parameter, such as the mean, that we estimate from the sample. For sample sizes greater than 30, using n as the divisor makes a negligible difference in the results.

square root is taken. Engineers call this statistic the "root mean square," because it is the square root of a form of the average of the squared deviations. This statistic is always in the same unit of measurement as the variable itself. For example, the standard deviation of a variable measured in dollars is also expressed in dollars.

The standard deviation (s) is computed by the formula

$$s = \sqrt{\frac{\sum^{n} (y_i - \bar{y})^2}{n-1}}$$

Usually we compute the standard deviation by using the following short-cut formula, in which

$$\sum^{n} y_i^2$$

equals the sum of the squares of each of the sample values.

$$s = \sqrt{\frac{\sum^{n} y_i^2 - n\bar{y}^2}{n-1}}$$

A work sheet for computing the sum of the values and the sum of the squares of the values is illustrated in table 3.2. Putting values into the formula, we obtain

$$s = \sqrt{\frac{1{,}065{,}797 - (30)(182.30)^2}{30-1}}$$

$$s = 48.71$$

Thus, the sample standard deviation is $48.71.

The next step is to calculate the sampling error of the mean ($E_{\bar{y}}$) at a specified level of confidence. This is done by multiplying the standard deviation by a t value corresponding to the stipulated level of confidence and dividing by the square root of the sample size. (Appendix I contains a table of t factors for commonly used confidence levels.) The formula is

$$E_{\bar{y}} = \frac{ts}{\sqrt{n}}$$

Suppose that we have previously decided that the confidence level for the precision of our estimate should be 95 percent. The t factor for 95 percent is 1.96 (or 2 for all practical purposes). Using this value as well as the others, we obtain

$$E_{\bar{y}} = \frac{(1.96)(48.71)}{\sqrt{30}}$$

$$E_{\bar{y}} = 17.43$$

Thus, the sampling error of the mean is $17.43.

We mentioned earlier that in most sampling applications, we are interested in the estimated total. How do we compute the sampling error of the total? We computed the estimated total by multiplying the sample mean by the number of items in the universe. The computation of the sampling error of the total ($E_{\hat{Y}}$) is a parallel procedure. We simply multiply the sampling error of the mean by the number of items in the universe. The formula is $E_{\hat{Y}} = NE_{\bar{y}}$. For our example, we obtain

$$E_{\hat{Y}} = (100)(17.43)$$

$$E_{\hat{Y}} = 1{,}743$$

Thus, the sampling error of the total is $1,743.

The interpretation of this value parallels the interpretation of the sampling error of the mean. For practical purposes, using the 95-percent confidence level, we state that if all small purchase orders were reviewed, the chances are 19 in 20 that the results of a review would differ from the estimate obtained from the sample by less than the sampling error. Note that sampling errors are always stated at a certain confidence level. The best estimate, the point that is likely to be closest to the true population total, is $18,230.

When we sample for variables, a single sample may be used to develop estimates for many different variables. In principle, if different variables are to be estimated from the sample, the sampling errors should be adjusted upward to account for the increased exposure to possibly bad estimation. (See Dixon and Massey, 1969, and Snedecor and Cochran, 1980.)

<u>Calculating sample size</u>

Whether we sample for variables or sample for attributes, one advantage of statistical sampling is that it permits us to determine objectively the sample size required to achieve a given degree of precision at a specified confidence level. To make this computation, we need to estimate the standard deviation of the universe. It may seem paradoxical that we must obtain information about the universe when we are sampling in order to estimate its characteristics. However, when we look at other types of information-gathering, this is really not so strange. To look up a word's correct spelling in a dictionary, we must have some idea of how the word is spelled. To locate our position on a map, we must know "about where we are." And to compute a vessel's exact position by celestial navigation, we use an assumed position that has to be a fairly accurate estimate.

In computing the sample size, we must consider three factors. The evaluators specify two factors--the confidence level and precision. The third factor, the standard deviation, is based on the characteristics of the universe. A formula can bring these three factors together for computing the sample size. (The universe size is not taken into account, as we explain below.)

To use the formula in computing the sample size when sampling for variables, suppose we want to reduce the sampling error of the total small purchases in the example given above from $1,743 to $1,400 at the 95-percent confidence level. The first step is to convert the precision that is wanted (or tolerable error) of the total to the tolerable error of the mean (E). The computation is

$$E = \frac{\text{desired precision of estimated total}}{N}$$

In our example,

$$E = \frac{1{,}400}{100}$$

$$E = 14$$

Thus, the tolerable error of the mean is $14.

Once we have the tolerable error of the mean, we compute the required sample size by using the formula

$$n = \left(\frac{ts}{E}\right)^2$$

$$n = \left[\frac{(2)(48.71)}{14}\right]^2$$

$$n = 48.42 \quad \text{or} \quad 49 \text{ (rounding up)}$$

This means that 19 purchase orders, in addition to the first sample of 30, would have to be sampled in order to achieve the required precision.

Notice that we have used the standard deviation obtained from our first sample as an estimate of the true universe standard deviation. If the true standard deviation were known, we would not be sampling.

The best method of estimating the standard deviation is to take a small, random, preliminary sample and calculate the standard deviation from it. The sample should be random so that, if it must be increased to obtain the precision that is wanted, which usually happens, the preliminary sample can be included in the final sample and no work will have been wasted. The preliminary sample should consist of at least 30 cases; otherwise, the laws discussed in appendix I will not apply or will not work as well.

Sometimes it is possible to use the results of samples taken at other times as an estimate of the standard deviation. This is usually satisfactory if a similar review has been made and no major change in the distribution of the universe values is suspected. Another possibility is that subject matter experts may be able to guess the size of the standard deviation from their knowledge of the field or previous work.

Occasionally, it is incorrectly stated that a larger universe requires a larger sample or that the sample must always be a certain percentage of the universe. This is not true. As we noted in the formula for calculating sample size, the size of the universe does not enter into the calculations. The universe size and the sample size are slightly related, but before explaining this further, we need to discuss the concepts of sampling with replacement and sampling without replacement.

When we sample with replacement, an item selected for the sample is returned to the universe and can be selected again. Since the sample item is replaced, the universe from which the sample is drawn can be regarded as infinite. (In theory, when we sample with replacement, the entire sample could consist of the same item.) When we sample without replacement, an item selected for the sample is "used up" and cannot be selected again. Thus, each item can appear in the sample only once.

Sampling without replacement is used in GAO, except in special circumstances. Because we are gradually using up the universe, sampling becomes more efficient as we go along. If the sample size is large in relation to the universe size, we can use this efficiency to reduce both the sample size and the sampling error. We do this through the finite population correction (FPC) factor. Considering a practical matter, when we do manual calculations, we need use the FPC only when the sample size is greater than 5 percent of the universe.

To use the FPC to reduce the sampling error of the mean (and, by extension, the sampling error of the total), we multiply the sampling error by the factor

$$\sqrt{\frac{N-n}{N}}$$

to get the complete formula for the sampling error of the mean:

$$E_{\bar{y}} = \frac{ts}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

Using the data from our random sample of 30 purchase orders, drawn from the universe of 100 purchase orders, we have

$$E_{\bar{y}} = \frac{(1.96)(48.71)}{\sqrt{30}} \sqrt{\frac{100-30}{100}}$$

$$E_{\bar{y}} = 14.58$$

Thus, the "corrected" sampling error of the mean is \$14.58, and the "corrected" sampling error of the total is

$$E_{\hat{Y}} = NE_{\bar{y}}$$

$$E_{\hat{Y}} = (100)(14.58)$$

$$E_{\hat{Y}} = 1{,}458$$

By using the FPC, we can reduce the sampling error by about 16 percent (\$1,458 versus \$1,743 without the FPC).

To use the FPC to reduce the sample size, we first calculate the sample size. If it is greater than 5 percent of the universe, we enter this first estimate of the sample size (n_e) in the following formula to determine the final sample size (n_f):

$$n_f = \frac{n_e}{1 + \frac{n_e}{N}}$$

We calculated above that a sample size of 49 purchase orders would be required in order to reduce the sampling error of the total to \$1,400. Since 49 purchase orders obviously make up more than 5 percent of the universe, we calculate as follows:

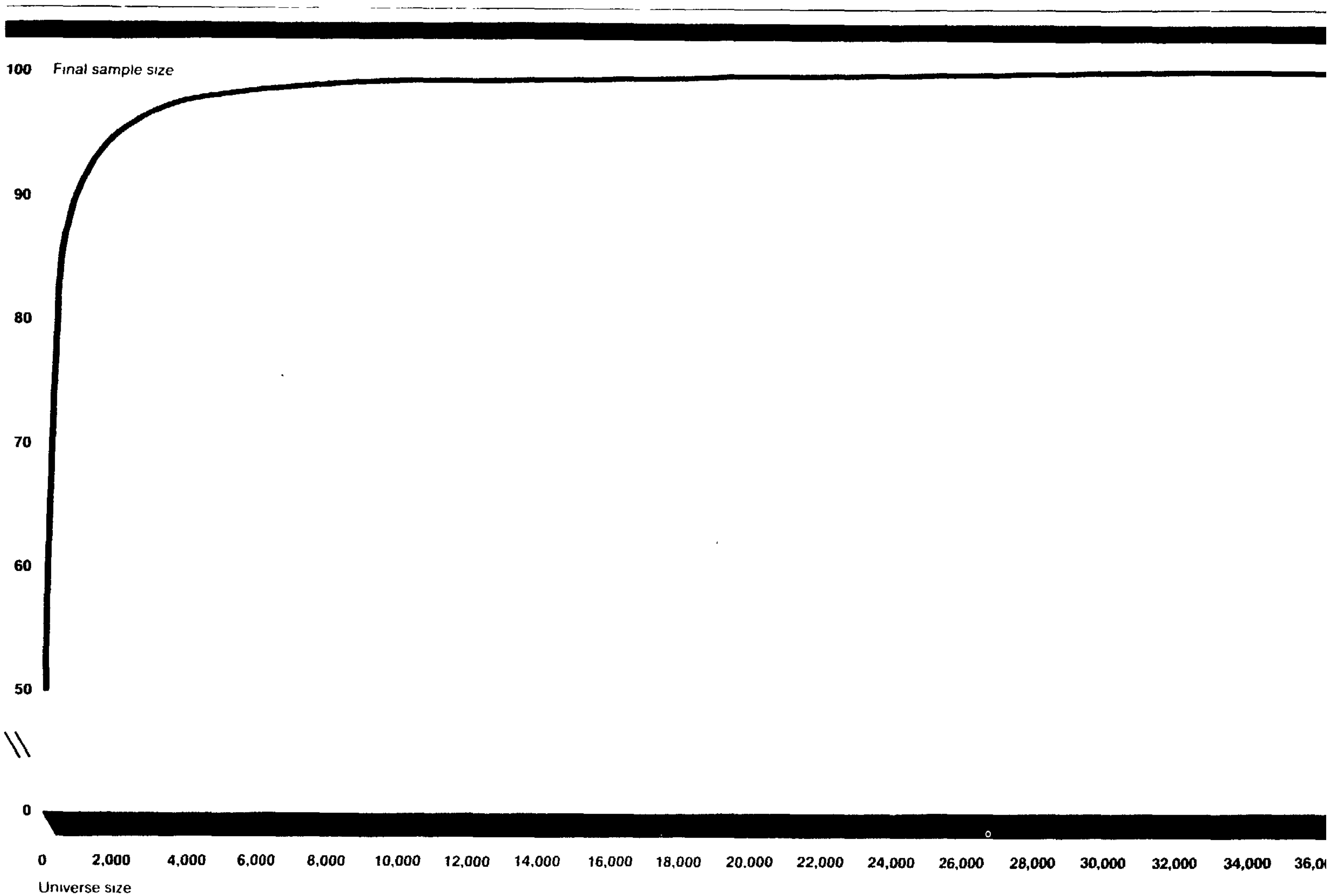
$$n_f = \frac{49}{1 + \frac{49}{100}}$$

$$n_f = 32.89 \text{ or } 33$$

Thus, the final sample size is 33 purchase orders. By using the FPC, we have reduced the sample size by about 33 percent. If the FPC is used to compute the required sample size, it should be used to compute the sampling error. Otherwise, the sampling error will be greater than the specified precision. If the sample size is 5 percent of the universe or less, the FPC can be ignored.

Figure 3.1 on the next page highlights the weak relationship between universe size and sample size, especially when the

Figure 3.1: Final Sample Size as Universe Size Increases: First Estimate of Sample Size Is 100



universe is large compared to the sample size. The figure shows the final sample sizes for various universe sizes, assuming that the first estimate of each sample size was 100.

Universe size	Final sample size	Universe size	Final sample size
100	50	4,000	98
200	67	6,000	98
400	80	8,000	99
600	86	10,000	99
800	89	20,000	100
1,000	91	40,000	100
2,000	95		

Note that if the first estimate of the sample size is less than 1 percent of the universe size, the final sample size equals the first estimate and remains constant.

SAMPLING FOR ATTRIBUTES

Sometimes we want to estimate the proportion, percentage, or total number of items in a universe that possess some characteristic or attribute or that fall into some defined classification. Examples are the percentage of the labor force that is unemployed, the percentage of people older than 65, and the number of low-income households in a county, and the like.

Assume that the evaluators are reviewing a supply depot's efficiency of operations. For this evaluation, they want to estimate the number of requisitions the depot was unable to fill during the past fiscal year because requisitioned items were out of stock and the rate at which the depot was unable to fill them. The universe consists of all 12,000 requisitions received by the depot that year.

Again, we use N to represent the size of the universe and n to represent the sample size. The sample consists of a simple random sample of 100 requisitions. Therefore, N equals 12,000 and n equals 100.

The characteristic of interest is, of course, a requisition that was not filled because the item was out of stock. We will let a represent the number of items in the sample that have the characteristic of interest. In this case, assume a equals 36. With these data, we use these formulas for calculating the estimated rate of occurrence (p) and the estimated number of occurrences ($\hat{A}$):

$$p = \frac{a}{n} \quad \text{and} \quad \hat{A} = Np$$

We calculate:

$$p = \frac{36}{100}$$

$$p = 0.36 \quad \text{or} \quad 36\%$$

Note that the computation of the estimated number of occurrences is identical to the computation of the estimated total when sampling for variables. We compute a rate of occurrence (analogous to the arithmetic mean) from the sample data, assume that it is an estimate of the rate of occurrence in the universe, and then multiply it by the universe size.

Given an estimated rate of occurrence of 36 percent for unfilled requisitions in our universe of 12,000, after multiplying the universe size times the rate of occurrence, we estimate that the number of unfilled requisitions is 4,320. Do these estimates of the rate of occurrence and the number of unfilled requisitions actually represent the true parameters in the universe? Can we specify a confidence level for these estimates? Can we calculate their precision? Yes. The laws governing large samples discussed in appendix I for variables also apply to attributes estimation.

Calculating the sampling error

To calculate the sampling error of the estimated rate of occurrence (E_p) and of the estimated number of occurrences ($E_{\hat{A}}$), we use the formulas below. Let q equal 1 - p.

$$E_p = t\sqrt{\frac{pq}{n}}$$

In the supply depot example, the sampling errors at the 90-percent and 95-percent confidence levels are calculated as follows:

90-percent confidence

$$E_p = 1.645\sqrt{\frac{(0.36)(0.64)}{100}}$$

$$E_{\hat{A}} = 947.52 \quad \text{or} \quad 948$$

95-percent confidence

$$E_p = 1.96\sqrt{\frac{(0.36)(0.64)}{100}}$$

$$E_{\hat{A}} = 1{,}128.96 \quad \text{or} \quad 1{,}129$$

In the manner that we calculated the estimated number of occurrences by multiplying N times p, we now calculate the sampling error of the number by multiplying the universe size by the sampling error of the rate.

Based on the results obtained above, we can say that the number of unfilled requisitions is within 948 of the 4,320 at the 90-percent confidence level; that is, the number of unfilled requisitions falls between 3,372 and 5,268 at the 90-percent confidence level, or the best estimate of the number of unfilled requisitions is 4,320 within a sampling error of 948 as stated at the 90-percent confidence level. Similar statements can be made regarding sampling errors calculated at the 95-percent confidence level.

In these computations, we have used the rate of occurrence found in the sample to represent the unknown rate of occurrence in the universe. If we increase the sample size to 400 items, the estimated rate of occurrence of unfilled requisitions should be about the same, but the sampling error at the 95-percent confidence level would be reduced to 4.7 percent, or 560 requisitions.

The unknown percentage we are trying to estimate is fixed, a constant; it does not move around. Only the estimates from different samples vary.

If a large number of samples were taken from the same universe, 68 percent of them would be within 1 sampling error of the percentage, about 95 percent would be within 2 sampling errors, and 99 percent would be within 2.58 sampling errors. At the 95-percent confidence level, we could state (rather crudely) that if all 12,000 requisitions in the universe were examined, the chances are 19 in 20 that the results would differ from the estimate obtained from the sample by less than the sampling error.

Calculating sample size

The concepts and formula used in calculating sample sizes for attributes sampling are the same as those for variables sampling, discussed above, except that $\sqrt{pq}$ can be substituted for the standard deviation, as shown below. Let E equal the tolerable error of the proportion.

$$n = \left(\frac{t\sqrt{pq}}{E}\right)^2 \quad \text{or} \quad n = \frac{t^2pq}{E^2}$$

We can give an example of the use of this formula. Suppose we wanted to reduce the sampling error of the percentage of unfilled requisitions to 4 percentage points at the 95-percent confidence level. Since E equals 0.04, we have

$$n = \frac{(2)^2(0.36)(0.64)}{(0.04)^2}$$

$$n = 576$$

Thus, we would have to sample 476 requisitions in addition to the 100 already sampled.

To use the formula, we need some "estimate" of the expected rate of occurrence of the characteristic of interest. This may be obtained from a preliminary random sample of 30 to 50 cases, from prior experience in a similar review, from experts in the field, or from information supplied by the agency being evaluated. If the evaluators suspected that the agency's estimate was too low, they could increase it by 10 to 20 percentage points. This would give a larger sample size. The largest sample sizes are needed when the percentage is around 50 percent. Smaller sample sizes can be used to obtain the same precision when the estimated value moves away from 50 percent in either direction. When no other estimate is available, the sample size can be estimated with a 50-percent rate.

With attributes sampling, just as with variables sampling, if the sample size is more than 5 percent of the universe, we should use the FPC. The formula is

$$E_p = t\sqrt{\left(\frac{pq}{n}\right)\left(\frac{N-n}{N}\right)}$$

In an example, we can use the FPC to reduce the sampling error. Suppose the supply depot received only 1,200 requisitions during the fiscal year. The sampling error of the rate of occurrence of unfilled requisitions would be 9.01 percent. The sampling error of the estimated number of unfilled requisitions would be 108. The computation is shown below.

$$E_p = 1.96\sqrt{\left[\frac{(0.36)(0.64)}{100}\right]\left(\frac{1{,}200-100}{1{,}200}\right)}$$

$$E_p = 0.09008 \quad \text{or} \quad 9.01\%$$

$$E_{\hat{A}} = (1{,}200)(0.09008)$$

$$E_{\hat{A}} = 108.10 \quad \text{or} \quad 108$$

We can also use the FPC to reduce the sample size. As we noted above, we needed a sample size of 576 requisitions to reduce the sampling error of the percentage to 4 percentage points. If the universe were only 1,200 requisitions, we could do as follows:

$$n_f = \frac{n_e}{1+\frac{n_e}{N}}$$

Thus, if the 576 became the estimated sample size and the above formula computed, an additional 290 requisitions would have to be sampled.

As with variables sampling, if the FPC is used to compute the sample size, it should be used to compute the sampling error. Otherwise, the computed sampling error will be greater than the specified tolerable error.

A DISTINCTION BETWEEN PRECISION AND ACCURACY

We use the word "precision" rather than "accuracy" throughout this chapter. Precision refers to the maximum amount, stated at a certain confidence level, that we can expect the estimate from a single sample to deviate from the results obtained by applying the same measuring procedures to all the items in the universe. Accuracy refers to the difference between the mean of the universe from which the sample is selected and the true characteristic that we intend to measure.

We will use a simple example to illustrate the distinction. Suppose we want to estimate the average weight of all men employed in GAO today. We could select a sample of the men, weigh them, and compute their average weight. If the sample were large enough, we could estimate the average weight very precisely. But if the list of male employees from which we drew the sample were a year old, we would not be estimating the average weight today; we would be estimating the average weight of men employed a year ago. The problem is that the universe from which the sample was selected is different from the universe we defined. Thus, the estimate, regardless of how precise it might be, would be inaccurate.

Then suppose we are able to select a sample of 100 male employees from an up-to-date list of all male employees and weigh them on a single scale. If we find that the average weight is 170 pounds with a standard deviation of 34 pounds, this estimate would have a precision of about 6.7 pounds at the 95-percent confidence level. If we wanted the estimate to be more precise, all we would have to do is increase the sample size. A sample of 400 would give a precision of about 3.3 pounds at the 95-percent confidence level. Or, if we weighed all the men, the result would be perfectly precise; there would be no sampling error at all, because this sample size is equal to the universe size. However, if the scale were 1 pound off, and we did not know this, the results, regardless of the sample size and the degree of precision, would not be accurate. The average weight we computed would be 1 pound less (or more) than the true average weight. Because it is so difficult to ensure that no unsuspected bias (or inaccuracy) enters into the estimate, we speak of the precision of the estimate rather than the accuracy.

CHAPTER 4

ADVANCED ESTIMATION PROCEDURES

Chapter 3 discussed expansion estimation, in which we calculate a sample mean or proportion and multiply it by the universe size to obtain the estimated total or the estimated number of occurrences in the universe. In this chapter, we describe ratio, regression, and difference estimation, which take into account other information that we can obtain from the samples and that we may obtain about the universe. These three procedures permit evaluators to develop more information from the sample data and frequently yield more efficient (or precise) estimates. However, the possibility of using these procedures must be considered before the data are collected. Otherwise, the benefits may be lost or it may be necessary to return to the location where the data were collected.

RATIO ESTIMATION

Sometimes, in a sampling application, the summary statistic we want to estimate is a ratio between two variables, both of which can vary from sampling unit to sampling unit. For example, we may want to estimate the ratio of costs of replacement parts sold under foreign military sales agreements to the amounts received or the ratio of Medicare reimbursements for prescription drugs to total reimbursements.

In other applications, we may want to estimate the total value of an unknown variable that is related to another variable for which we already know the universe total value. For example, we may want to estimate the total subsistence cost claimed on an agency's travel vouchers for a year, when we already know the total amount of travel reimbursement (the universe total) for the year and the number of vouchers paid (the universe size).

In this case, we select a random sample of vouchers and record, for each voucher, the amount paid, which corresponds to the universe total we already know and is referred to as the auxiliary (X) variable, and the amount of the claimed subsistence cost, which corresponds to the total we want to estimate and is referred to as the primary (Y) variable. Thus, we record two variables for each sample travel voucher. The summary statistic is the ratio in which the total of the primary variables (the subsistence cost) is the numerator and the total of the auxiliary variables (the amount paid) is the denominator.

We should also use ratio estimation when we suspect that there is a positive correlation between the two variables, even if we are interested not in the ratio but only in the estimated total of the primary variable. If the correlation is positive and if it is strong enough, the estimate that can be obtained will be more precise than the estimate that can be obtained with simple expansion estimation.

The ratio estimate has a slight mathematical bias that can usually be ignored if the sample size is large, say 50 or more. The bias results from the assumption that the line representing the ratio is a straight line passing through the origin. This almost never happens in practice.

Ratio estimation can be used with more complex sample designs such as stratified samples and single-stage and multistage cluster samples. However, the use of ratio estimation with complex sample designs is beyond our present scope.

Calculating the estimates

For an example of ratio estimation, consider this situation. An agency has made 10,000 small purchases totaling \$5,100,000 over 1 year. The evaluators, suspecting that purchases could have been made at a lower cost, decide to estimate the savings that would have resulted if suppliers offering lower costs had been used. The confidence level is set at 95 percent, and a preliminary random sample of 50 cases is taken. The results of the preliminary sample are shown in the first three columns of the work sheet in table 4.1 on the next page. The total for the second column is the sum of the purchase costs found in the sample, and the total for the third column is the sum of the savings found in the sample.

The first step is to compute the ratio of savings to agency purchase costs (R) by dividing the sum of the savings in the sample (Σx_i) by the sum of the agency purchase costs in the sample (Σy_i) :

$$R = \frac{\Sigma y_i}{\Sigma x_i}$$

$$R = \frac{3{,}600}{24{,}800}$$

$$R = 0.14516$$

The estimated ratio of savings to purchase costs can be expressed as 0.145, or 14.5 cents on the dollar, or 14.5 percent.

The next step is to estimate the total savings (Y) by multiplying the total purchase costs (X) of \$5,100,000 by 0.145, the ratio of savings to costs.

$$\hat{Y} = XR$$

$$\hat{Y} = (5{,}100{,}000)(0.14516)$$

$$\hat{Y} = 740{,}316$$

Note that with ratio estimation, we obtain two estimated results, ratio and total. For many purposes, the ratio is more

Table 4.1

Work Sheet for Computing the Sampling Error of Ratio for 50 Small Purchases

Sample item (i)	Agency cost (x_i)	Saving (y_i)	x_i^2	y_i^2	x_iy_i
1	$ 300	$ 33	$ 90,000	$ 1,089	$ 9,900
2	900	126	810,000	15,876	113,400
3	300	45	90,000	2,025	13,500
4	200	26	40,000	676	5,200
5	900	90	810,000	8,100	81,000
6	700	140	490,000	19,600	98,000
7	1,000	180	1,000,000	32,400	180,000
8	100	20	10,000	400	2,000
9	900	135	810,000	18,225	121,500
10	700	70	490,000	4,900	49,000
11	700	70	490,000	4,900	49,000
12	400	68	160,000	4,624	27,200
13	300	45	90,000	2,025	13,500
14	100	16	10,000	256	1,600
15	200	32	40,000	1,024	6,400
16	100	12	10,000	144	1,200
17	600	72	360,000	5,184	43,200
18	400	60	160,000	3,600	24,000
19	900	153	810,000	23,409	137,700
20	1,000	200	1,000,000	40,000	200,000
21	1,000	138	1,000,000	19,044	138,000
22	600	96	360,000	9,216	57,600
23	800	152	640,000	23,104	121,600
24	200	24	40,000	576	4,800
25	200	28	40,000	784	5,600
26	1,000	110	1,000,000	12,100	110,000
27	900	108	810,000	11,664	97,200
28	600	60	360,000	3,600	36,000
29	500	75	250,000	5,625	37,500
30	200	36	40,000	1,296	7,200
31	500	75	250,000	5,625	37,500
32	200	28	40,000	784	5,600
33	200	22	40,000	484	4,400
34	500	80	250,000	6,400	40,000
35	500	100	250,000	10,000	50,000
36	400	76	160,000	5,776	30,400
37	200	40	40,000	1,600	8,000
38	600	60	360,000	3,600	36,000
39	500	75	250,000	5,625	37,500
40	300	36	90,000	1,296	10,800
41	900	135	810,000	18,225	121,500
42	100	16	10,000	256	1,600
43	100	15	10,000	225	1,500
44	900	90	810,000	8,100	81,000
45	300	60	90,000	3,600	18,000
46	500	85	250,000	7,225	42,500
47	500	75	250,000	5,625	37,500
48	300	63	90,000	1,089	9,900
49	500	65	250,000	4,225	32,500
50	100	14	10,000	196	1,400
Total	$24,800	$3,600	$16,620,000	$365,422	$2,400,400

meaningful than the total, because it permits us to make such statements as 14.5 cents of every dollar could have been saved had purchases cost less.

Computing the sampling error

The formulas for computing the sampling errors of the estimated ratio and total are more complicated for ratio estimation than for expansion estimation. Before computing the ratio's sampling error, we must first introduce the concept of the standard deviation of the difference variate (S_d). We let d_i equal the difference between the observed value of the primary variable for sampling unit number i and the predicted value that would be obtained if the value of the auxiliary variable for sampling unit number i were multiplied by the estimated ratio. For example, in the case of sample purchase number 1:

$$d_i = y_i - Rx_i$$

$$d_1 = y_1 - Rx_1$$

$$d_1 = 33 - (0.14516)(300)$$

$$d_1 = -10.548$$

Then the standard deviation of the difference variate is given by the formula

$$S_d = \sqrt{\frac{\sum (y_i - Rx_i)^2}{n-1}}$$

The standard deviation of the difference variate can be computed by using the following short-cut formula:

$$S_d = \sqrt{\frac{\sum y_i^2 - 2R\sum x_i y_i + R^2 \sum x_i^2}{n-1}}$$

Most of these symbols are familiar from earlier chapters. However, it should be noted that

$$\sum x_i y_i$$

equals the sum of the cross products. This means that for each sample unit, the agency purchase cost is multiplied by the savings, and the sum of all the products is obtained. For ease of computation, it is best to prepare a work sheet like the one in table 4.1.

Substituting the sums of the squares, cross products, and other values into the formula, we obtain

$$S_d = \sqrt{\frac{365{,}422 - (2)(0.14516)(2{,}400{,}400) + (0.14516)^2(16{,}620{,}000)}{50-1}}$$

$$S_d = 19.56$$

That is, the standard deviation of the difference variate is $19.56. This value is used in the formula for computing the sampling error of the ratio (E_R):

$$E_R = \frac{tS_d}{\bar{x}\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

Since the sample size is less than 5 percent of the universe, the finite population correction expressed as

$$\sqrt{\frac{N-n}{N}}$$

can be ignored in this example. Thus, the formula can be simplified to

$$E_R = \frac{tS_d}{\overline{X}\sqrt{n}}$$

Most of these symbols are familiar from earlier chapters. However, $\overline{X}$ equals the universe mean obtained by dividing the universe total (X) by the universe size (N).

$$\overline{X} = \frac{5{,}100{,}000}{10{,}000} = 510$$

$$E_R = \frac{(1.96)(19.56)}{(510)\sqrt{50}}$$

$$E_R = 0.01063$$

Thus, the sampling error of the ratio is 0.01063 at the 95-percent confidence level.

The computation of the sampling error of the total ($E_{\hat{Y}}$) obtained by ratio estimation parallels the computation of the estimated total. We merely multiply the known universe total of the auxiliary variable by the ratio's sampling error. For our example, this would be

$$E_{\hat{Y}} = XE_R$$

$$E_{\hat{Y}} = (5{,}100{,}000)(0.01063)$$

$$E_{\hat{Y}} = 54{,}213$$

Thus, the estimated total is $740,316, plus or minus $54,213 at the 95-percent confidence level.

If we are estimating only the ratio, the universe total for the auxiliary variable need not be known. We merely substitute the sample mean ($\overline{x}$) for the population mean ($\overline{X}$) in the sampling-error formula. This does make a slight change in the ratio's sampling error.

If we had calculated the total and the sampling error of the total by expansion estimation, we would have obtained an estimated total of $720,000 and a sampling error of $129,057. This is approximately 2.4 times the sampling error obtained by using ratio estimation. Ratio estimation yields a smaller sampling error because the positive correlation between the auxiliary and primary variables reduces the sampling error.

Computing the sample size

Suppose we want to improve the precision of our estimate. Is there a formula to compute sample size for a desired level of precision? The answer is, Yes. The formula for computing sample sizes with expansion estimation can be adapted to ratio estimation. In place of the standard deviation of the values, we substitute the standard deviation of the difference variate, described in the section above.

To use the standard deviation of the difference variate in the formula for sample size, the evaluators must specify the confidence level and the precision that is wanted (tolerable error) of either the estimated total or the ratio. Assume that they specified a precision of $30,000 for the total. Then the desired precision of the mean is obtained by the following formula:

$$E = \frac{\text{desired precision of the estimated total}}{N}$$

$$E = \frac{30{,}000}{10{,}000}$$

$$E = 3$$

Once we have the value for E, we use the formula

$$n = \left(\frac{tS_d}{E}\right)^2$$

Substituting the values into the formula, we obtain

$$n = \left[\frac{(2)(19.56)}{3}\right]^2$$

$$n = 170$$

(Note that 2 was used as the value for t, for ease of computation.)

Since we already have taken a sample of 50, an additional 120 small purchases would have to be sampled. If the first estimate of the sample size were more than 5 percent of the universe, the finite population correction would be applied.

If instead of specifying a precision for the total, the evaluators had specified a precision for the ratio, the following formula would be used:

$$n = \left(\frac{tS_d}{\overline{X}E}\right)^2$$

The only difference between these formulas is the inclusion of the universe mean ($\overline{X}$) of the auxiliary variable. If the universe mean is not known, the sample mean ($\overline{x}$) of the auxiliary variable is used instead.

Assume the evaluators specified that the precision of the ratio should be 0.01. Thus, E equals 0.01. Substituting the required values in the formula given above, we have

$$n = \left[\frac{(2)(19.56)}{(510)(0.01)}\right]^2$$

$$n = 58.84 \quad \text{or} \quad 59$$

In addition to the 50 in our preliminary sample, 9 more small purchases would have to be sampled.

REGRESSION ESTIMATION

Like ratio estimation, regression estimation is an attempt to increase precision by using additional information that we know about the universe and can obtain from our sample. We obtain two measurements, one of the primary variable and the other of the auxiliary variable, on a single sampling unit. (The primary variable and auxiliary variable are defined exactly the same as they were with ratio estimation.) For regression estimation, we must know the universe total for the auxiliary variable. In this technique, we use the regression model, which is well known from data analysis, to make statistical estimates.

Like ratio estimation, regression estimation is subject to a slight mathematical bias that can usually be ignored if the sample size is 50 or more. The bias results from the assumption that the regression line is straight. Regression estimation can also be used with more complex sample designs.

Calculating the estimates

Going back to the example at the beginning of the chapter, let us develop the regression estimate of the total. We use the following formula, in which we let b represent the coefficient of regression predicting y from x, and $\hat{Y}_{lr}$ represents the total for the primary variable estimated by linear regression (the coefficient of regression is simply a mathematical way of calculating the amount of change in one variable when the amount of a related variable is changed): $\hat{Y}_{lr} = N[\overline{y} + b(\overline{X} - \overline{x})]$.

To compute the regression coefficient b, we use the following formula:

$$b = \frac{\Sigma x_i y_i - n\overline{x}\overline{y}}{\Sigma x_i^2 - n\overline{x}^2}$$

In this formula, $\overline{x}$ is the sample mean of the savings, and $\overline{y}$ is the sample mean of the purchase cost. $\overline{x}$ = 24,800/50, or 496, and $\overline{y}$ = 3,600/50, or 72. Substituting into the formula the values from table 3.1 and the two means we calculated above gives the following:

$$b = \frac{2{,}400{,}400 - (50)(496)(72)}{16{,}620{,}000 - (50)(496)^2}$$

$$b = 0.142341$$

When the value for the regression coefficient, 0.142341, and the values for the other symbols are substituted in the formula, we obtain

$$\hat{Y} = (10{,}000)\,[72 + (0.142341)(510 - 496)]$$

$$\hat{Y} = 739{,}928$$

Thus, the estimated total savings is $739,928. If we wanted to obtain the mean savings by using regression estimation ($\overline{y}_{lr}$), we would use a slightly simpler formula: $\overline{y}_{lr} = \overline{y} + b(\overline{X} - \overline{x})$. This formula gives $73.99 for the mean.

Note that with regression estimation, as with ratio estimation, we get two results, the estimated total and the regression coefficient. However, the latter is not the same as a ratio. The regression coefficient measures the change in the primary variable that results from a unit change in the auxiliary variable; the ratio measures the proportional relationship between the sum of the primary variables and the sum of the auxiliary variables.

In our example, we would interpret the regression coefficient as follows. For every $1 increase (or decrease) in the agency purchase cost, the available savings increase (or decrease) by 14.23 cents. If the proportional relationship between the primary variable and the auxiliary variable is needed, the ratio estimate must be used.

Computing the sampling error

The sampling error of the regression estimate is computed by the formula below. Let r represent the coefficient of correlation between the primary and auxiliary variables.

$$E_{\hat{Y}} = \frac{N t S_y}{\sqrt{n}} \sqrt{\frac{N-n}{N}} \sqrt{\frac{(n-1)}{(n-2)}(1 - r^2)}$$

Since the sample size is less than 5 percent of the universe, the finite population correction

$$\sqrt{\frac{N-n}{N}}$$

can be eliminated and the formula can be simplified to

$$E_{\hat{Y}} = \frac{NtS_y}{\sqrt{n}}\sqrt{\frac{(n-1)}{(n-2)}(1-r^2)}$$

To compute the coefficient of correlation, we use the formula

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}}$$

Substituting the appropriate values into the formula, we obtain

$$r = \frac{2{,}400{,}400 - (50)(496)(72)}{\sqrt{[16{,}620{,}000 - (50)(496)^2][365{,}422 - (50)(72)^2]}}$$

$$r = 0.907664$$

When the value of r (0.907664), the standard deviation of the savings calculated using the formula in chapter 3, and the other values are substituted in the sampling error formula, we obtain

$$E_{\hat{Y}} = \frac{(10{,}000)(1.96)(46.560)}{\sqrt{50}}\sqrt{\frac{(50-1)}{(50-2)}[1-(0.907664)^2]}$$

$$E_{\hat{Y}} = 54{,}726$$

The sampling error of the total is $54,726 at the 95-percent confidence level. This is slightly larger than the sampling error obtained by the ratio estimate, partially because of the use of the correction factor (n - 1)/(n - 2). The correction factor is introduced because both the mean and the regression coefficient were calculated from sample data. This leaves only n - 2 observations for measuring the variation in the universe. With a larger sample, the sampling error of the regression estimate would have been slightly smaller than that of the ratio estimate, which almost always happens.

If the sampling error of the mean is wanted, we simply divide the sampling error of the total ($54,726) by the universe size (10,000), and we obtain $5.47.

Computing the sample size

To compute sample sizes with regression estimates, we can use the same formula that we used with expansion estimation, except that we need the standard deviation of the difference variate. The difference variate is the difference between the observed value of the primary variable and the value predicted from the regression equation. That is, $d_i = y_i - \bar{y} - b(x_i - \bar{x})$.

The standard deviation of the difference variate is given by the formula

$$S_d = \sqrt{\frac{\sum [y_i - \bar{y} - b(x_i - \bar{x})]^2}{n-1}}$$

The standard deviation of the difference variate can also be computed by the following formula:

$$S_d = S_y\sqrt{1-r^2}$$

It is readily apparent that both S_y and $\sqrt{1-r^2}$ are parts of the formula for computing the sampling error of the estimate.

Thus, the standard deviation of the difference variate is

$$S_d = (46.560)\sqrt{1-(0.907664)^2}$$
$$S_d = 19.54$$

Assume that the evaluators specify a precision of $40,000 for the estimated total at the 95-percent confidence level. First, we compute the precision of the mean.

$$E = \frac{\text{desired precision of the estimated total}}{N}$$
$$E = \frac{40,000}{10,000}$$
$$E = 4$$

Substituting this into the sample size formula, we obtain

$$n = \left(\frac{tS_d}{E}\right)^2$$
$$n = \left[\frac{(2)(19.54)}{4}\right]^2$$
$$n = 95.45 \quad \text{or} \quad 96$$

Thus, 46 small purchases, in addition to the original 50, will have to be examined.

If the final sample size were more than 5 percent of the universe, the finite population correction would be applied.

DIFFERENCE ESTIMATION

Difference estimation is used when we want to obtain a "corrected" estimate of a previously stated "book" value. For example, suppose we wanted to estimate the "correct" total value of an inventory when we know the value per book and can take a sample from the inventory items and correct the items examined in the sample, if necessary. It is also an attempt to increase precision by obtaining two measurements on a single sampling unit. However, difference estimation will increase precision only if the differences between the primary and auxiliary variables are very small. To use difference estimation, we must know the universe total for the auxiliary variable. (Primary and auxiliary variables are defined the same as with ratio estimation.)

Calculating the estimates

Going back to our example of the small purchases, assume that the evaluators decide to audit the accuracy of the payments that were made. To save work, they decide to use the same sample of 50 purchases. The purchase costs, the audited amounts that should have been paid, and the differences are shown in table 4.2. Note that when we compute the difference, the amount paid is considered the base and that we subtract it from the audited amount.

The mean difference ($\bar{d}$) equals the net difference divided by the sample size. The net difference is the algebraic sum of the fourth column of table 4.2.

$$\bar{d} = \frac{\Sigma d_i}{n}$$

$$\bar{d} = +\frac{1{,}048}{50}$$

$$\bar{d} = +20.96$$

The mean difference, \$20.96, is then multiplied by the universe, or 10,000, to obtain the estimated total difference ($\hat{D}$).

$$\hat{D} = N\bar{d}$$

$$\hat{D} = (10{,}000)(20.96)$$

$$\hat{D} = +209{,}600$$

The estimated total difference, \$209,600, is then added to or

Table 4.2

Audited Results of the Preliminary Sample of 50 Small Purchases

Sample item (i)	Amount paid (x_i)	Audited amount (y_i)	Difference (d_i)
1	$ 300	$ 300	$ 0
2	900	900	0
3	300	340	+40
4	200	200	0
5	900	900	0
6	700	700	0
7	1,000	1,169	+169
8	100	100	0
9	900	1,013	+113
10	700	700	0
11	700	700	0
12	400	560	+160
13	300	300	0
14	100	100	0
15	200	200	0
16	100	100	0
17	600	673	+73
18	400	400	0
19	900	900	0
20	1,000	1,141	+141
21	1,000	1,000	0
22	600	600	0
23	800	800	0
24	200	200	0
25	200	200	0
26	1,000	1,145	+145
27	900	900	0
28	600	600	0
29	500	735	+235
30	200	200	0
31	500	500	0
32	200	200	0
33	200	200	0
34	500	538	+38
35	500	500	0
36	400	400	0
37	200	200	0
38	600	587	-13
39	500	500	0
40	300	211	-89
41	900	900	0
42	100	100	0
43	100	100	0
44	900	900	0
45	300	336	+36
46	500	500	0
47	500	500	0
48	300	300	0
49	500	500	0
50	100	100	0
Total	$24,800	$25,848	+$1,048

subtracted from the universe total of the auxiliary variable in order to obtain the estimated corrected total payments. In this example, the estimated total difference is positive, so it is added to the universe total, or $5,100,000:

$$\hat{Y} = X + \hat{D}$$

$$\hat{Y} = 5{,}100{,}000 + 209{,}600$$

$$\hat{Y} = 5{,}309{,}600$$

Thus, using the difference method, we can estimate the correct amount that should have been paid at $5,309,600.

Table 4.3

Differences and Squares of Differences

Sample item (i)	Difference (d_i)	(d_i^2)	Sample item (i)	Difference (d_i)	(d_i^2)
1	0	0	26	+145	21,025
2	0	0	27	0	0
3	+40	1,600	28	0	0
4	0	0	29	+235	55,225
5	0	0	30	0	0
6	0	0	31	0	0
7	+169	28,561	32	0	0
8	0	0	33	0	0
9	+113	12,769	34	+38	1,444
10	0	0	35	0	0
11	0	0	36	0	0
12	+160	25,600	37	0	0
13	0	0	38	-13	169
14	0	0	39	0	0
15	0	0	40	-89	7,921
16	0	0	41	0	0
17	+73	5,329	42	0	0
18	0	0	43	0	0
19	0	0	44	0	0
20	+141	19,881	45	+36	1,296
21	0	0	46	0	0
22	0	0	47	0	0
23	0	0	48	0	0
24	0	0	49	0	0
25	0	0	50	0	0
				+1,048	180,820

Computing the sampling error

To compute the sampling error of the difference estimate, we first compute the standard deviation of the difference variate (S_d), as shown in the formula below. The first step in computing the standard deviation of the difference variate is to square the difference for each sample item and then obtain the sum of the squares, as shown in table 4.3. Let

$$\sum d_i^2$$

equal the sum of the squares of the differences.

$$S_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}$$

Substituting the sum of the squares from table 4.3 and other values into the formula, we obtain $56.94.

$$S_d = \sqrt{\frac{180,820 - (50)(20.96)^2}{50-1}}$$

$$S_d = 56.94$$

The general formula for computing the sampling error of the mean difference $E_{\bar{d}}$ is

$$E_{\bar{d}} = \frac{tS_d}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

Since the sample size is less than 5 percent of the universe, we can drop the finite population correction

$$\sqrt{\frac{N-n}{N}}$$

and simplify the formula to

$$E_{\bar{d}} = \frac{tS_d}{\sqrt{n}}$$

The value of t is 1.96, obtained from appendix I. Substituting the appropriate values into the formula, we compute

$$E_{\bar{d}} = \frac{(1.96)(56.94)}{\sqrt{50}}$$

$$E_{\bar{d}} = 15.78$$

Thus, the sampling error of the mean difference is $15.78. The sampling error of the total difference ($E_{\hat{D}}$) is obtained by multiplying the sampling error of the mean difference by the universe size.

$$E_{\hat{D}} = NE_{\bar{d}}$$

$$E_{\hat{D}} = (10{,}000)(15.78)$$

$$E_{\hat{D}} = 157{,}800$$

Since the total difference was applied to the amount paid in order to estimate the amount that should have been paid, the sampling error of the total difference is also the sampling error of the estimated amount that should have been paid. Thus, the amount that should have been paid is $5,309,600, plus or minus $157,800 at the 95-percent confidence level.

Computing the sample size

To compute the sample sizes for difference estimation, using simple random sampling, we first specify the confidence level and the tolerable error of the estimated total difference. Then we compute the tolerable error of the mean difference. Assume that for our small purchases example we have specified a precision of the total difference of $100,000 at the 95-percent confidence level.

$$E = \frac{\text{desired precision of estimated total difference}}{N}$$

$$E = \frac{100{,}000}{10{,}000}$$

$$E = 10$$

We compute

$$n = \left(\frac{tS_d}{E}\right)^2$$

$$n = \left[\frac{(2)(56.94)}{10}\right]^2$$

$$n = 129.69 \quad \text{or} \quad 130$$

The required sample size is 130. Besides our preliminary sample of 50, an additional 80 small purchases would have to be audited. (We would use the finite population correction if the sample size were more than 5 percent of the universe.)

OTHER ADVANTAGES OF RATIO, REGRESSION, AND DIFFERENCE ESTIMATION

One big advantage of ratio, regression, and difference estimation is that they adjust the sample results to known universe data when we compute totals. If the sample mean for the auxiliary variable turns out to be lower than the universe mean, as in our example, the sample results are adjusted upward. If, however, the sample mean for the auxiliary variable turns out to be higher than the universe mean, the sample results are adjusted downward.

When the calculations must be done manually, difference estimation has one advantage over ratio and regression estimation: the formulas for computing estimates and sampling errors are simple. Also, because of the simplicity of the formulas, difference estimation can be easily adapted to stratified sampling and to various cluster sampling plans.

CHAPTER 5

SAMPLING IN THE AUDIT ENVIRONMENT

This chapter is directed more to the field of financial and management auditing than to program evaluation. However, some of the points here also apply to program evaluation.

In this chapter, we briefly discuss discovery and acceptance sampling, the relationship between audit judgment and sampling, and the characteristics of a good sample.

DISCOVERY SAMPLING

Discovery sampling is a type of sampling that has a specified probability of including at least one item that occurs very rarely in the universe. It is used when there is a possibility of finding such things as fraud and avoidance of internal controls. In discovery sampling, the evaluators can specify the probability of including in the sample at least one item with a particular characteristic, if the characteristic occurs at a specified rate in the universe. If the sample does not turn up an item with this characteristic, the evaluators can make a probability statement that the characteristic's rate of occurrence is less than that specified.

Discovery sampling can be regarded as a special case of attribute sampling. However, in its usual applications, it does not yield an estimated rate of occurrence, and usually it is used only if the particular characteristic's rate of occurrence is thought to be very small--that is, close to zero. For example, discovery sampling is usually used in financial audits to guard against an intolerable rate of fraud.

The evaluators must specify two things: the rate of error, fraud, and so on that would be intolerable and the probability of finding at least one occurrence in the sample (if the rate of occurrence is even this high). The required sample size can usually be looked up in published tables like tables 5.1 and 5.2 on pages 50 and 51. Evaluators who are familiar with logarithms can easily calculate the required sample size for any specified probability level and rate of occurrence.

For the same assumptions that we used in our example above, we would use the following formula, in which we let "antilog" represent the antilogarithm; log is the common logarithm; $Pr(a \geq 1)$ is the probability that the number of occurrences in the sample is equal to or greater than one; and A is the number of occurrences in the universe that would be intolerable.

$$n = N\left\{1 - \text{antilog}\left[\frac{\log(1 - Pr(a \geq 1))}{A}\right]\right\}$$

Table 5.1

The Probability of Finding at Least One Error in a Universe of 500 for Various Numbers of Errors and Sample Sizes

Sample size	Total errors in universe size of 500																
	1	2	3	4	5	10	15	20	25	30	40	50	75	100	200	300	500
5.	1.0	2.0	3.0	4.0	4.9	9.6	14.2	18.5	22.7	26.7	34.2	41.1	55.8	67.4	92.3	99.0	100.0
10.	2.0	4.0	5.9	7.8	9.6	18.4	26.5	33.8	40.4	46.5	56.9	65.5	80.6	89.5	99.4	100.0	
15.	3.0	5.9	8.7	11.5	14.2	26.5	37.1	46.3	54.2	61.0	71.9	79.9	91.6	96.7	100.0		
20.	4.0	7.8	11.5	15.1	18.5	33.8	46.3	56.5	64.9	71.7	81.8	88.4	96.4	99.0	100.0		
25.	5.0	9.8	14.3	18.6	22.7	40.4	54.2	64.9	73.2	79.5	88.2	93.3	98.5	99.7	100.0		
30.	6.0	11.7	17.0	22.0	26.7	46.5	61.0	71.7	79.5	85.3	92.4	96.2	99.4	99.9	100.0		
35.	7.0	13.5	19.6	25.3	30.5	51.9	66.9	7[illegible].3	84.4	89.4	95.2	97.8	99.7	100.0			
40.	8.0	15.4	22.2	28.4	34.2	56.9	71.9	81.8	88.2	92.4	96.9	98.8	99.9	100.0			
45.	9.0	17.2	24.7	31.5	37.7	61.4	76.2	85.4	91.1	94.6	98.0	99.3	100.0				
50.	10.0	19.0	27.1	34.5	41.1	65.5	79.9	88.4	93.3	96.2	98.8	99.6	100.0				
55.	11.0	20.8	29.6	37.4	44.3	69.2	83.0	90.7	95.0	97.3	99.2	99.8	100.0				
60.	12.0	22.6	31.9	40.1	47.4	72.5	85.7	92.6	96.2	98.1	99.5	99.9	100.0				
65.	13.0	24.3	34.2	42.8	50.3	75.5	88.0	94.2	97.2	98.7	99.7	99.9	100.0				
70.	14.0	26.1	36.5	45.4	53.1	78.2	89.9	95.4	97.9	99.1	99.8	100.0					
75.	15.0	27.8	38.7	47.9	55.8	80.6	91.6	96.4	98.5	99.4	99.9	100.0					
80.	16.0	29.5	40.8	50.3	58.3	82.8	93.0	97.2	98.9	99.6	99.9	100.0					
85.	17.0	31.1	42.9	52.7	60.0	84.8	94.2	97.8	99.2	99.7	100.0						
90.	18.0	32.8	44.9	54.9	63.1	86.5	95.1	98.3	99.4	99.8	100.0						
95.	19.0	34.4	46.9	57.1	65.3	88.1	96.0	98.7	99.6	99.9	100.0						
100.	20.0	36.0	48.9	59.2	67.4	89.5	96.7	99.0	99.7	99.9	100.0						
125.	25.0	43.8	57.9	68.5	76.4	94.5	98.8	99.7	99.9	100.0							
150.	30.0	51.0	65.8	76.1	83.3	97.3	99.6	99.9	100.0								
175.	35.0	57.8	72.6	82.3	88.5	98.7	99.9	100.0									
200.	40.0	64.0	78.5	87.1	92.3	99.4	100.0										
225.	45.0	69.8	83.4	90.9	95.0	99.8	100.0										
250.	50.0	75.1	87.6	93.8	96.9	99.9	100.0										
275.	55.0	79.8	91.0	96.0	98.2	100.0											
300.	60.0	84.0	93.7	97.5	99.0	100.0											
325.	65.0	87.8	95.8	98.5	99.5	100.0											
350.	70.0	91.0	97.3	99.2	99.8	100.0											
375.	75.0	93.8	98.5	99.6	99.9	100.0											
400.	80.0	96.0	99.2	99.8	100.0												
425.	85.0	97.8	99.7	100.0													
450.	90.0	99.0	99.9	100.0													
475.	95.0	99.8	100.0														
500.	100.0	100.0															

(Probability is 100.0)

Source: U.S. Air Force, Auditor General, Handbook of Practical Sampling Procedures for Internal Auditors (Norton Air Force Base, Calif.: 1966), p. 125.

Then $Pr(a \geq 1)$ is set at 0.95, and A is 5. Substituting these values in the formula, we get a sample size of 225, as shown below:

$$n = 500\left\{1 - \text{antilog}\left[\frac{\log(1-0.95)}{5}\right]\right\}$$

$$n = 500\left[1 - \text{antilog}\left(\frac{8.69897-10}{5}\right)\right]$$

$$n = 500(1 - 0.5493)$$

$$n = 225.35 \quad \text{or} \quad 225$$

For example, assume that the universe size is 500 and the evaluators want to be 95-percent certain that, if the error rate in the universe is 1 percent, they will find at least 1 error in the sample. If the specified intolerable error rate is 1 percent, A is 1 percent of 500, or 5. Instead of making the calculations, we can usually use published tables like tables 5.1

Table 5.2

The Probability of Finding at Least One Error in a Universe of 600 for Various Numbers of Errors and Sample Sizes

Sample size	Total errors in universe size of 600: 1	2	3	4	5	10	15	20	25	30	40	50	75	100	200	300	500
5.	0.8	1.7	2.5	3.3	4.1	8.1	11.9	15.6	19.2	22.7	29.3	35.4	48.8	59.9	86.9	96.9	100.0
10.	1.7	3.3	4.9	6.5	8.1	15.6	22.5	28.9	34.9	40.4	50.1	58.4	74.0	84.1	98.3	99.9	100.0
15.	2.5	4.9	7.3	9.7	11.9	22.5	31.9	40.2	47.6	54.1	64.9	73.3	86.8	93.7	99.8	100.0	
20.	3.3	6.6	9.7	12.7	15.6	28.9	40.2	49.8	57.9	64.8	75.4	83.0	93.4	97.6	100.0		
25.	4.2	8.2	12.0	15.7	19.2	34.9	47.6	57.9	66.3	73.0	82.8	89.2	96.7	99.1	100.0		
30.	5.0	9.8	14.3	18.6	22.7	40.4	54.1	64.8	73.0	79.4	88.0	93.1	98.4	99.6	100.0		
35.	5.8	11.3	16.5	21.4	26.0	45.4	59.9	70.5	78.4	84.3	91.7	95.7	99.2	99.9	100.0		
40.	6.7	12.9	18.7	24.2	29.3	50.1	64.9	75.4	82.8	88.0	94.3	97.3	99.6	99.9	100.0		
45.	7.5	14.4	20.9	26.9	32.4	54.4	69.4	79.5	86.3	90.9	96.0	98.3	99.8	100.0			
50.	8.3	16.0	23.0	29.5	35.4	58.4	73.3	83.0	89.2	93.1	97.3	98.9	99.9	100.0			
55.	9.2	17.5	25.1	32.0	38.3	62.1	76.8	85.9	91.4	94.8	98.1	99.3	100.0				
60.	10.0	19.0	27.1	34.5	41.1	65.4	79.8	88.3	93.2	96.1	98.7	99.6	100.0				
65.	10.8	20.5	29.1	36.9	43.7	68.5	82.5	90.3	94.7	97.1	99.1	99.8	100.0				
70.	11.7	22.0	31.1	39.2	46.3	71.4	84.8	92.0	95.8	97.8	99.4	99.8	100.0				
75.	12.5	23.5	33.1	41.5	48.8	74.0	86.8	93.4	96.7	98.4	99.6	99.9	100.0				
80.	13.3	24.9	35.0	43.7	51.2	76.4	88.6	94.6	97.4	98.8	99.7	99.9	100.0				
85.	14.2	26.3	36.8	45.8	53.5	78.6	90.2	95.5	98.0	99.1	99.8	100.0					
90.	15.0	27.8	38.6	47.9	55.8	80.6	91.5	96.3	98.4	99.3	99.9	100.0					
95.	15.8	29.2	40.4	49.9	57.9	82.4	92.7	97.0	98.8	99.5	99.9	100.0					
100.	16.7	30.6	42.2	51.9	59.9	84.1	93.7	97.6	99.1	99.6	99.9	100.0					
125.	20.8	37.4	50.4	60.8	69.0	90.5	97.1	99.1	99.7	99.9	100.0						
150.	25.0	43.8	57.9	68.5	76.4	94.5	98.7	99.7	99.9	100.0							
175.	29.2	49.9	64.5	74.9	82.3	96.9	99.5	99.9	100.0								
200.	33.3	55.6	70.4	80.3	86.9	98.3	99.8	100.0									
225.	37.5	61.0	75.7	84.8	90.6	99.1	99.9	100.0									
250.	41.7	66.0	80.2	88.5	93.3	99.6	100.0										
275.	45.8	70.7	84.2	91.5	95.4	99.8	100.0										
300.	50.0	75.0	87.6	93.8	96.9	99.9	100.0										
325.	54.2	79.0	90.4	96.5	98.0	100.0											
350.	58.3	82.7	92.8	97.0	98.8	100.0											
375.	62.5	86.0	94.8	98.1	99.3	100.0											
400.	66.7	88.9	96.3	98.8	99.6	100.0											
425.	70.8	91.5	97.5	99.3	99.8	100.0											
450.	75.0	93.8	98.5	99.6	99.9	100.0											
475.	79.2	95.7	99.1	99.8	100.0												
500.	83.3	97.2	99.5	99.9	100.0												
550.	91.7	99.3	99.9	100.0													
600.	100.0	100.0	100.0														

(Probability is 100.0)

Source: U.S. Air Force, Auditor General, Handbook of Practical Sampling Procedures for Internal Auditors (Norton Air Force Base, Calif.: 1966), p. 126.

and 5.2. We could use table 5.1 for a universe of 500. Reading down the column for 5 errors to the row that corresponds to a probability of 95 percent, we find that a sample size of 225 is required.

Then the evaluators select a simple random sample of items and examine each item until they find one with an error or until they have examined the entire sample and found none. If they find an error, they know that the error rate is at least as great as the specified intolerable rate and can extend the review perhaps to the entire universe. If they find no deficiencies, they can conclude that the rate of occurrence of deficiencies is less than that specified as intolerable.

An advantage of discovery sampling is that the probability of finding at least one error will increase if the rate of occurrence of deficiencies is greater than the intolerable rate specified by the evaluators. Thus, the likelihood of more quickly finding the one error in the sample is increased, and the average sample size that actually has to be examined is smaller.

ACCEPTANCE SAMPLING

Acceptance sampling is a type of sampling that provides us with the decision of whether to accept or reject a specific universe. It also provides us with assurance that "on the average" very bad universes will be rejected and very good universes will be accepted. In acceptance sampling, a random sample of items is drawn from a universe (or "lot"), and the sample is examined or tested. On the basis of this examination, the decision is made to accept or reject the entire lot. The decision may also be made to draw one or more additional samples if the results of the first sample are inconclusive. These are called "double," or "multiple," "acceptance sampling plans," terminology that is derived from the field of industrial quality control. If the number of deficiencies found in the sample is greater than a predetermined number, the entire lot is rejected. If the number of deficiencies in the sample is equal to or less than the predetermined number, the entire lot is accepted.

Acceptance sampling is also a variation of attribute sampling, but it does not permit the estimation of the rate of occurrence of deficiencies.

To use acceptance sampling, the evaluators must specify four criteria:

1. the limit of acceptable quality, or the maximum percentage of deficiencies that can be considered satisfactory on the average over the long run. It should have a high probability of acceptance or a low probability of rejection;

2. the lot tolerance percent defective (LTPD), or the maximum percentage of defects that can be tolerated in a lot. It should have a low probability of acceptance or a high probability of rejection;

3. the probability of incorrectly rejecting a lot of acceptable quality (sometimes called "producer's risk");

4. the probability of incorrectly accepting a lot of unacceptable quality (sometimes called "consumer's risk").

Once the evaluators have specified these criteria, they consult tables of acceptance sampling plans for the plan that comes closest to the criteria. (For an example, see table 5.3.) The plan will give the sample size required and the maximum number of acceptable defectives in the sample, referred to as the "acceptance number."

To give an example of an acceptance sampling plan, let us assume that the evaluators want to verify the accuracy of key-

Table 5.3

An Acceptance Sampling Table[a]

Lot size	Acceptable quality limit 1.21-1.60% Sample size	Acceptance number[b]	LTPD	Acceptable quality limit 1.61-2.00% Sample size	Acceptance number[b]	LTPD
1- 15	All	0		All	0	
16- 50	14	0	13.6	14	0	13.6
51-100	16	0	12.4	16	0	12.4
101-200	35	1	10.5	35	1	10.5
201-300	37	1	10.2	37	1	10.2
301- 400	38	1	10.0	60	2	8.5
401- 500	60	2	8.6	60	2	8.6
501- 600	60	2	8.6	60	2	8.6
601- 800	65	2	8.0	85	3	7.5
801-1,000	65	2	8.1	90	3	7.4
1,001-2,000	95	3	7.0	120	4	6.5
2,001-3,000	120	4	6.5	180	6	5.8
3,001-4,000	155	5	6.0	210	7	5.5
4,001-5,000	155	5	6.0	245	8	5.3
5,001-7,000	185	6	5.6	280	9	5.1
7,001- 10,000	220	7	5.4	350	11	4.8
10,001- 20,000	290	9	4.9	460	14	4.4
20,001- 50,000	395	12	4.5	720	21	3.9
50,001-100,000	505	15	4.2	955	27	3.7

Source: H. F. Dodge and H. G. Romig, Sampling Inspection Tables, Single and Double Sampling, 2nd ed. (New York: John Wiley and Sons, 1959).

[a]For this table, according to Dodge and Romig, the risk of incorrectly rejecting a lot whose average quality limit is indicated by the column headings is about 5 percent. The risk of incorrectly accepting a lot whose percentage of defectives equals an entry in the LTPD columns is, at most, 10 percent.

[b]According to Dodge and Romig, accept the lot if the number of defectives does not exceed this value.

punching. Each keypuncher's work is batched into boxes of 2,500 cards each, after the cards have been punched. The evaluators have decided to define an error card as any card containing one or more errors. The evaluators want to take no more than a 5-percent risk of incorrectly rejecting a batch of cards whose error rate is 1.75 percent, and they want to take only a 10-percent risk of incorrectly accepting a batch whose error rate is 5.8 percent. Therefore, the lot size is 2,500, the acceptance quality limit is 1.75 percent, the LTPD is 5.8 percent, the producer's risk is 5 percent, and the consumer's risk is 10 percent.

As shown in table 5.3, the sampling plan for lot size 2,001 to 3,000, and an acceptable quality limit of 1.61 to 2 percent will meet these criteria. The evaluators will take a random sample of 180 cards from each box, key-verify each sample card, and accept the lot if the number of error cards is 6 or less; otherwise, the lot will be rejected.

At first glance, acceptance sampling seems attractive because tabulated plans are available and the sample sizes are usually smaller than those required for estimation sampling. However, the assumptions underlying acceptance sampling make it unsuitable for most GAO work. The major assumption is that many samples are drawn from a continuous stream of homogeneous items (such as ball bearings or artillery shells) being produced or processed by a system under some form of control. Thus, the decision whether to accept or reject many lots over the long run is generally in accordance with the four criteria described above, but this may not be so with a single lot. The policy or oversight researcher's opinion must be based on the results of a single test; the lot being examined may have been produced by several different processes or departments; and controls, if any, may vary from department to department.

Also, the industrial sampler has little concern about moderately bad situations, but they are important to evaluators because they may indicate fraud or collusion. Further, evaluators, unlike industrial samplers, cannot merely send back a "bad" lot to be reworked at no cost. If they reject a lot because of a moderately bad situation, they may require an unnecessary extension of the test that may cost GAO, or the agency, money. The evaluators should take all these factors into account before deciding to use acceptance sampling.

Acceptance sampling formulas and tables do not permit the development of statistical estimates. However, once an entire acceptance sample has been evaluated, it can be used to develop attributes or variables estimates, if it is large enough.

AUDIT JUDGMENT AND STATISTICAL SAMPLING

The charge has been made that statistical sampling prevents auditors from using professional judgment in conducting reviews. This is not correct; statistical sampling is merely a tool to help them make wise decisions. The auditors still decide what type of review to make, how and when to use sampling, and how to interpret the results. In applying statistical sampling techniques to audit testing, auditors must make the following decisions that involve professional judgment.

1. The auditors must define the problem. They must decide what to measure, what type of information will provide sufficient facts for the formation of an opinion, and what testing procedures to use.

2. The level of confidence must be specified. This is precision, or the probability that an estimate made from the sample will fall within a stated interval of the true value for the universe as a whole. Auditors may think of it as the percentage of times that a correct decision (within the specified precision limits) will result from using an estimate based on a sample.

3. The auditors must define the universe for size and other characteristics. They decide what type of items will be included and excluded and specify the time period to be covered.

4. The areas susceptible of sampling must be determined. They should comprise numerous items or similar transactions that can be measured. The auditors' assessment of the internal control system for an area may determine whether statistical sampling is appropriate. A strong internal control system, for example, may reduce testing to the minimum necessary for verification and may, therefore, call for a different sampling plan or no statistical sampling at all. Prior experience, as well as information from prior audits, plays a role here. Prior audits may suggest that certain kinds of records are more prone to error and need higher verification rates than other kinds of records. Thus, auditors may have to stratify the universe between records likely to have a high error rate and those likely to have a low error rate.

5. The maximum error rate that the auditors will consider acceptable must be decided, as well as the definition of an error. Or, if the auditors are attempting to estimate the value of some balance sheet amount, they must determine the required precision of the estimate in terms of the materiality of the amount being examined and the overall objective.

6. Conclusions about the universe must be drawn from the sampling results. In arriving at these conclusions, the auditors must judge the significance of the errors they have discovered.

Because statistical sampling provides more and better information, it permits greater use of professional judgment and enables auditors to more effectively analyze the results of tests. And by reducing the work load statistical sampling allows more time to use professional judgment.

THE CHARACTERISTICS OF A GOOD SAMPLE

In traditional estimation sampling, the ideal sample is characterized as representative. That is, the sample produces an unbiased estimate of the true universe characteristic, and this estimate is as precise as possible given the resources available for designing the sample and collecting the data.[1]

When we sample for audit purposes, we should expect a good sample to be not only representative but also corrective, protective, and preventive.

As we noted above, "representative" means that the sample estimates the true universe characteristic as accurately as

[1]This section is based on Ijiri and Kaplan (1970), pp. 42-44. Their article is highly recommended for everyone doing sample design for financial audits and for economy, efficiency, and effectiveness audits.

possible. For example, if we were taking a sample from an inventory of 2,000 items whose true (but unknown) total value was $860,000, we would like the estimated total computed from our sample data to be as close to $860,000 as possible. If the same inventory had a true (but unknown) error rate of 5 percent, we would like our sample to estimate an error rate as close to 5 percent as possible. From our previous discussion, we know that if the universe is defined correctly, if the list of universe items from which we draw the sample is correct, and if we use random procedures to select the sample items, the sample will be representative. The estimate will be unbiased and have a measurable precision and confidence level.

"Corrective" means that the sample will locate as many error items as possible, so that they can be corrected. Even if the system that generated the errors is not corrected, as many specific instances as possible will be. (Regarding this characteristic, some may state that it is not GAO's job to do the agency's work. However, if we suspect a problem and if, by careful sample design, we can disclose a large number of problem items, our findings may be more effective and more likely to bring about corrective action.)

In the example above, we assumed that the true error rate in the universe was 5 percent. If we selected a random sample of 100 items from the universe, we would expect to find only 5 error items. However, if we could identify in advance the error-prone items--that is, the items most likely to contain errors based on what they were, how they were stored, how they were accounted for, or the like--we could perhaps isolate them from the other items and take all our sample or the largest part of our sample from these items. Thus, we could maximize the number of errors disclosed by the sample.

"Protective" means that the person who does the sampling attempts to include the maximum number of high-value items in the sample. This approach is common in auditing when auditors isolate the high-value items from the rest of the universe, gather data on all these items, and gather data from a sample of the remaining items. Continuing with our inventory example, if we knew that 100 of the items had values in excess of $1,000, we might audit all these items and audit a sample of the remaining items. Or if, in addition to knowing that 100 of the items had values in excess of $1,000, we knew that 500 items had values of $100 to $1,000, we might review half of the 500 items and all the items that had values in excess of $1,000.

"Preventive" means that the sampling method gives agency managers no idea which items will be selected during our review.

When designing a sampling plan, evaluators should keep in mind the desirability of obtaining a sample that is as representative, corrective, protective, and preventive as possible. To do so, they should stratify the universe on the basis of dollar value and the likelihood that the items contain errors, use some random method to select the sample from each stratum, and weight the results from each stratum to compute overall estimates for the universe.

It is not possible, however, to optimize all four characteristics in a single sample. Instead, a balance must be struck, depending on which characteristic is most important in view of the audit objective. Also, in certain types of jobs, one or more of the characteristics may not require consideration at all.

CHAPTER 6

SAMPLE SELECTION PROCEDURES

One of the most critical parts of every sampling operation is the actual random selection of the units to be examined. Mistakes in estimating the sample size or in evaluating the sample results can be corrected or appropriate adjustments can be made before, or sometimes even after, data collection has been completed. However, a mistake in the sample selection process can materially distort or even invalidate the sample results, particularly if the mistake is not detected, and can sometimes make it necessary to redo or abandon the work. This chapter describes the steps involved in sample selection. A more detailed description of selection procedures, the various problems that may confront the sampler, and the methods of overcoming these problems is in appendix II.

As we have noted, developing a sampling plan is iterative. However, we must assume that certain components of the plan are fixed, even though they might change in a later iteration. Therefore, we assume that the

- o audit or evaluation questions have been formulated,
- o audit or evaluation strategy has been chosen,
- o universe of interest has been defined,
- o sample design has been chosen,
- o sampling unit has been defined,
- o sample size has been determined, and
- o data collection and analysis plans have been made.

In order to make intelligent decisions about the last five of these points, we must have rather detailed knowledge about the physical location and accessibility of the universe and a good estimate of the number of items in the universe, if the exact number is not available. We must know the practical aspects of gathering the data. We must know whether the sampling units are documents, school pupils, spare parts, or the like; whether the universe is at one location or at several locations some distance apart; and whether the sampling units are in file drawers, storage bins, or neighborhoods. And we must know how the measurements will be administered. To individuals? To groups? In person? By telephone? By mail?

THE SELECTION OF SAMPLING UNITS

There are three basic procedures for selecting statistical samples: random number sampling, systematic selection with a random start, and selection based on randomly selected combinations of digits in the lower order positions of socially assigned identification numbers. Computer programs can also be used. We present manual methods of sample selection below because evaluators may not always have automatic data processing equipment available and because we believe that evaluators should understand how the computer produces the sample numbers.

Systematic sampling, or systematic selection with a random start

Because of its simplicity and usefulness in many situations, systematic sampling, or systematic selection with a random start, will be discussed first. In this selection procedure, the sample is selected from the universe on the basis of a fixed, or uniform, interval between sampling units, after a random starting point has been determined. The uniform interval between units is obtained by dividing a given sample size into the universe size and dropping any decimals in the result. The random start is selected from a table of random digits and is the first combination of digits that is between 1 and the uniform sampling interval, inclusive (see appendix II). The random start ensures that, for all practical purposes, every sampling unit has an equal opportunity of being selected.

For example, assume that we want to draw a sample of 200 items from a file containing 10,100 items. Dividing the sample size into the universe size gives a quotient of 50.5. Rounding downward to the nearest whole number gives a sampling interval of 50. From a table of random digits, the first number between 1 and 50 is selected to obtain the starting point. (Note that since the sampling interval is a two-digit number, we look at two-digit combinations in the table of random digits. Methods for obtaining a "random" starting point in the table are explained in the section of this chapter entitled "Random Number Sampling.") Suppose the random starting number between 1 and 50 is 36. We start with item number 36 and pull every 50th item thereafter; the 36th item, 86th item, 136th item, and so on will constitute the sample.

Situations calling for systematic sampling

Systematic sampling may be used when the sampling units are not numbered or when it would be too cumbersome to attempt to match the sampling units against random numbers. Here are some circumstances in which systematic selection may be used advantageously:

1. The sampling units are long lists or pages of lists.

2. The sampling units are filed on index cards that are not serially numbered, or if they are numbered, they are not in numerical sequence.

3. The sampling units are not suitably numbered and are intermingled with other items that are not to be included in the sample.

4. The sampling units are numbered in blocks of numbers, and some blocks are not used.

Cautions in using systematic selection

To obtain a sample size that is neither too large nor too small, the evaluators must know or be able to closely estimate the universe size. Before using systematic selection, the evaluators should make sure that the sample will be drawn from the entire universe. If the sample is to be drawn from a list of items, the list must be complete; if the sample is to be drawn from a file, all the folders must be in the file, or charge-out cards or a similar system must be used to mark the position of missing folders. Otherwise, the evaluators must make special arrangements to ensure that missing sampling units have the opportunity of being selected.

In certain types of universe, the items are arranged so that certain significant characteristics recur at regular intervals. This is called "periodicity." Some examples are daily highway traffic passing through a certain intersection during the day and department store sales during the week. A systematic sample might consist of a certain time of day from the former type of universe or a certain day of the week from the latter, even though every time point and every day would have an equal opportunity of being selected. Obviously, samples like these would be unrepresentative of the entire universe.

Another example is a universe consisting of a payroll list on which every 25th employee is a superintendent. A systematic sample of every 25th name or multiple thereof could result in a sample consisting entirely of superintendents or, more likely, in a sample that excluded all superintendents. Obviously, neither situation is desirable. Discussions with agency personnel commonly disclose situations of this type.

In general, this problem can be minimized if the sample is taken from lists of persons arranged alphabetically by name or in order of Social Security number or from lists of inventory items in sequence by stock numbers or by the dates the items were first stocked. Before using systematic selection, it is imperative that the evaluators determine whether there is a relationship between the arrangement of the universe and the characteristic being measured.

When systematic selection with a random start is used, the selection process must be continued throughout the entire universe, as originally defined by the evaluators, even though the universe size may have been underestimated when the sampling interval was calculated and even though the selection will produce a larger sample than required. Under no circumstances should the evaluators stop when they reach the required sample size. This is equivalent to "throwing out" part of the universe and could result in an unrepresentative sample. If the sample turns out to be too large, it can be reduced by using one of the procedures described in appendix II.

Once the sample has been selected, it is not permissible to substitute other items for sample items that are missing (that are out of the file or the like) or for sample items that may not have adequate supporting material to permit measurement. Every effort should be made to locate the missing items or supporting material. If they cannot be located, this fact should be noted and reported as one of the sample results.

Evaluators should also be aware that, if the sampling units are arranged in ascending or descending order of magnitude, a systematic sample will yield a smaller estimate of the sampling error than a random number sample (as we discuss later in this chapter).

<u>Selection based on randomly selected combinations of terminal digits in identification numbers</u>

This is really another method of systematic selection with a random start, but the mechanical procedures for selecting the sample are different. Certain types of sampling units have been assigned consecutive identification numbers. Examples are Social Security numbers, inventory stock numbers, and transaction numbers assigned in the order in which documents were received or processed. The important feature of the identification numbers is that the terminal digits (usually the last three, sometimes the last four) can usually be assumed to be random with respect to the characteristics the evaluators want to measure.

A sample can be selected from a universe of units having such identification numbers by selecting all the items (or persons) having identification numbers ending in a certain randomly selected digit or combination of digits. Because there are 10 digits from 0 through 9, each digit will appear in the last position in 10 percent of the identification numbers. Thus, all identification numbers having a terminal digit that matches a randomly selected digit from 0 through 9 will constitute a random 10-percent sample. Similarly, there are 100 possible combinations of pairs of digits between 0 and 99. Each pair of digits will appear in the last two positions of 1 percent of the identification numbers. Thus, all identification numbers whose last 2 digits match a randomly selected pair of digits between 0 and 99 will constitute a random 1-percent sample.

The steps in this selection procedure are

1. determining the required sample size,

2. dividing the sample size by the universe size to obtainthe sampling rate (or percentage), and

3. selecting the required quantity of random digits or combinations of random digits by using a table of random digits or some other suitable source of random numbers.

The percentage indicates the number of digits that should be in the randomly selected combination of digits and the number of combinations that should be selected. For a 20-percent sample, we would match against 2 randomly selected digits between 0 and 9; for a 30-percent sample, against 3 randomly selected digits between 0 and 9; for a 1-percent sample, against 2 randomly selected digits between 0 and 99; and for a 3-percent sample, against 3 pairs of randomly selected digits between 0 and 99.

For example, to measure the accuracy of payroll records at an installation employing 6,000 persons, the evaluators determine that a sample of 240 records will be adequate. They decide to draw the sample by selecting the payroll records of employees whose Social Security numbers end in certain randomly selected pairs of digits. Because a sample of 240 from a universe of 6,000 is a 4-percent sample, the evaluators will need 4 pairs of digits. From a table of random digits, they select the following pairs of digits between 0 and 99: 01, 26, 85, and 94. Then they examine the payroll records of all employees whose Social Security numbers end in those digits.

In this type of sampling, selection should not be based on the leading digits in the identification number, because these digits frequently are codes and are not assigned in serial order.

(This type of sampling is sometimes called digital selection, digital sampling, or junior digit sampling.)

<u>Random number sampling</u>

For a simplified example of random number selection of sampling units, suppose that we want to make a random selection of one pay record from a universe of 10 pay records. The evaluators know that the likelihood of selecting any specific record is 1 in 10. The probability is usually expressed as a proportion, 0.10, or as a percentage, 10 percent. The probability is known because the only factor involved in random selection is chance. Subjective considerations (conscious or otherwise), such as selecting new-looking pay records, people who look approachable, or military installations in nearby locations, are completely avoided.

If the universe is very small, such as the 10 pay records, the sample could be selected by recording the serial number or some other identification symbol of each of the 10 pay records on a separate slip of paper. The slips of paper could then be placed in a container and mixed thoroughly, and a blindfolded person could withdraw a quantity of slips equal to the specified sample size. The identification numbers on the slips of paper would indicate which pay records to select. Although this selection method is practical only when the universe is very small, most random selection methods are merely extensions of it.

For most projects, the procedures are to (1) have a set of random numbers generated on a large computer system or (2) use the computer to select randomly from records in machine-readable form. At this writing, however, the random number generators on microcomputers are usually inadequate for both these procedures.

Programmed random number generators are available for use with most large computer systems. These generators are designed to produce a selection of random numbers that will be suitable for, or can be adapted to, most numbering systems, including compound numbering systems. For most sampling applications, such generators reduce to minutes the time required for random number selection.

In its simplest form, random number sampling is a selection procedure in which a quantity of random numbers equal to the specified sample size is first selected from a table of random digits, then matched against the serial numbers, stock numbers, transaction numbers, or whatever, and finally assigned to the sampling units in the universe. If the sampling units are not numbered, the evaluators may develop a numbering system for identifying each unit. For example, if documents are entered in a register with 25 lines to the page, documents 1 through 25 could be assigned to the first page, 26 through 50 to the second page, and so on. If the items have their own numbers, the sampling process will be greatly simplified if they are arranged in numerical sequence or if a list of the cases in numerical sequence is available. The sampling units having numbers that correspond to the selected random numbers constitute the sample.

First, the beginning and ending numbers of the items in the universe are determined. Then numbers falling between the beginning and ending numbers equal to the specified sample size are selected from a table of random digits. For example, if we want to select a sample of 200 items from a universe of 8,894 items numbered from 265 through 9,158, we start at some random point in the table. Going either down the table columns or across the rows, we select the first 200 four-digit numbers that fall between 0265 and 9,158, inclusive. Note that we must look at four-digit numbers because the largest number in the universe has four digits. The quantity of digits in the numbers that are read must always equal the quantity of digits in the largest number of the universe. For most applications, numbers that duplicate a number that has already been drawn are discarded, and the quantity of additional random numbers that will achieve the required sample size is selected. Any table of random digits can be used with any purely numerical numbering system, and the table can be adapted for use with an alphabetical-numerical numbering system.

In most situations, the use of random number sampling is not as simple as this. Appendix II gives detailed descriptions of how to determine the starting point in a table of random digits, how to record the random numbers that are used, how to adapt random number sampling to compound numbering systems, and other complicated situations. Statisticians can provide guidance on these procedures.

When preliminary results indicate that the sample is larger than needed, the evaluators may want to decrease the sample size. Basically, this is done by taking a random sample of the random sample. Details on this procedure are also given in appendix II.

THE APPLICATION OF SELECTION PROCEDURES

The selection procedures described in this chapter are applicable to simple random sampling, stratified sampling, and cluster sampling. In simple random sampling, we have only a simple universe, and only one procedure is used to select the entire sample.

With stratified sampling, the universe is divided up into two or more separate subuniverses, or strata. Thus, a different procedure could be used to select the samples in the various subuniverses. Depending on the arrangement of the items in the subuniverses and the numbering systems employed, it might be advisable to use random number sampling in some of the strata and systematic selection with a random start in others.

In the application of these procedures to cluster sampling, one procedure might be used to select the clusters; if it were necessary to select a sample of items within these clusters, a different procedure might be used to do this. For example, systematic selection with a random start could be used to select the clusters and random number sampling could be used to select the sample items within the clusters.

SELECTION WITH PROBABILITY PROPORTIONAL TO SIZE

When evaluators apply random selection procedures to cluster sampling, the clusters can be selected with probability proportional to size (PPS) or to a related variable that can be used as a measure of size. This sampling method is based on the assumption that the variable to be measured is highly correlated

Table 6.1

Selection with Probability
Proportional to Size

Office	Number of claims	Range of cumulative numbers Lower	Upper	Random numbers
New York	2,936	1	2,936	01038; 02770; 02471; 01174
Hicksville	1,245	2,937	4,181	
Paterson	471	4,182	4,652	
Bronx	2,335	4,653	6,987	
Atlanta	1,775	6,988	8,762	
Pittsburgh	1,254	8,763	10,016	09745; 09094
Tampa	636	10,017	10,652	
Charlestown	174	10,653	10,826	10679
Chicago	2,562	10,827	13,388	12993
Springfield	1,630	13,389	15,018	14922; 13547; 14300
Cincinnati	687	15,019	15,705	15150
South Bend	139	15,706	15,844	
St. Paul	1,818	15,845	17,662	17237
St. Louis	1,114	17,663	18,776	17850
Columbia	148	18,777	18,924	
Detroit	2,159	18,925	21,083	
Cleveland	1,327	21,084	22,410	
Fort Worth	668	22,411	23,078	22638; 22952
Waco	163	23,079	23,241	23223
San Antonio	430	23,242	23,671	
Nashville	625	23,672	24,296	
Chattanooga	202	24,297	24,498	
Jackson	187	24,499	24,685	
Oakland	1,469	24,686	26,154	25972; 25402
Portland	723	26,155	26,877	
Fresno	281	26,878	27,158	
Los Angeles	2,162	27,159	29,320	
Van Nuys	361	29,321	29,681	
San Diego	597	29,682	30,278	29841
Honolulu	169	30,279	30,447	

Source: Interstate Commerce Commission, Bureau of Transport Economics and Statistics, Table of 105,000 Random Decimal Digits (Washington, D.C.: 1949), p. 20, col. 7, line 971, through col. 9, line 952.

with some data already known about the cluster, such as number of inhabitants, dollar volume of transactions, or number of students in a school system. If the assumption is correct, this selection method will yield a smaller sampling error than other methods would.

For example, table 6.1 lists claims-paying offices and the number of claims each one paid in 1982. Suppose we want to estimate the dollar value of the claims that were paid. It is reasonable to assume that the dollar value of claims that were paid is approximately proportional to the number of claims that were paid. Assume that the maximum number of offices that can be audited is 20.

First, we set up a range of cumulative numbers of claims for each office, as shown in the third and fourth columns of the table. Next, we select 20 random numbers between 1 and the total number of claims paid--that is, between 1 and 30,447. Then we enter each random number on the line for the office whose range of paid claims includes the random number. (For example, random number 9,745 is entered on the line for Pittsburgh.) This identifies the sample office. Note that some of the offices are included in the sample more than once; this is characteristic of

PPS sampling. Sampling with replacement is used. Thus, when the random numbers are selected, duplicates should not be eliminated.

Each office's probability of selection is proportional to the number of paid claims. Yet each office, from the smallest to the largest, has an opportunity of being selected.

This example shows only how the sample would be selected. The major use of PPS sampling is in two-stage sampling when the cluster sizes vary greatly, as they do here. If clusters were chosen with equal probability, the variation in cluster sizes would increase the computed variation between cluster totals. Using two-stage sampling and PPS sampling to select the primary units, the evaluator can calculate subsampling rates within primary units in a way such that the second-stage sample sizes within each primary unit are equal and, at the same time, the sample is self-weighting. Therefore, the sample can be treated as if it were a simple random sample of clusters, which greatly simplifies the calculations.

The formulas for computing estimates and sampling errors for samples selected with PPS are beyond the scope of this paper.

A FINAL CHECK

Once the selection procedure has been decided upon, the procedures to be followed should be written in a sampling plan. Ordinarily, no deviations from the plan should be permitted during the selection process. If unforeseen circumstances make it necessary to modify the sampling plan, the circumstances as well as the modified procedure should be described in the working papers.

Before leaving the field, the evaluators should review the entire sample selection process to ensure that the

1. correctly defined the universe,
2. drew the sample from the entire universe as originally defined,
3. did not substitute readily accessible items for items that would have been difficult to locate or question, and
4. correctly recorded the pertinent information on each item.

CHAPTER 7

DATA COLLECTION AND ANALYSIS CONSIDERATIONS

RELATED TO SAMPLING

Sampling is a precursor to data collection. In this chapter, we briefly review some basic ideas about data collection. Other transfer papers, giving much greater detail about data collection, are cited.

Data collection methods and the types of data gathered vary with the type of application. Examples of applications from various disciplines follow.

- In evaluation and policy analysis, the sampling units are often people who are interviewed, either in person or by telephone, or who are asked to fill out mailed questionnaires.

- When data are collected from the general public, the housing unit is often the only means of getting at the persons in the sample. When this is so, we have a cluster sampling situation in which the housing unit is the cluster. All the people in the household come into the sample at once.

- In accounting, data are quite frequently gathered from documents such as vouchers, purchase orders, and ledgers or from computer files. Sometimes the data are gathered by actually counting or measuring and pricing items, as in verifying a physical inventory.

- To verify various types of book balances, evaluators may have to send letters to the persons listed on an agency's records in order to determine that the agency's information is correct. In verifying inventory held in public warehouses, the evaluators might send confirmation letters to the management at each warehouse, asking them to say how much inventory is held in the warehouse.

- When agricultural data are collected, quite often measurements are made in the field, such as measuring the growth of plants treated with different fertilizers or the prevalence of blight on various types of plants. Alternatively, such data as the price farmers received for grain may be collected from farmers' records or from grain elevator operators.

- In industry, data are frequently collected by physical measurements, such as testing the tensile strengths of wire or measuring the temperature at which a thermostatic switch will operate.

CHECKING THE RELIABILITY OF DATA COLLECTION

When data are collected, every effort should be made to examine the correct sampling units, to collect and record the data correctly, and to have another person independently verify all measurements and computations. If the data collection operation is large, it is usually advisable to have qualified, well-trained personnel make a quality review of a randomly selected data subsample. This permits a statistical measurement of the data collection errors that can be used to check the accuracy of the final sample results.

If the data are to be collected by questionnaire (including confirmation letters), the evaluators must remember that once they have put the questionnaire in the mail, for all practical purposes it is impossible to further explain the questions and instructions. What comes back depends on how the respondents react to the questionnaire and how well they understand it. Therefore, the questionnaire should be as short as possible, it should look easy to answer, and the questions and instructions should be as simple as possible. The questionnaire should be pretested under conditions that are as similar as possible to those under which it will be answered. With rare exceptions, questionnaires should be designed by an expert in questionnaire design. More details will be covered in PEMD's forthcoming transfer paper on developing and using questionnaires.

If the data are to be collected in person or by telephone, the interviewers should be given a complete set of instructions or an interviewer's manual. They should use an interview form, a questionnaire designed by an expert that is read to the respondent, so that all the interviewers ask the same questions. The interviewers should be trained to read each question exactly as it was written and to provide explanations and clarifications if the respondents do not understand it. Also, interviewers should be trained to maintain a friendly but neutral attitude and not inject personal opinions that might lead the respondents into giving certain answers. (For more details, see GAO, 1985.)

Even with the best training and in the best circumstances, each interviewer has mannerisms that affect people differently and thus affect their responses. This is called "interviewer bias." To attempt to overcome it, interviewer assignments should be randomized so that a single interviewer is not responsible for all the interviews in a particular area, town, or section of a city. (Travel costs are, of course, a big factor in deciding whether this can be done and how.)

For example, assume that interviews are to be conducted in three small cities A, B, and C, which are fairly close together and that three interviewers X, Y, and Z are available. To randomize assignments, we randomly assign one third of the interviews in city A to interviewer X, one third to interviewer Y, and one third to interviewer Z, and we follow the same procedure in cities B and C. (More sophisticated schemes for randomizing interviewer assignments and testing interviewer bias are described in the literature.) Randomization of telephone interviewer assignments, of course, is much more easily accomplished.

THE PROBLEM OF MISSING DATA AND NONRESPONSES

One of the most troublesome problems in any type of data collection operation--whether it is a sample or complete universe coverage--is missing data, sampling units for which the data cannot be collected. Examples are persons who do not bother to fill in and send back questionnaires and those who send back incomplete questionnaires, persons who are not at home or prefer not to be interviewed in person or by telephone, and documents missing from a file from which the sample is selected.

It can be a serious mistake to assume that the respondents are representative of those who do not respond. For example, suppose that we want to know whether prisoners at a state penitentiary favor abolition of the death penalty. The universe is the 2,500 prisoners at the penitentiary on a certain date. We send each prisoner a questionnaire asking only one question:

> "Do you favor abolition of the death penalty?
> (Check one) yes ___ no ___ ."

We receive 1,500 completed questionnaires (a 60-percent response rate). A tabulation of the responses shows

Response	Number	Percent
Yes	900	60
No	600	40
Total	1,500	100

From this, we might conclude that the prisoners favors abolition of the death penalty by a majority of 3 to 2. However, if information were available that permitted us to analyze the type of sentence the respondents had received, we might discover the following:

Response	Awaiting execution	Other sentence	Total
Yes	600	300	900
No	0	600	600
Total	600	900	1,500

This additional information might lead us to revise our conclusions about the prisoners' attitudes.

The effect of nonresponses can be overcome by intensive follow-up in order to try to reduce them to, say, less than 5 percent, unless this percentage includes potential respondents who may have a disproportionate effect on the results (such as several very large firms that fail to respond in an economic survey).

In questionnaire surveys, we can take a random sample (say 25 to 33 percent) of the nonrespondents and attempt to interview them by telephone or in person to obtain answers to at least the key questions. The responses from this sample then can be weighted so that they are representative of all nonrespondents. The resulting overall estimate will be unbiased, but the sampling error will be increased, because responses would not have been obtained from the full sample originally selected.

Various statistical tests based on known data can be made for differences between respondents and nonrespondents to questionnaires. Examples of the types of data that can be tested for significant differences are mean or median income, educational level, mean age, and race. If there are no significant differences between the two groups, sometimes it is safe to assume that the respondents are representative of the nonrespondents, at least with respect to the characteristics we can measure. In personal interview surveys, a more experienced interviewer or supervisor can attempt to interview the persons who were not at home and to convert refusals into responses, or at least try to find out why they refused to participate in the survey. This may enable us to decide whether they have characteristics that are different from those of the typical respondents.

A major mistake is to make arbitrary substitutions for missing sample documents, nonrespondents, and the like. An example of arbitrary substitution is to take the file folder immediately following or preceding a sample folder that is missing from the file; another is to take the household next door to a sample household in which no one was home. The fallacy of this approach is that the record may be missing from the file for a reason, such as fraud or collusion. The household in which no one is home may be occupied by a single person or by a couple who are both employed and have no children. The responses of these people may be different from the responses of those who are usually at home at the time the survey is made. Also, some people may refuse to be interviewed because they are emotionally involved in the survey's subject, and these may be the very people whose answers we want.

Sometimes, random procedures are used to select substitutes for missing or unavailable records or for nonrespondents. Although this is better than arbitrarily selecting the closest available record from a file or the household next door, we are really obtaining more data about the same type of sampling unit--the households that will respond or the records that are available.

Another technique is to assume the worst or most extreme value for the missing information and see how it affects our estimates.

If nonresponse cannot be overcome or if missing records cannot be located, this should be disclosed in the published report. Also, the rate of nonresponse to individual variables should be reported. The user of the information can thus know the actual results and can evaluate them accordingly.

CHAPTER 8

THE STRENGTHS AND LIMITATIONS OF STATISTICAL SAMPLING

When a design is developed for answering an audit or evaluation question, the question arises as to whether or not to sample. If the universe is small or the individual sampling units in the universe are very important, it is often advisable to examine every item in the universe. However, if the universe is large, a sample is preferred to a complete enumeration of the universe because the information that is wanted can be obtained more cheaply, more quickly, and often more accurately and in more detail. In some instances, only one of these benefits applies, and in some extreme situations, not one does. These points deserve some explanation.

Sampling is usually cheaper than a complete review of the universe because, by definition, it usually deals with only a small group selected from the universe. The total cost of getting information includes a variable cost related to examining the individual items. By reducing the number of items to be examined, sampling permits a substantial reduction in that variable cost. However, a good sampling plan may add some costs that would not be present in a complete review. Although almost always much smaller than the savings, such costs should not be ignored. They usually cover

1. developing the sampling plan,
2. selecting the sample,
3. monitoring the sample selection process,
4. processing the data and calculating estimates and sampling errors (manually or by computer), and
5. providing special training or instructions necessary for items 2 through 4.

With regard to speed, sometimes a recommendation must be prepared or a decision made within a relatively short time. No matter how good the quality, information is of no help unless it is received in time to be used in making the recommendation or arriving at the decision. The measurement or examination process takes time; so does the summarization of results. Because a sample has fewer items than a complete review, these processes can be done more quickly in order to make them more useful to the decisionmakers.

Sometimes an attempt will be made to obtain more information from the sample than it was originally designed for. An example of this is taking a sample that was designed to evaluate the effectiveness of a runaway-youth shelter program as a whole and then attempting to develop estimates for different domains of interest (e.g., classifying the youths by sex, age, or race or the marital status of their parents). In some cases, attempts to use a sample for purposes other than those for which it was designed can lead to estimates with sampling errors as large as, or larger than, the estimates themselves.

Similarly, many believe that sampling may furnish less accurate answers than a complete analysis. We would like to assert the opposite view, even though it may not, at first glance, seem reasonable. The basis for this suggestion is the reduction in the risk of errors of measurement, recording, processing, and reporting. Because sampling involves the observation of fewer items, it frequently allows us to use personnel who have been better trained to collect, process, and evaluate the data than would be practicable in a full examination of the universe. In fact, it has been found that measuring physical inventory by sampling is more accurate than counting and pricing every item. Also, because fewer observations are needed in sampling, the measuring process can be done more nearly simultaneously and the result is more likely to present correctly the status of the universe at a given time than would the result of a complete review, during which changes may take place. For instance, by the time the last item has been measured, the first one may have been used up or materially changed.

By suggesting that sampling permits more detailed information to be obtained, we mean that if an attempt is made to measure all the items in the universe, it may be possible to make only one or a few observations on each sample item. However, if sampling is used and fewer items are measured, it may be possible to collect much more data about each item and thus develop more information about the universe. For example, the Bureau of the Census is mandated to count the population (complete enumeration) in the decennial census, but the detailed statistics on the demographic characteristics of the population and the nation's housing are developed from samples of households counted in the census.

Additionally, we may discover during data collection that we are taking the wrong measurements, that additional data are needed, or possibly that the investigation's objective has been poorly defined. If sampling is used to collect the data, as opposed to making a complete enumeration, mistakes are much easier to correct.

In sum, if statistical sampling is feasible and is carried out correctly, it usually has advantages over complete enumeration and nonstatistical sampling.

THEORETICAL BACKGROUND

This appendix explains the bases for sample estimates and the concepts of confidence and sampling error. It is intended for those who have not completed a college course in statistics or who have completed one but need a refresher.

As we noted in chapter 3, we use measurement or observations of a sample to draw inferences about the universe from which the sample was drawn. The sample is used to estimate totals, rates of occurrence, or both. Assuming that the sample mean estimates the universe mean, we multiply the sample mean by the number of items in the universe to estimate the universe total. Similarly, when estimating rates of occurrence, we assume that the rate of occurrence found in the sample estimates the rate of occurrence in the universe. Such assumptions are based on the laws of probability.

CALCULATING PROBABILITIES

Let us consider a dice-throwing experiment that uses a pair of perfect dice. In table I.1 (on the next page), we tabulate, for each point that can be thrown, the number of ways the point can be "made" and the probability of making the point in a single throw. With this table, a person can calculate the probabilities of all possible outcomes even before picking up the dice. For example, the probability of throwing 2, 3, or 12 (called "craps" or "crapping out") is 11.11 percent (2.78 + 5.55 + 2.78). The probability of throwing 7 or 11 (a "natural") is 22.2 percent (16.67 + 5.55). The probability of throwing a 7 is 16.67 percent. The probability of throwing any one of the points 5 through 9 is 66.7 percent, and the probability of throwing any one of the points 3 through 11 is 94.4 percent.

The ability to calculate probabilities for a dice-tossing experiment depends on mechanical conditions--the way the dice are constructed and roll. Similarly, it has been found that the probabilities for sampling experiments also depend on certain conditions--not mechanical but conditions of sample selection. Tabulated in table I.2 (on the next page) is the outcome of a sampling experiment in which 400 samples, each consisting of 400 beads, were drawn at random from a jar containing 20,000 beads, of which 4,000 (20 percent) were red and 16,000 (80 percent) were blue.

If each sample were an exact image of the universe, each would contain 80 red beads (20 percent of 400). Instead, as we can see, the samples vary from as few as 55 red beads (13.75 percent) to as many as 103 red beads (25.75 percent), with the remaining samples having varying quantities of red beads between those two extremes. However, most samples contain quantities of red beads that are close to the number we would expect, knowing

APPENDIX I

Table I.1

Dice-Throwing Probabilities

Point	Number of ways point can be made	Probability of making point in a single roll
2	1	2.78%
3	2	5.55
4	3	8.33
5	4	11.11
6	5	13.89
7	6	16.67
8	5	13.89
9	4	11.11
10	3	8.33
11	2	5.55
12	1	2.78
Total	36	99.99%

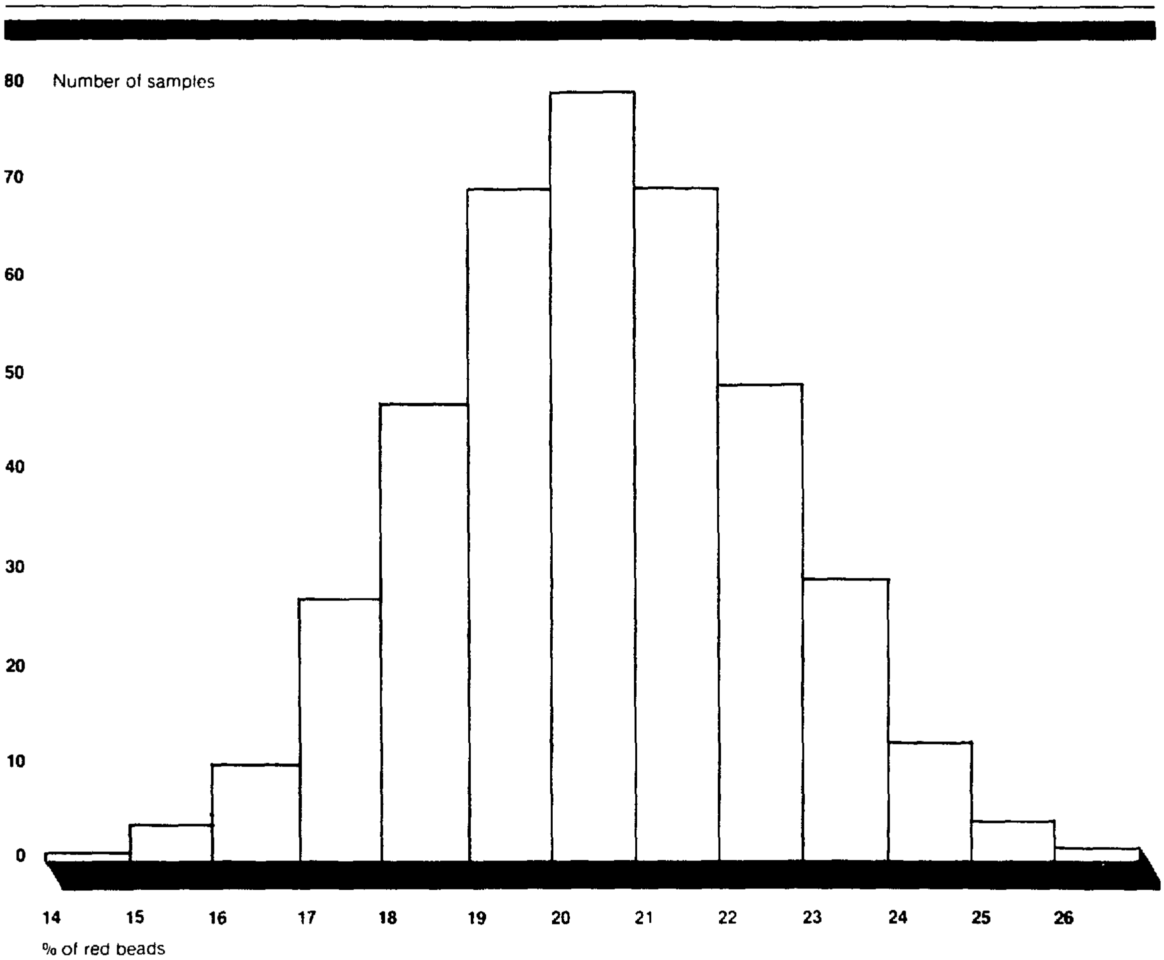
Table I.2

Results of Bead Sampling Experiment

Number of samples each with 400 beads	Red beads in sample: Number	Red beads in sample: %	Number of samples each with 400 beads	Red beads in sample: Number	Red beads in sample: %
1	55	13.75	23	80	20.00
1	58	14.50	23	81	20.25
2	59	14.75	20	82	20.50
1	61	15.25	16	83	20.75
2	62	15.50	19	84	21.00
2	63	15.75	14	85	21.25
4	64	16.00	9	86	21.50
2	65	16.25	11	87	21.75
6	66	16.50	14	88	22.00
8	67	16.75	15	89	22.25
7	68	17.00	6	90	22.50
6	69	17.25	9	91	22.75
10	70	17.50	7	92	23.00
14	71	17.75	6	93	23.25
12	72	18.00	4	94	23.50
11	73	18.25	2	95	23.75
19	74	18.50	2	96	24.00
15	75	18.75	4	97	24.25
17	76	19.00	1	98	24.50
18	77	19.25	1	100	25.00
15	78	19.50	2	101	25.25
18	79	19.75	1	103	25.75

what we do about the universe. The two categories in which the largest number of samples (23) fell contain 80 and 81 red beads, or 20 and 20.25 percent, respectively. Further, 274, or 68.5 percent, of the samples contain between 18 and 22 percent red beads and 382, or 95.5 percent, of the samples contain between 16 and 24 percent red beads. The results of the samples are shown in the form of the bar chart in figure I.1. As can be seen, the samples are arranged almost symmetrically about the category that contains the true percentage of red beads.

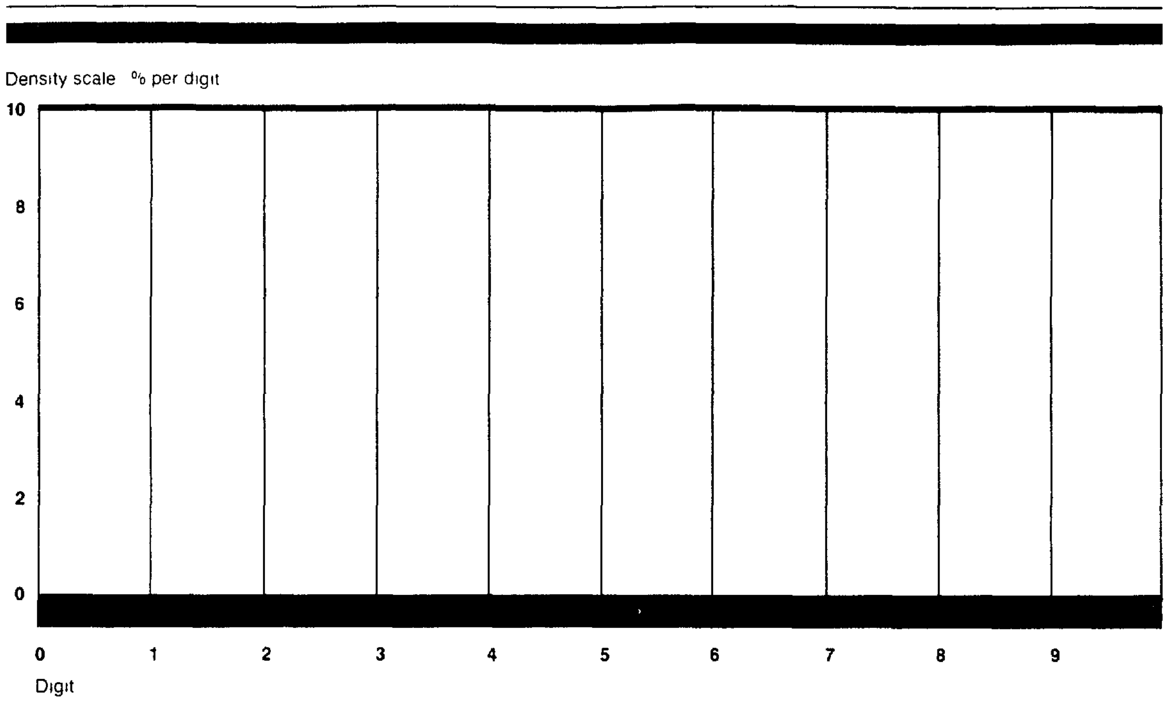
Figure I.1: Number of Samples Classified by Percentage of Red Beads



Effect of sampling variation

We might view the percentage of red beads in a sample of 400 beads as a shot at knowing the percentage of red beads in the jar, but we would have to understand that the shot is affected by sampling variation. Figure I.1 indicates the confidence that should be associated with each level of precision for this shot. For instance, the shot (the sample percentage of red beads) should be between 18 and 22 percent with about 68-percent confidence. That is, the sample percentage of red beads will "likely" fall within 2 percentage points of the universe percentage of red beads. The shot should be between 16 and 24 percentage points with about 95-percent confidence. That is, the sample percentage of red beads will "very likely" fall within 4 percentage points of the universe percentage of red beads.

Figure I.2: The Relative Frequency of a Large Population of Single Digits Between 0 and 9



If we look at the red beads as ones and the blue beads as zeros, the percentage of red beads in the universe or sample is the mean (in percentage terms) of the ones and zeros. The lesson from figure I.1 is that a sample mean of a large random sample from the jar is a reliable shot at the universe mean.

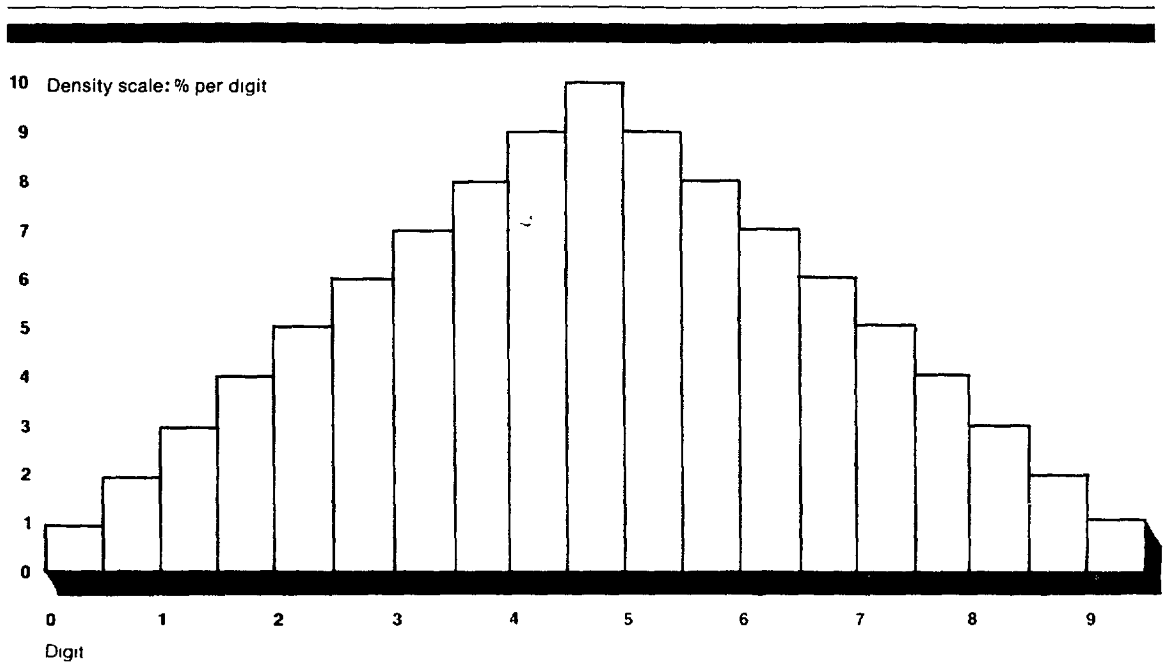
We can illustrate the situation in which we are sampling for variables by considering a very large universe of single digits. Each of the 10 digits (0 to 9) constitutes 10 percent of the universe. We summarize this universe in figure I.2 by setting each digit as a separate class interval.[1] The universe mean is 4.5; note that the arithmetic mean is the point of balance, if the chart is a physical object. A random sample of a single item from this universe gives no useful information about the mean. This one item could with equal chance be any of the 10 digits.

A sample of two elements (n = 2) provides more information about the universe mean. To see this, refer to the histogram in figure I.3 for the means of samples of two elements from this universe.

---

[1]We are indebted to Donald T. Gantz of the Department of Mathematical Sciences, George Mason University, Fairfax, Virginia, for developing this example.

Figure I.3: The Relative Frequency of Means of Samples of Two from a Population of Digits Between 0 and 9



Interval	Density	Interval	Density
0	1	5.0	9
0.5	2	5.5	8
1.0	3	6.0	7
1.5	4	6.5	6
2.0	5	7.0	5
2.5	6	7.5	4
3.0	7	8.0	3
3.5	8	8.5	2
4.0	9	9.0	1
4.5	10		

If we use simple random sampling, all the two-item samples have an equal chance of being selected. Of the 19 possible values (0, 0.5, 1, 1.5, ..., 8.5, 9) for the sample mean of a sample of 2 items, the most likely value is 4.5 (the universe mean).
The chance that the sample mean will be 4.5 is 10 percent, and the chance that it will be in the range from 4 to 5 is 28 percent. Correspondingly, each of the extreme values 0 and 9 has only a 1-percent chance of being the mean of a sample of two items. Note the strong centralizing effect of averaging only two randomly selected items from the universe.

Now, consider depictions of the means of random samples of three items from the universe, as shown by the frequency data here and the histogram in figure I.4.

Interval	Density	Interval	Density	Interval	Density
0	0.1	3.000	5.5	6.000	5.5
0.333	0.3	3.333	6.3	6.333	4.5
0.667	0.6	3.667	6.9	6.667	3.6
1.000	1.0	4.000	7.3	7.000	2.8
1.333	1.5	4.333	7.5	7.333	2.1
1.667	2.1	4.667	7.5	7.667	1.5
2.000	2.8	5.000	7.3	8.000	1.0
2.333	3.6	5.333	6.9	8.333	0.6
2.667	4.5	5.667	6.3	8.667	0.3
				9.000	0.1

Figure I.4: The Relative Frequency of Means of Samples of Three from a Population of Digits Between 0 and 9

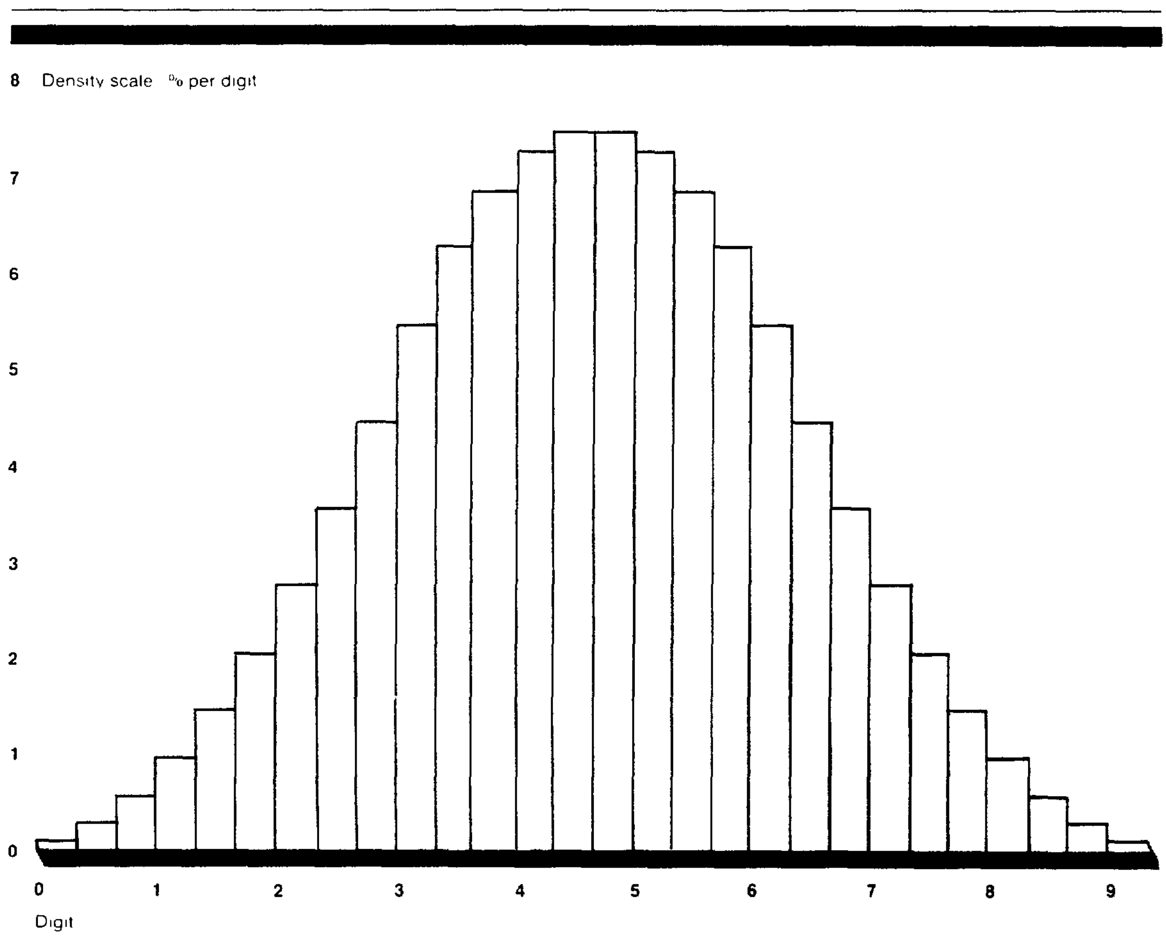
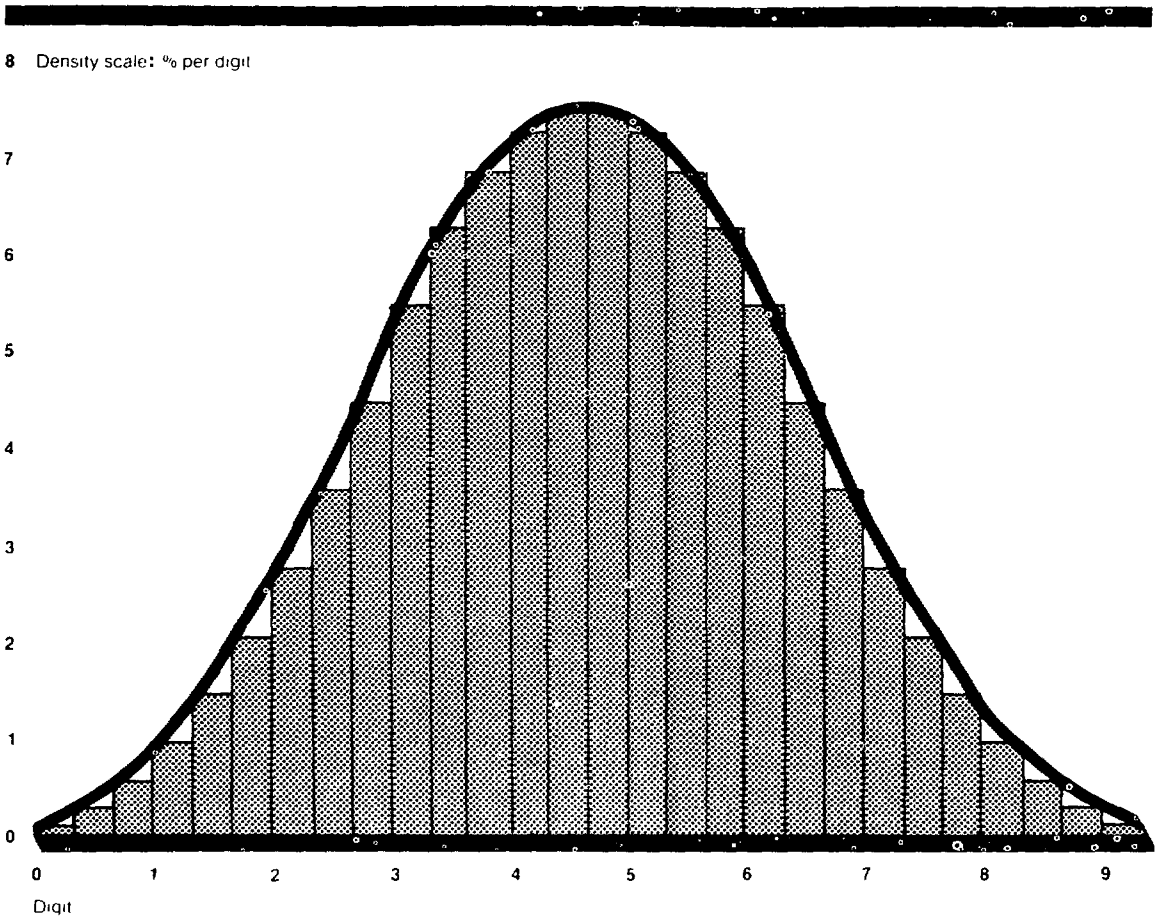


Figure I.5: The Smoothed Curve of the Relative Frequency of Means of Samples of Three from a Population of Digits Between 0 and 9



If we use simple random sampling, we find that all the possible 3-item samples have the same chance of being selected. Figure I.5 presents the same information as figure I.4 but a smooth curve gives the heights (i.e., the density) of the rectangles over the class intervals. Of the 28 possible values (0, 0.333, 0.667, 1, ..., 8.667, 9) for the sample mean of a 3-item sample, 4.333 and 4.667 (the values closest to 4.5) are most likely. There is about a 30-percent chance that the sample mean will be in the range from 4 to 5. Correspondingly, of the extreme values 0 and 9, each has only a 0.1-percent chance of being the mean of a 3-item sample. Further, the two extreme values 0 and 1 (as well as the tail values 8 and 9) have only a 3.5-percent chance of being the mean of a 3-item sample.

Central limit theorem

As the sample size, n, increases, the approximating smooth curve, analogous to the one in figure I.5, approaches the normal distribution density curve with mean 4.5 and standard deviation $2.87/\sqrt{n}$. This tendency of the histogram of the means of all simple random samples from any universe to approximate a normal density curve, if the sample size n is sufficiently large, is called the "central limit theorem."

To illustrate the power of the central limit theorem, we can note that the normal density curve has the following two properties: 68 percent of the area under the curve is within one standard deviation of the mean (or the center of the curve), and 95 percent is within two standard deviations of the mean.

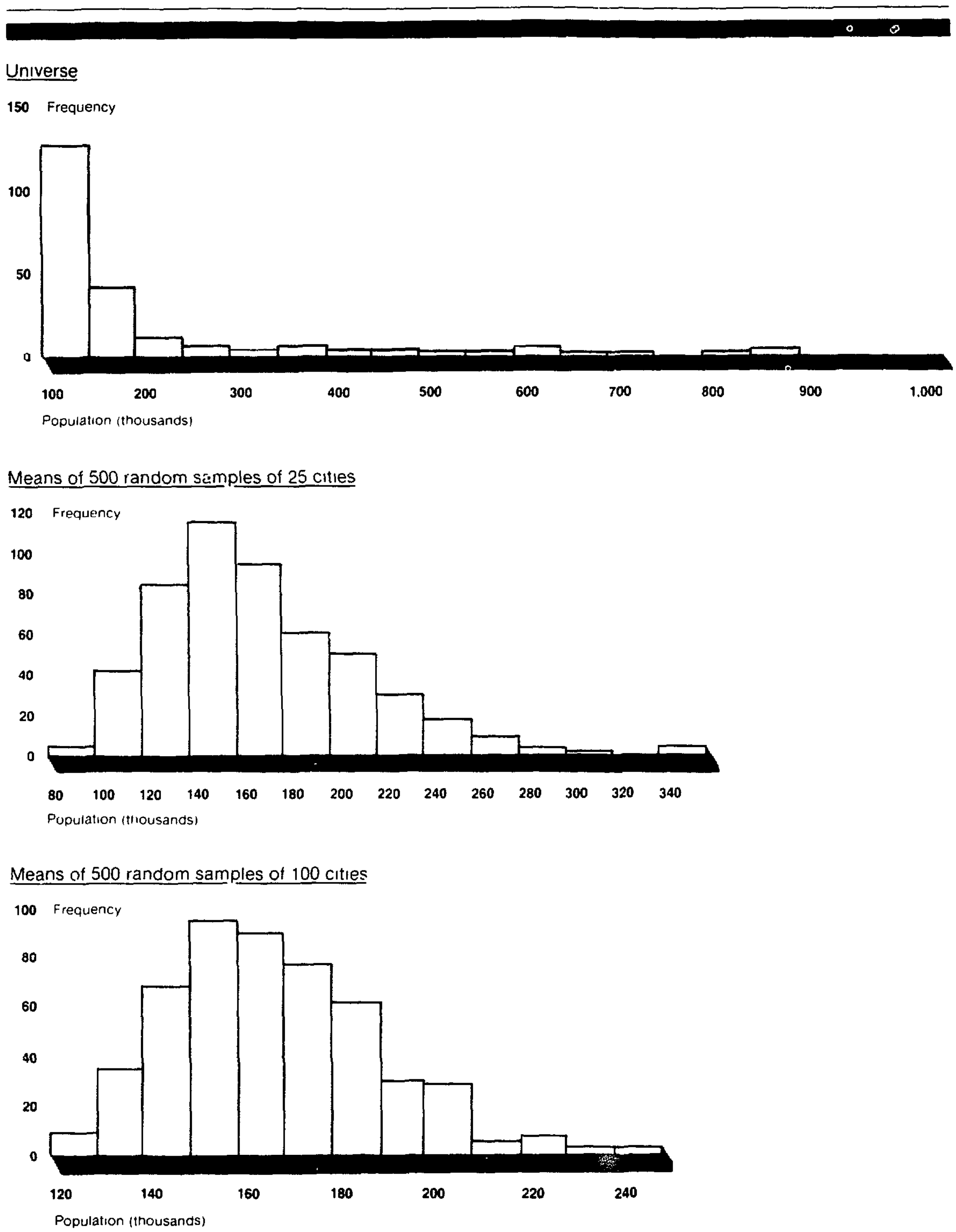
Further, note that the normal curve that approximates the sample mean histogram for a large n has a standard deviation of $2.87/\sqrt{n}$. Here, 2.87 is the standard deviation of the universe histogram in figure I.2. Hence, when n is large, the standard deviation of the sample mean curve (analogous to figure I.5) is quite small; this says that the sample mean of a large random sample will almost certainly be quite close to the mean of the approximating normal curve, which is also the universe mean. For example, if n equals 36, and the normal curve is therefore a very good approximation of the sample mean curve in the current example, then 95 percent of the possible samples of 36 will have sample means between 3.543 and 5.457 [$4.5 - (2)(2.87)/\sqrt{36}$ and $4.5 + (2)(2.87)/\sqrt{36}$].

To further illustrate the central limit theorem behavior, we can consider the frequency distribution of 228 cities whose populations were more than 50,000 in 1950, as shown in figure I.6. (The four largest cities have been excluded.) This frequency distribution is definitely not symmetrical. Technically, it is called a "reverse J-shaped distribution." The smallest class, cities with populations 50,000 to 100,000, contains more cities than all the other classes combined. One city with 1,850,000 inhabitants is not even shown on this histogram because it would require making the horizontal axis twice as long.

What happens if we take 500 random samples, each consisting of 25 cities, from this universe of 228 cities and prepare a histogram of the distribution of the sample means (also shown in figure I.6)? We note that the distribution, although by no means symmetrical, is approaching symmetry about the true universe mean. This is remarkable, considering the shape of the original universe distribution.

If we take random samples of 100 cities each and prepare a frequency distribution of the means (as in the third graph in figure I.6), the distribution shows some additional improvement in the direction of symmetry, and the sample means cluster more closely about the true universe mean, although the distribution is certainly not normal. If the city with 1,850,000 persons had been removed from the universe, the distribution of the sample means for a sample of 100 would be more nearly normal. Thus, regardless of the shape of the original universe distribution, the distribution of sample means approaches normality as the sample size increases.

Figure I.6: Frequency Distributions of the Means of Samples Drawn from a Universe of 228 U.S. Cities with Populations Greater than 50,000 in 1950[a]



[a] The four largest cities are excluded

Source Adapted from G W Snedecor and W G Cochran, Statistical Methods, 7th ed (Ames, Iowa Iowa State Univ Press, 1980)

CALCULATING THE STANDARD DEVIATION AND SAMPLING ERROR

The standard deviation of a distribution of sample means represents the expected differences between the true mean and the sample means because of variation from sample to sample. It is called the "standard error" (or "sampling error") in order to distinguish it from the standard deviation of the individual universe values. (Note that this is strictly a mathematical concept. It does not imply that there has been a human mistake or mechanical failure in some operation.) The standard error is related to the variation in the universe distribution. To show how this relationship works, we need to develop further the concept of the standard deviation.

As we stated in chapter 3, the standard deviation is merely a numerical measurement of the dispersion of a group of values about their arithmetic mean. Although other measures of dispersion exist, the standard deviation has the advantage that it can be manipulated arithmetically--that is, multiplied and divided (but in general not added or subtracted).

The theoretical formula for computing the standard deviation of a universe S is shown below. Let

$$\sum^{N}$$

represent the sum over the entire universe, and let $\bar{Y}$ represent the universe mean computed by the formula

$$\bar{Y} = \frac{\sum^{N} y_i}{N}$$

Then,

$$S = \sqrt{\frac{\sum^{N} (y_i - \bar{Y})^2}{N}}$$

What this formula tells us is that we take each value in turn, subtract the mean from it, square the difference, then add all the squared differences, divide by the universe size, and finally take the square root of the quotient. Thus, the standard deviation is the square root of the average squared deviation from the mean.

However, after some algebraic manipulation, the formula can be simplified to

$$S = \sqrt{\frac{\sum^{N} y_i^2}{N} - \bar{Y}^2}$$

Thus, it is necessary only to square the values and the mean, eliminating the calculation of the differences. For example, we can calculate the standard deviation of the universe of 100 small purchase orders, discussed in chapter 3. A work sheet to compute the sum of the values, the squares of the values, and the sum of the squares is illustrated in table I.3.

Table I.3

Work Sheet for Computing the Mean and Standard Deviation of Universe Data

Purchase No. (i)	Purchase Amount (y_i)	y_i^2	Purchase No. (i)	Purchase Amount (y_i)	y_i^2	Purchase No. (i)	Purchase Amount (y_i)	y_i^2
1	$157	$24,649	34	$194	$37,636	67	$205	$ 42,025
2	147	21,609	35	197	38,809	68	226	51,076
3	259	67,081	36	195	38,205	69	236	55,696
4	152	23,104	37	277	76,729	70	250	62,500
5	144	20,736	38	74	5,476	71	203	41,209
6	187	34,969	39	217	47,089	72	161	25,921
7	192	36,864	40	215	46,225	73	152	23,104
8	189	35,721	41	237	56,169	74	202	40,804
9	165	27,225	42	184	33,856	75	143	20,449
10	88	7,744	43	176	30,976	76	169	28,561
11	166	27,556	44	169	28,561	77	218	47,524
12	192	36,864	45	184	33,856	78	89	7,921
13	190	36,100	46	142	20,164	79	248	61,504
14	185	34,225	47	177	31,329	80	160	25,600
15	164	26,896	48	80	6,400	81	175	30,625
16	279	77,841	49	191	36,481	82	224	50,176
17	230	52,900	50	231	53,361	83	159	25,281
18	150	22,500	51	178	31,684	84	138	19,044
19	297	88,209	52	125	15,625	85	158	24,964
20	199	39,601	53	172	29,584	86	194	37,636
21	187	34,969	54	159	25,281	87	228	51,984
22	137	18,769	55	225	50,625	88	187	34,969
23	261	68,121	56	241	58,081	89	177	31,329
24	132	17,424	57	177	31,329	90	190	36,100
25	159	25,281	58	232	53,824	91	178	31,684
26	218	47,524	59	170	28,900	92	164	26,896
27	200	40,000	60	123	15,129	93	135	18,225
28	134	17,956	61	175	30,625	94	147	21,609
29	259	67,081	62	169	28,561	95	96	9,216
30	125	15,625	63	192	36,864	96	187	34,969
31	204	41,616	64	193	37,249	97	256	65,536
32	177	31,329	65	233	54,289	98	163	26,569
33	172	29,584	66	198	39,204	99	181	32,761
						100	59	3,481
							$18,257	$3,532,617

We will first compute the universe mean, Y:

$$\bar{Y} = \frac{18,257}{100}$$

$$\bar{Y} = 182.57$$

Thus, the universe mean is $182.57. Next, we will compute the universe standard deviation:

$$S = \sqrt{\frac{3{,}532{,}617}{100} - (182.57)^2}$$

$$S = 44.66$$

As can be seen, the standard deviation of these values is $44.66.

The standard deviation of the sample means, or the standard error ($E_{\bar{y}}$), can be calculated easily from the universe standard deviation. The computation (assuming sample size n equals 30) is

$$E_{\bar{y}} = \frac{S}{\sqrt{n}}$$

$$E_{\bar{y}} = \frac{44.66}{\sqrt{30}}$$

$$E_{\bar{y}} = 8.154$$

Even before a large sample is drawn, we can say that the chances are about 2 in 3, or the probability is approximately 68 percent, that the sample mean will lie within one standard error of the true mean. Likewise, we can say before the sample is drawn that the chances are approximately 19 in 20, or the probability is approximately 95 percent, that the sample mean will lie within two standard errors of the true mean. In general, these statements will be true regardless of the shape of the universe distribution, if the sample size is at least 30 items.

The question that may be asked is, What good is this? The evaluators do not know what the true mean and standard deviation are. If they did, they would not have to sample. The answer is that we take a sample from the universe of values and calculate the sample mean and sample standard deviation. We then let the sample standard deviation represent the universe standard deviation to calculate the sampling error. In chapter 3, we calculated a standard deviation of $48.71 for our sample of 30 small purchase orders. The estimated sampling error is then computed:

$$E_{\bar{y}} = \frac{48.71}{\sqrt{30}}$$

$$E_{\bar{y}} = 8.89$$

Thus, the sampling error of the mean, estimated from the sample, is $8.89. Note that the size of the sampling error is inversely proportional to the square root of the sample size. The larger the sample, the smaller the sampling error.

Once we have computed the sampling error, we can say the estimated mean calculated in chapter 3 is $182.30 with a sampling error of $8.89. How closely does this estimate the true mean? We know that if a large number of samples is taken, approximately 68 percent of the sample means will be within one sampling error of the true mean. Thus, the probability is approximately 68 percent that our sample mean is within one sampling error of the true mean. From these statements, we can infer that the true mean is within one sampling error of the sample mean.

The sample gives us an estimate of the true mean, and the sampling error tells us how precise the estimate is. Although we can say before the sample is drawn that the probability is 95 percent that a sample mean will be within two sampling errors of the true mean, we cannot say the probability is 95 percent that the true mean is within two sampling errors of the sample mean after drawing the sample and computing the estimates. Either the true mean is included within the interval given above, in which case the probability is 1, or the true mean is not included in the interval, in which case the probability is zero.

Statements such as "68 percent of the sample means are within one sampling error on either side of the true mean" and "95 percent of the sample means are within two sampling errors on either side of the true mean" are known as "confidence statements." The confidence level to be used is set by management or the evaluators, and they base it on the risk they are willing to take that the sample estimate may miss the mark. If management is willing to take a 5-percent risk of being wrong, for example, the confidence level should be set at 95 percent. If management sets the risk of being wrong at 1 percent, the confidence level should be set at 99 percent.

Table I.4

Table of t Factors

Confidence level as %	Multiplier for sampling error (t factor)
50	0.6745
68	1.000
80	1.282
90	1.645
95	1.960 (or 2)
99	2.578
99.73	3.000

The person who does the sampling then selects the proper t factor for the specified confidence level and multiplies it by the sampling error to get the maximum possible difference between the sample mean and the true mean for that confidence level. (Some t factors are shown in table I.4.) The t factors are based on the normal distribution, which describes the variability in sample means of large size samples. For this example, at 68-percent confidence the t factor equals 1, so the sampling error is $8.89. At 95-percent confidence, the sampling error equals 1.96 times $8.89, or $17.43.

A COMPREHENSIVE DESCRIPTION

OF SAMPLING PROCEDURES

This appendix, an extension of the material in chapter 6, provides a comprehensive description of sampling procedures, problems that may confront the sampler, and methods for overcoming these problems. We have attempted to make this material as comprehensive as possible, but it obviously cannot cover every situation that may be encountered nor does it cover every method of random selection.

RANDOM NUMBER SAMPLING

As mentioned in chapter 6, random number sampling is, in its simplest form, a procedure by which a quantity of random numbers equal to a specified sample size is selected from a universe of random digits, called a "table of random digits," and then matched against the serial numbers, stock numbers, transaction numbers, or the like that have been assigned to the sampling units in the universe of interest. The sampling units having numbers that correspond to the selected random numbers constitute the sample.

Although the description of the procedures to be used in various situations makes the use of random number sampling seem tedious, the work can be done quite quickly by relatively inexperienced personnel once they understand the procedures. The major pitfalls are the failure to define the universe correctly and to ensure that the sample is drawn from the entire universe.

A table of random digits is a universe of thousands of digits from 0 through 9 and is generated by electronic or electromechanical procedures designed to ensure that all the digits have an equal probability of being generated, regardless of which digit was previously generated. The random digits are printed in tables in the same order in which they are generated. Thus, we can use the tables to select any required sample with complete assurance that the sample was drawn at random.

The best known tables of random digits are

- o Interstate Commerce Commission (ICC), Bureau of Transport Economics and Statistics, Table of 105,000 Random Decimal Digits (Washington, D.C.: 1949), and
- o Rand Corp., A Million Random Digits (New York: Glencoe Free Press, 1955).

Table II.1 contains excerpts from the ICC publication. Note that the digits are printed in groups of 5 with 14 columns to the page and that each line has a unique number. (In the whole table, the line numbers go from 1 to 1,500.) Thus, any group of five digits can be located by the line number and the column number.

Table II.1

Random Decimal Digits

Line \ Col	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
801	33993	51249	78123	16507	57399	77922	38198	63494	00278	30782	33119	64943	17239	69020
802	39041	05779	74278	75301	01779	60768	22023	07510	67883	55288	67391	54188	31913	29733
803	56011	26839	38501	03321	43259	73148	43615	49093	91641	77179	50837	48734	85187	41210
804	07397	95853	45764	43803	76659	57736	44801	45623	23714	69657	87971	24757	94493	78723
805	74998	53337	13860	89430	95825	65893	96572	73975	19577	87947	23962	78235	64839	73456
806	59572	95893	69765	43597	90570	60909	06478	77692	30911	08272	81887	57749	02952	51524
807	74645	13940	28640	00127	04261	17650	34050	78788	57948	36189	88382	72324	59253	30258
808	42765	23855	38451	11462	32671	52126	23800	02691	57034	34532	19711	71567	90495	55980
809	66561	56130	30356	54034	53996	98874	78001	29707	91938	72016	16429	69726	41990	33673
810	50670	13172	31460	20224	34293	59458	24410	01366	68825	22798	52873	18370	15577	63271
811	53971	08701	38356	36149	10891	05178	55653	31553	20037	39346	28591	13505	04446	92130
812	47177	03085	37432	94053	87057	61859	97943	81113	62161	11369	54419	58886	89956	12857
813	41494	89270	48063	12253	00383	96010	41457	54657	46881	75255	29242	07537	53186	95083
814	07409	32874	03514	84943	74421	86708	34267	66071	62262	99391	61245	95839	75203	93984
815	03097	12212	43093	46224	14431	15065	18267	60039	62089	38572	70988	17279	05469	28591
816	34722	88896	59205	18004	96431	41366	50982	92400	59369	43605	26404	04176	05106	08366
817	48117	83879	52509	29339	87735	97499	42848	81449	80024	81312	59469	91169	70851	90165
818	14628	89161	66972	19180	40852	91738	23920	75518	32041	13411	61334	52386	33582	72143
819	61512	79376	88184	29415	50716	93393	96220	82277	64510	43374	09107	28813	41848	08813
820	99954	55656	01946	57035	64418	29700	99242	42586	11583	82768	44966	39192	82144	05810
821	61455	28229	82511	11622	60786	18442	36508	98936	19050	57242	33045	54278	21720	87812
822	10398	50239	70191	37585	98373	04651	67804	84062	27380	75486	63171	24529	60070	66939
823	59075	81492	40669	16391	12148	38538	73873	68596	25538	83646	61066	45210	24182	18687
824	91497	76797	82557	55301	61570	69577	23301	31921	09862	73089	69329	41916	41165	34503
825	74619	62316	80041	53053	81252	32739	65201	92165	93792	30912	59105	76944	70998	00317
826	12536	80792	44581	12616	49740	86946	41819	85104	25705	92481	95287	61769	29390	05764
827	10246	49556	07610	59950	34387	70013	64460	96719	43056	24268	23303	19863	43644	76986
828	92506	24397	19145	24185	24479	70118	42708	54311	95989	08402	77608	98356	47034	01635
829	65745	27223	22831	39446	65808	95534	03348	11435	24166	62726	99878	59302	81164	08010
830	01707	04494	48168	58480	74983	63091	81027	72579	67249	48089	34219	71727	86665	94975
831	66959	80109	88908	38757	80716	36340	30082	43295	37551	18531	43903	94975	31049	19033
832	79278	02746	50718	90196	28394	82035	03255	39574	41483	12450	32494	65192	54772	97431
833	11343	22312	41379	22297	71703	78729	65082	57759	79579	41516	46248	37348	34631	88164
834	40415	10553	65932	34938	43977	39262	95828	98617	27401	50226	17322	44024	23133	57899
835	72774	25480	30264	08291	93796	22281	51434	66771	20118	00502	07738	31841	90200	46348
836	75886	86543	47020	14493	38363	64238	16322	45503	90723	35607	43715	85751	15888	80645
837	64628	20234	07967	46676	42907	60909	73293	38588	31035	12226	37746	45008	43271	32015
838	45905	77701	98976	70056	80502	68650	24469	15574	40018	90057	96540	47174	03943	37553
839	77691	00408	64191	11006	39212	26862	99863	58155	66052	96864	61790	11064	49308	94510
840	39172	12825	43379	57590	45307	72206	53283	75882	93451	44830	06300	45456	49567	51673
841	67120	01558	99762	79752	17139	52265	97997	66806	55559	62043	51324	32423	88325	99634
842	88264	85390	92841	63811	64423	50910	38189	88183	56625	22910	58250	70491	71111	37202
843	78097	59495	45090	74592	47474	56157	88287	47032	66341	38328	70538	91105	12056	36125
844	41888	69798	82296	09312	04150	07616	34572	83202	58691	27354	37015	11278	49697	65667
845	46610	07254	28714	18244	53214	39560	68753	16825	48639	38228	95166	53649	05071	26894
846	29213	42101	25089	11881	77558	72738	57234	28458	74313	29665	97366	94714	48704	07033
847	38601	25735	04726	36544	67842	93937	68745	62979	97750	28293	75851	08362	71546	17993
848	92207	10011	64210	77096	00011	79218	52123	29841	76145	82364	55774	15462	44555	26844
849	30610	13236	33241	68731	30955	40587	45206	11949	28295	12666	98479	82498	49195	46254
850	74544	72806	62236	65685	37996	00377	59917	91100	07993	15046	51303	19515	25055	56386

Source: Interstate Commerce Commission, Bureau of Transport Economics and Statistics, Table of 105,000 Random Decimal Digits (Washington, D.C.: 1949), p. 17.

(Table II.1 continued)

Line \ Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
851	76385	05431	82252	79850	31192	86315	75612	59985	76421	39300	64976	27951	17855	02220
852	08059	15958	10514	86124	29817	19044	03555	80725	67857	31395	68780	16560	79952	41739
853	30636	03463	50326	69684	38422	59826	47858	90601	50834	88109	43882	15687	06212	19886
854	23794	51463	67574	48953	73512	46239	10953	04622	60650	35048	34705	90502	31011	81004
855	01117	60216	29314	65537	84029	00741	40851	96344	13861	43421	57107	60813	06877	52161
856	29527	19577	01414	35290	70174	37019	80223	62206	22928	63414	03940	02188	20345	13183
857	64236	24229	17970	92022	64164	17873	41189	28240	60697	26495	87634	75899	09741	84939
858	92331	30325	61918	71623	38040	51375	91127	93903	83715	93244	04366	57679	70829	90088
859	93454	37190	23790	40058	03758	01774	90696	81674	53791	15559	42798	46892	57960	06575
860	17101	42181	45798	68745	24190	16539	32330	21732	65547	94356	38651	35102	16327	17886
861	30742	93358	95730	52535	34404	76057	21325	87526	93020	94861	83865	61393	89645	00773
862	02472	01280	67106	47893	93551	76697	56598	67982	77376	33312	58893	69370	59118	95277
863	80718	72187	67178	77179	06212	37409	48788	68930	21672	88783	59304	82369	19410	93050
864	85406	73687	02116	57637	94701	46754	54019	96344	72780	47764	57490	21321	29075	40086
865	00563	67156	88141	13491	92592	35746	72117	37593	93343	78271	75915	85972	58615	71755
866	89190	58965	55213	24337	58807	36123	09235	95541	96979	03336	34380	66288	98659	46572
867	01438	81590	83758	45361	76209	65081	34785	68423	04408	73827	78494	02765	46174	83192
868	79127	53282	50510	80129	23960	78423	31988	78571	79458	95043	23997	97528	21631	63898
869	33952	92823	32840	94420	51193	69652	04332	81675	64644	68673	33718	02256	34414	87710
870	57146	14126	71734	56942	83371	31526	13444	11912	03152	66411	42853	08437	35667	26251
871	33158	61761	73207	01764	81696	55137	41834	81860	81310	14711	36599	78042	62086	41752
872	63615	69083	00118	47991	99521	88655	94451	67445	99377	75528	40794	30140	82298	85868
873	89010	46915	70186	55657	76955	25430	91951	56473	34225	68103	25353	89595	04715	20102
874	32547	43398	30909	62599	53105	27460	56734	41954	47696	82113	38508	88941	49983	36899
875	61992	28258	27359	61002	16882	44018	85376	66756	14395	66865	67036	78374	43612	44134
876	78326	74541	22198	48380	45919	76160	10974	03127	58980	18350	22089	54977	94019	84739
877	35493	53008	78622	38329	27611	12327	52541	00861	62380	65890	79729	99710	64836	43706
878	19130	59917	28850	76593	02389	80759	18481	02724	57578	35705	89265	25033	13767	95888
879	00317	05769	03497	42174	32653	23663	29569	36342	85908	29572	60063	41170	59957	14755
880	84122	36454	70776	17000	83017	07027	98058	41274	22476	27436	30798	62287	21235	00249
881	76320	32120	91585	39640	23470	86000	68204	23980	17625	53197	35128	76385	02848	61680
882	09234	36233	94404	42812	39210	25967	12232	38195	16649	96739	64610	96067	89561	15772
883	16206	70598	95378	70573	42636	53862	81334	65439	28858	07619	59608	61460	00581	43226
884	04071	51662	67884	73911	08708	66287	89261	73451	81146	77733	70162	42449	44755	56401
885	97545	87732	83795	38027	90239	80044	64677	47912	82144	85918	93508	05816	57549	74831
886	53253	56120	42720	25660	36921	30891	42042	80370	97880	62507	01218	19202	22323	81363
887	66817	18439	53188	35155	24309	88284	74644	25454	19606	61460	52684	36568	68108	45653
888	28077	26409	11443	22200	23129	32407	52401	78416	63693	35633	77724	86835	89829	81383
889	18889	00291	13701	12401	26466	67700	55805	63818	16067	95185	97241	66126	16774	39342
890	10598	64974	66296	33329	30560	73380	94905	04959	80213	14228	97242	94826	64276	37466
891	18656	81152	45498	14400	92435	67664	86229	74358	76537	87066	42293	91743	49462	42808
892	79044	10440	25777	05486	65659	22183	82080	04351	19530	49941	29181	34667	28910	70927
893	74042	20365	42672	34850	60670	56980	88333	75288	64996	26913	62379	55068	91239	85752
894	87249	06640	09090	03242	68467	85678	23411	53443	19526	03205	29261	36061	34325	03761
895	82839	52537	00518	60559	43669	44297	75071	17146	35492	60718	38106	06409	75657	66013
896	07749	62249	01611	43795	17129	44447	95197	25088	22150	54427	11578	77560	26460	55002
897	37171	34598	87234	28324	85927	23465	80833	62872	40826	10066	64858	33605	24848	30881
898	55432	45030	85336	49128	40487	63959	25879	60415	26744	38584	51543	17333	47300	13834
899	43658	35437	83506	11209	24770	87123	21494	85056	56630	75919	26005	90077	50380	09261
900	59536	79475	04874	50831	16996	04750	02246	08846	82410	50997	45824	55547	08168	16679

The digits are printed in groups of five merely to make the table easier to read. The evaluator may read down the columns or across the rows, whichever is more convenient. If the number to be selected consists of fewer than five digits, the evaluator may read from either the left or right edge of the group of five. If the numbers to be selected consist of more than five digits, the digits in one group of five can be combined with the digits from one or more adjacent groups to form a number with the required quantity of digits. It is extremely important that decisions about the combination of digits, about whether to read from the left of the group or the right, and about what to do when the bottom of the page is reached be made before the selection process is begun and followed consistently throughout.

Locating a starting point in a random decimal digit table

Take care to avoid starting at the same place each time the table of random digits is used and to avoid a purposive selection of random numbers. There are two basic methods of locating a random starting place in the table.

1. The simplest method is to start at the beginning of the table the first time it is used and mark lightly through each group of digits read. The next time the table is used, start with the first group of digits immediately following the last group lined through. Continue this process until the entire table is used up. At that point, reading can start again at the beginning of the table.

2. More complex is the so-called "random stab" method, in which the table pages are allowed to fall open and, without looking, the evaluator stabs the page with a pencil point and begins with the digit closest to it.

A refinement of the random-stab method is to locate the starting point by two stabs of the pencil. On the first random stab, read down the column closest to the pencil point, reading either the four lefthand digits or the four righthand digits, until a number between 0001 and 1,500 is reached. This becomes the line number of the starting place. Then allow the table to fall open again at random, make a second random stab, and again read down the column closest to the pencil point, reading either the lefthand digits or the two righthand digits, until a number between 01 and 14 is reached. This becomes the column number of the starting point.

Determining the proper quantity of digits to read

Random number sampling can be used most conveniently if the sampling units in the universe are already numbered or can be numbered easily. However, actual physical numbering of the sampling units is not necessary if their location can be established by counting (e.g., lines in a ledger or folders in a file drawer).

Suppose you want to select a sample of 50 documents from a universe of documents numbered from 1 through 360. After locating the starting point in the table and deciding which way to proceed through the table and whether to read the digits from the left or the right, record the first 50 numbers from 001 through 360. Note that because you are selecting three-digit numbers, you must always read three digits at a time. The lowest number eligible for inclusion in the sample is the three-digit group 001. Disregard the number 000 and all numbers greater than 360. If you are sampling without replacement--that is, including an item in the sample only once (which is usually the case in GAO work)--do not use random numbers that duplicate a number previously selected. Instead, use the next available random number from 001 through 360.

If the 360 documents in the universe are not numbered but a numbering system can be established by counting the documents, almost the same procedure can be used. The only difference is that the random number selected corresponds not to the document's number but to its location in the file, ledger, or list (e.g., the 3rd, 8th, or 25th).

To select a sample of 100 items from a universe numbered 458 through 15,936, select the first 100 five-digit numbers from 00458 through 15,936. Numbers less than 00458 or from 15,937 to 99,999 and numbers that duplicate previously selected numbers are disregarded.

Examples of random number sampling

Suppose you want to select 20 numbers between 1 and 89. Assume that the starting point is page 17, line 836, column 6, of the ICC table. Because 89, the highest number, is a two-digit number, you must read, two digits at a time, all combinations of digits between 01 and 89. If you choose to read the left pair of digits in the five-digit groups, start with 64238, read down the column to the bottom of the page, go to the top of column 7 and down that column to line 809, and finish with group 78001. Thus, you read numbers 64, 60, 68, 26, 72, 52, 50, 56, 07, 39, 72 (not used because already selected), 93 (not used because outside the specified range), 79, 40, 00 (not used because outside the specified range), 38, 22, 43, 44, 96 (not used because outside the specified range), 06, 34, 23, and 78.

If you choose to read the right pair of digits and proceed through the table in the same manner, you read numbers 38, 09, 50, and so on.

Suppose you want to select 150 vouchers from a group of vouchers numbered 23,427 through 28,965. Two possible methods of simplifying the random number selection in this type of situation follow:

1. Subtract the lowest number (23,427) from the highest number (28,965), obtaining 5,538. Select 150 random numbers from 0000 through 5,538, skipping all numbers from 5,539 through 9,999. Add 23,427 to each of the 150 random numbers selected to obtain the numbers of the sample vouchers.

2. Select 150 four-digit numbers from 3,427 through 8,965. Add 20,000 to each number selected to obtain the sample voucher numbers.

Recording selected random numbers

Random numbers should be recorded in a manner that will simplify arranging them in numerical sequence for identifying duplicate numbers and locating the sample items. The best way to do this is to record the random numbers on index cards or on a work sheet like the one in figure II.1

Using index cards

To use index cards, number the cards in advance, from 1 through the total quantity of items required for the sample, plus some extra cards to substitute for duplicate random numbers. These assigned numbers become the order of selection of the random numbers. As each eligible random number is read from the table, it is entered on one of the numbered index cards. The cards are kept in the order of their assigned numbers, and the extra cards are kept separate from the cards for the original sample and in their order of selection. After random numbers have been entered on all the index cards, sort the original cards into the numerical sequence of the random numbers to help identify duplicates. When a duplicate card is discovered, insert the extra card with the lowest previously entered number in the pack of original cards in the proper sequence of its random number, if it does not duplicate a random number in the original pack. If the random number on the extra card duplicates a random number previously selected, substitute the next extra card in order of selection. The duplicate original card can be either removed from the pack or retained and clearly identified as a duplicate. Repeat this procedure until all duplicates in the original pack have been identified and replaced by extra cards.

You will now have a numerically sequenced set of unduplicated random numbers equal to the specified sample size. Furthermore, since the order in which each random number was selected is also entered on the card, arranging the cards into their selection

Figure II.1: Random Number Selection Work Sheet

Random Number Selection Work Sheet

Source: ______________________________

Start: Page____, Line____, Col.____. Stop: Page____, Line____, Col.____

Thousands	Hundreds									
	0	1	2	3	4	5	6	7	8	9
0										
1										
2										
3										
4										
5										
6										
7										
8										
9										

order, if necessary, is a simple matter. Any remaining extra cards can be discarded or retained for later use if ineligible items or refusals come up in the sample or other problems occur.

Using a work sheet

Another system for recording random numbers in order to help sort them into numerical order and identify duplicates is to enter the numbers as they are selected on a work sheet like the one in figure II.1. This work sheet will accommodate random numbers from 0 through 9,999.

The work sheet should identify as the "source" the title of the table or document from which the random numbers were selected. Also, the work sheet heading should show the places in the table of random digits where selection began and ended. The row captions identify the digits in the thousands position, and the column headings show the digits in the hundreds position. Enter the random numbers read from the table into the boxes formed by the intersection of the appropriate rows and columns. For example, enter random numbers between 0000 and 0099 in the box formed by the intersection of row 0 and column 0; random numbers between 0700 and 0799, in the box formed by the intersection of row 0 and column 7; and random numbers between 8,000 and 8,099, in the box formed by the intersection of row 8 and column 0. Since the first two digits of every number entered on the work sheet can be easily determined from the row and column numbers, only the last two digits of each number need be entered.

If you use the work sheet as shown in figure II.2, duplicate random numbers can be easily identified as soon as they are recorded. Circle such duplicates to indicate that they duplicate a random number previously selected and are not to be used.

This type of work sheet can be adapted to record three-digit random numbers by letting the rows define the digits in the hundreds position and the columns define the digits in the tens position. For recording two-digit random numbers, the row can be eliminated and the columns used to define digits in the tens position.

We can illustrate the use of the work sheet for the selection of 50 random numbers from 3,427 through 8,965. The starting point is page 17, line 801, column 7, of the ICC table. Select eligible random numbers from the four lefthand digits in each column, reading down the column to the bottom of the page and then to the top of the next column, and so on.

Note that since only numbers between 3,427 and 8,965 are eligible for selection, the work sheet does not have rows 0, 1, 2, and 9 and that columns 0 through 3 of row 3 have been crossed off. This is to prevent needless writing. When numbers less than 3,400 or greater than 8,999 are encountered in the table, they

Figure II.2: Random Number Selection Work Sheet Used for Selecting 50 Random Numbers

Random Number Selection Work Sheet

Source: ____________________

Start: Page____, Line____, Col.____. Stop: Page____, Line____, Col.____

Thousands	Hundreds									
	0	1	2	3	4	5	6	7	8	9
3					~~05~~ ~~26~~ 57		50		19 18	57
4		45 81	84 70 58	61 29	80	20 62				09
5	98	43	12	28	65 31	65		23		91
6	03			49	46	20 08	07	80	75 74 59	
7			57	87 29 97		51		69	00 78	
8		02 11 44	27		06	10			28	

should be ignored. Numbers between 3,400 and 3,426 and between 8,966 and 8,999, if encountered, may be entered in the proper boxes and lined through later to indicate they are outside the range of eligible numbers. (They need not be recorded at all, but there is little point in remembering the exact cutoff points while reading numbers from the table.)

In reading through the table, the first set encountered is 3,819. This number is within the specified range and is recorded by writing 19 in the box formed by the intersection of row 3 and column 8. The next number is 2,202; it is obvious that this number is outside the specified range, because this work sheet does not have a row 2. The next number is 4,361, which is recorded by writing 61 in row 4, column 3. This procedure is continued until 50 eligible random numbers have been selected. Note that in row 3, numbers 3,405 and 3,426 are lined through because they are less than the lowest eligible number, which is 3,427.

If you had read any duplicates from the table, you would have recorded them in the appropriate box and circled them to identify them as duplicates. The random numbers can be sorted into ascending numerical order by simply sorting the relatively few entries in each box in ascending order.

Special problems in random number sampling

Certain situations present special problems in random sampling. If these problems are anticipated through careful research into the characteristics of the numbering system, they can usually be resolved with little difficulty.

Gaps in the numbering system

Occasionally, a numbering system has gaps in it; that is, certain blocks of numbers are not used. When selecting random numbers from a universe numbered in such a manner, simply ignore random numbers that correspond to the gaps. These are the equivalent of out-of-range numbers.

For example, assume that a sample of 20 documents is to be drawn from a universe of documents numbered from 1 to 95. However, the numbers 1 through 9, 25 through 29, and 40 through 49 are assigned to a class of documents not to be reviewed. For all practical purposes, this universe really consists of documents numbered from 10 through 24, 30 through 39, and 50 through 95.

Selection starts on page 17, line 812, column 7, of the ICC table. If we read the two righthand digits, going down the column, the following numbers are obtained: 43 (ineligible), 57, 67, 67 (duplicate), 82, 48 (ineligible), 20, 20 (duplicate), 42 (ineligible), 08 (ineligible), 04 (ineligible), 73, 01 (ineligible), 01 (ineligible), 19, 60, 08 (ineligible), 48 (ineligible), 27 (ineligible), 82 (duplicate), 55, 82 (duplicate), 28 (ineligible), 34, 22, 93, 69, 63, 83, 97 (ineligible), 89, 87, 72, 53, 34 (duplicate), 45 (ineligible), and 23. The sample would consist of documents 19, 20, 22, 23, 34, 53, 55, 57, 60, 63, 67, 69, 72, 73, 82, 83, 87, 89, 93, 94.

Caution should be taken in that gaps in the numbering system sometimes indicate that various groups of items have been assigned different blocks of numbers because the characteristics of the groups have important differences. If so, it might be advisable to define each group as a separate universe. Always investigate this possibility before proceeding as if there were only a single universe.

Ineligible items

Occasionally, certain items not eligible for inclusion in the sample are not identified until they have actually been examined. Some of the random numbers may correspond to documents that have been voided, to inventory times that are no longer stocked, and the like. Or certain types of items or entries may not be of interest. For example, payment and receipt entries may be intermingled when only payments are to be sampled.

If you were unaware that certain items would be ineligible for inclusion in the sample, you would use up the random numbers before finding a sufficient quantity of eligible items. To obtain the specified sample size, select additional random numbers, plus some extras to allow for additional ineligible items.

However, the most efficient approach to this problem is to estimate in advance the proportion of usable items, if not already known, by scanning a list of the universe items (if available), taking a small preliminary sample of the items, or questioning agency officials. The estimated proportion of eligible items is then divided into the required sample size to determine the quantity of random numbers to be selected; that is, the required quantity of random numbers equals the specified sample size divided by the proportion of usable items.

Assume that in a review of travel vouchers, a sample of 300 vouchers involving reimbursement for the use of personally owned vehicles will be required. From a small preliminary sample, the evaluators estimate that the proportion of vouchers with mileage claims is about 75 percent, or 0.75. The required quantity of random numbers is calculated as 300 divided by 0.75, or 400.

After the items corresponding to the random numbers have been examined, the actual sample of eligible items may differ from the specified sample size. A rule of thumb is that differences of less than 10 percent of the specified sample size can be ignored. If the difference is 10 percent or greater, compute the quantity of additional random numbers to be selected as follows:

1. Determine the actual proportion of eligible items in the first sample by dividing the quantity of random numbers selected into the quantity of eligible items found.

2. Divide this proportion into the quantity of additional eligible items needed to determine the quantity of additional random numbers required.

If, in the travel vouchers example, the first sampling operation found only 220 vouchers containing mileage claims among the 400 vouchers examined, the proportion of eligible documents would be 220 divided by 400, or 0.55. Since 80 additional vouchers with mileage claims would be needed to obtain a sample of 300 eligible documents, the quantity of additional random numbers to be selected would be 80 divided by 0.55, or about 145.

If the sample of eligible items is larger than required, the sample size can be reduced by using one of the techniques described in the last section of this appendix.

<u>Selecting random letters or months</u>

Sometimes it is necessary to select a series of random letters or random months to draw a random sample. Some publications (such as Arkin, 1984) contain tables of random letters and random months. However, if such tables are unavailable, the problem of selecting a group of random letters can be easily resolved by selecting a group of random numbers from 1 to 26 from a table of random digits and assigning the letters of the alphabet that correspond to the numbers selected, such as A for 1, B for 2, and C for 3. Similarly, random months can be selected by selecting random numbers from 1 to 12 and assigning the months corresponding to the numbers selected.

<u>Compound numbering systems</u>

Sometimes numbering systems use a letter as a prefix or suffix to the digits in a number. This is referred to as a "compound numbering system," of which there are two types: the quantity of items is the same for each letter or the quantity of items differs for the various letters.

<u>Same quantity of items for each letter used</u>. Following the procedures below will produce an unduplicated random sample drawn from a single universe. Assume that a sample of 20 items is required from a universe numbered as follows:

A-0001 to A-5,000
B-0001 to B-5,000
C-0001 to C-5,000
through
Z-0001 to Z-5,000

There are 5,000 items for letters A through Z.

The <u>first step</u> is to select 20 random numbers from 0001 to 5,000, plus some extras, say 5. Record the random numbers, including the extras, in the order in which they were selected, keeping the extras separate. Duplicate numbers should not be eliminated at this point. The <u>second step</u> is to select 25 random letters from A through Z. Do not eliminate duplicates. Assume that a selection of 25 four-digit random numbers and 25 random letters yields the results shown in table II.2.

Table II.2

Selecting Random Numbers and Random Letters in Compound Numbering Systems with the Same Quantity of Items for Each Letter[a]

Order of selection	Random letter	Order of selection	Random number
1	M	1	3284
2	U	2	1224
3	J	3	0199
4	X	4	0578
5	Y	5	1240
6	K	6	0750
7	S	7	0994
8	O	8	2055
9	V	9	4038
10	A	10	4976
11	Y	11	4815
12	K	12	2751
13	K	13	1946
14	F	14	2814
15	N	15	2055
16	P	16	1944
17	Y	17	1240
18	R	18	4684
19	F	19	1353
20	O	20	2021
(extras)		(extras)	
21	N	21	1959
22	O	22	1644
23	N	23	4768
24	L	24	3612
25	V	25	1347

[a]Although F, K, N, O, and V and 1240 and 2055 appear more than once, they are not eliminated at this step.

The random numbers and letters shown in table II.2 were deliberately chosen to illustrate the process, and the quantity of duplicates has been exaggerated. Normally, it would not be necessary to select a quantity of extras equal to 25 percent of the original sample. The quantity of extras needed depends on the anticipated number of ineligible items (if any) that may be found in the sample and the anticipated number of nonresponses in a personal interview or questionnaire survey for which evaluators may want to substitute other randomly selected sampling units. Normally, a quantity of extras equal to 10 percent of the original sample should be enough.

The <u>third step</u> is to match numbers and letters, keeping both in the original order of selection. The results of the matching process are illustrated in table II.3 on the next page.

Table II.3

Matching Random Numbers and Random Letters
in Compound Numbering Systems
with the Same Quantity of Items for Each Letter

Order of selection	Letter and number	Order of selection	Letter and number
1	M-3284	14	F-2814
2	U-1224	15	N-2055
3	J-0199	16	P-1944
4	X-0578	17	Y-1240
5	Y-1240	18	R-4684
6	K-0750	19	F-1353
7	S-0994	20	O-2021
8	Q-2055	(extras)	
9	Y-4038	21	N-1959
10	A-4976	22	O-1644
11	Y-4815	23	N-4768
12	K-2751	24	L-3612
13	K-1946	25	V-1347

Table II.4

Sorting Random Number-Letter Combinations
in Compound Numbering Systems
with the Same Quantity of Items for Each Letter

Order of selection	Letter and number	Remarks
10	A-4976	
19	F-1353	
14	F-2814	
3	J-0199	
6	K-0750	
13	K-1946	
12	K-2751	
1	M-3284	
21	N-1959	Extra; replaces duplicate Y-1240
15	N-2055	
20	O-2021	
16	P-1944	
17	Q-2055	
18	R-4684	
7	S-0994	
2	U-1224	
4	X-0578	
5	Y-1240	
~~17~~	~~Y-1240~~	Duplicate; replaced by N-1959
9	Y-4038	
11	Y-4815	

In the fourth step, sort the original 20 number-letter combinations into alphabetical-numerical order and eliminate duplicates. Use the extra number-letter combinations, in the original order of selection, to replace duplicates. Insert replacements into the original group in the proper sequence of their number-letter combinations. The sorted list of random number-letter combinations is shown in table II.4. Note that extra N-1959 was used to replace the duplicate Y-1240. This was the first extra selected. If another duplicate had been discovered, extra O-1644 would have been used to replace it, and so on.

Different quantity of items for each letter used. In some instances, the number of items is not the same for two or more letters of the alphabet. This is a more complicated variation of the situation described above. It is important that the procedures described below be followed exactly.

Assume that a sample of 20 items is required from a universe numbered as follows:

A-0001 to A-5,056	E-0001 to E-4,619
B-0001 to B-5,397	F-0001 to F-7,691
C-0001 to C-7,409	G-0001 to G-6,100
D-0001 to D-4,455	H-0001 to H-5,406

Although the letters here go up only to H, they could go through the entire alphabet or they could start with F and end with Q, for example. The lowest number is 0001, and the highest is 7,691. In some numbering systems, the lowest number for one or more of the groups may be greater than 1; however, the sampling procedure remains the same.

The steps in selecting the sample follow:

1. Select 20 random numbers, plus some extras, say 10, between 0001 and 7,691. Record the random numbers, including the extras, in the order of selection, keeping the extras separate. Do not eliminate duplicates.

2. Select 20 random letters, plus 10 extras, between A and H. Record the random letters in the order of selection, keeping the extras separate. Do not eliminate duplicates.

Assume that the selection of random numbers and letters yields the results shown in table II.5 on the next page. Again, the random numbers and letters shown here were deliberately chosen to illustrate the process. The quantity of duplicates and out-of-bound numbers has been greatly exaggerated. Normally, it is not necessary to select a quantity of extras equal to 50 percent of the original sample; 10 to 15 percent should suffice.

The next step is to match the letters and numbers, keeping both in the original order of selection, producing the results shown in table II.6 on the next page.

Next, sort the original 20 letter-number combinations into alphabetical-numerical order. (Keep the extras in the original order of selection in a separate group.) Then eliminate two types of combinations: (1) those having no corresponding item number (that is, the out-of-bounds combinations) and (2) those that duplicate a combination previously selected. The eliminated original combinations are replaced by eligible extras in their order of selection.

For example, combination A-7,146 is greater than the highest number for the A group of items. Therefore, replace this

Table II.5

Selecting Random Numbers and Random Letters in Compound Numbering Systems with a Different Quantity of Items for Each Letter[a]

Order of selection	Random letter	Order of selection	Random number
1	G	1	6,385
2	F	2	0718
3	C	3	2,472
4	H	4	1,117
5	E	5	4,236
6	H	6	2,331
7	E	7	3,454
8	C	8	7,101
9	F	9	0742
10	C	10	2,472
11	F	11	0718
12	B	12	5,406
13	C	13	0563
14	G	14	1,438
15	D	15	3,952
16	A	16	7,146
17	A	17	3,158
18	A	18	3,615
19	G	19	2,547
20	B	20	1,992
(extras)		(extras)	
21	E	21	5,493
22	B	22	0317
23	G	23	4,122
24	F	24	6,320
25	D	25	6,206
26	B	26	4,071
27	C	27	7,545
28	F	28	3,253
29	B	29	6,817
30	C	30	0598

[a]Although certain numbers and letters appear more than once, they are not eliminated at this step.

Table II.6

Matching Random Numbers and Random Letters in Compound Numbering Systems with a Different Quantity of Items for Each Letter

Order of selection	Letter and number	Order of selection	Letter and number
1	G-6385	17	A-3158
2	F-0718	18	A-3615
3	C-2472	19	G-2547
4	H-1117	20	B-1992
5	E-4236	(extras)	
6	H-2331	21	E-5493
7	E-3454	22	B-0317
8	C-7101	23	G-4122
9	F-0742	24	F-6320
10	C-2472	25	D-6206
11	F-0718	26	B-4071
12	B-5406	27	C-7545
13	C-0563	28	F-3253
14	G-1438	29	B-6817
15	D-3952	30	C-0598
16	A-7146		

out-of-bounds combination by extra B-0317, the first eligible extra in order of selection. Although combination E-5,493 was the first extra selected, it is also out of bounds and cannot be used. Of the original 20 combinations, B-5,406 is also out of bounds and is replaced by extra F-6,320. C-2,472 was selected

Table II.7

Sorting Random Number-Letter Combinations in Compound Numbering Systems with a Different Quantity of Items for Each Letter

Order of selection	Letter and number	Remarks
17	A-3158	
18	A-3615	
~~16~~	~~A-7146~~	Out of bounds; replaced by B-0317
22	B-0317	Extra; replaces A-7146
20	B-1992	
26	B-4071	Extra; replaces duplicate F-0718
~~12~~	~~B-5406~~	Out of bounds; replaced by F-6320
13	C-0563	
3	C-2472	
~~10~~	~~C-2472~~	Duplicate; replaced by G-4122
8	C-7101	
15	D-3952	
7	E-3454	
5	E-4236	
2	F-0718	
~~11~~	~~F-0718~~	Duplicate; replaced by B-4071
9	F-0742	
28	F-3253	Extra; replaces G-6385
24	F-6320	Extra; replaces B-5406
14	G-1438	
19	G-2547	
~~1~~	~~G-6385~~	Out of bounds; replaced by F-3253
23	G-4122	Extra; replaces duplicate C-2472
4	H-1117	
6	H-2331	

twice, the third and tenth combinations drawn. The duplicate C-2,472 is eliminated and replaced by extra G-4,122, the third eligible extra. When used as replacements, the extras are put in their proper alphabetical-numerical sequence in the original sample. Continue the process until the required quantity of unduplicated, within-bounds number-letter combinations has been obtained. In this example, all eligible extras except C-0598 are used. (Extras E-5,493, D-6,207, C-7,545, and B-6,817 are out of bounds and cannot be used.) The final list of number-letter combinations appears in table II.7.

The use of index cards will greatly simplify the sample selection process when numbering systems of this type are encountered. First, number enough cards in serial order 1, 2, 3, and up to take care of the original sample plus the extras. Keeping the cards in order, select the random numbers and enter them on each card in turn; then select the random letters and enter them on each card in the same order. Thus, the letters and numbers are automatically matched in order of selection. The original cards can be easily sorted into alphabetical-numerical sequence, duplicate and out-of-bounds number-letter combinations can be eliminated, and replacements can be inserted into the original pack in their proper sequence. The preassigned serial numbers indicate the order of selection.

Population of items listed in a book or on a computer printout

When items are listed in a book or on a computer printout, a random number sample can be drawn by using both the page number and the line number. The technique is almost the same as that described above for compound numbering systems, except that the prefix is the page number instead of a letter, and the line number is the remaining portion of the number.

For example, assume that the evaluators want to draw a sample of inventory items from an 80-page catalog listing all items in stock. The maximum number of items that can be listed on a page is 156. Each item in the catalog can be identified by a two-part numbering system: the page number (from 1 to 80) and the line number (from 1 to 156). To draw the sample, merely select the required quantity of random numbers between 1 and 156, plus some extras, and the required quantity of random numbers between 1 and 80, plus extras, without eliminating duplicates. Match the two series of numbers in order of selection and sort them into numerical order, then eliminate duplicates, out-of-bounds numbers, and numbers corresponding to ineligible items (if any). The eliminated numbers are replaced by the extras in order of selection. The result is a simple random sample from the catalog.

If the items are printed in two or more parallel columns, count the items instead of the line numbers. This procedure can also be extended to more complicated numbering systems. For example, the numbering system may consist of three groups of digits arranged as 10-450-39. The first two digits represent the folio or book number, the next three represent the page number, and the last two represent the line number. Both the number of pages per book and the number of lines per page can vary. For this type of universe, first select a sufficient quantity (plus extras) of two-digit numbers for the lines, three-digit numbers for the pages, and two-digit numbers for the book numbers. Match the numbers in order of selection and replace duplicates and out-of-bounds combinations by the extras in order of selection.

Caution should be taken when sampling from books or computer printouts: it is usually incorrect to assume that the same number of eligible cases will be listed on each page.

Periodic serial numbering system

Some numbering systems assign numbers serially (1, 2, 3, etc.) to documents for a certain period such as a week or a month. At the beginning of the next time period, the numbering process starts again with the numeral 1. This type of system appears to assign the same identification numbers to different items. If, however, the time periods are assigned numbers 1 through 12 for months or 1 through 52 for weeks, this type of numbering system becomes almost identical to the page-number-line-number system described in the previous section. The number of the time period corresponds to the page number, and the serial number corresponds to the line number. Thus, the sample can be drawn by the same procedure used for sampling from a book or computer printout. The number of items should be expected to differ from one time period to another.

SYSTEMATIC SELECTION WITH A RANDOM START

We explained the theory of systematic selection with a random start in chapter 6. In this section, we present some detailed examples of the procedure, including some unusual situations. To use systematic sampling with a random start, assume that you want a sample of 200 items from an unnumbered list of 4,500 items. Divide the number of items, 4,500, by the specified sample size, 200, to get 22.5; then drop the digits to the right of the decimal point, regardless of whether their value is more or less than 0.5. This ensures that the sample is at least the required size. A sampling interval of 22 results in a sample size of 204 or 205, depending on the starting place. An interval of 23 provides a sample of only 194 or 195 items.

Then, from a table of random digits, select a two-digit random number between 01 and 22 as the starting point. Assume that this number is 13. Count the items on the list until item 13, which will be the first sample item, is reached. Count off 22 more items and take the 35th item for the 2nd sample item; then count 22 more items and take the 57th item; repeat through the 4,479th item. With this method, the entire sample of 204 items can be selected without bothering to number the items.

Consider a second example. Assume that a sample of 180 items is required from a list of items numbered serially from 41,001 through 44,000. First determine that there are no gaps in the numbering system--that is, that all the numbers between 41,001 and 44,000 have been used. Then subtract the first number used from the last number used and add 1 to the difference to obtain the universe size:

Last number	44,000
First number	41,001
Difference	2,999
	+ 1
Universe size	3,000

Divide the universe size, 3,000, by the specified sample size, 180, to obtain the sampling interval. In this case, the sampling interval will be 16, after dropping the digits to the right of the decimal point.

From a table of random digits, select a two-digit random number between 01 and 16. Assume that the random number is 05. This number equals the last two digits of the serial number of the first sample item, so the first sample item is 41,005.

Then add the sampling interval, 16, to the serial number of the first sample item, 41,005, to obtain the serial number of the second sample item, 41,021. Continue adding the sampling interval to the serial number of the item previously selected to obtain the serial number of the next sample item, until you obtain a number larger than the last serial number in the universe. Thus, the items with serial numbers 41,005, 41,021 (41,005 + 16), 41,037 (41,021 + 16), etc., through 43,997 (43,981 + 16) will be selected. (The serial numbers of the sample items can be easily determined by using an adding machine with a paper tape and taking subtotals after addition.)

Consider a third example. Assume that records in a single universe are maintained in four groups numbered as follows:

Group	Assigned serial numbers
A	14,542 through 17,921
B	19,055 through 19,988
C	22,001 through 23,021
D	25,500 through 26,401

These records are numbered in a broken series. To select a single systematic sample from the four groups of records, subtract the beginning serial number of each group from the ending serial number and add 1 to the difference to obtain the total number of records in each group. Then add the total numbers of records to obtain the universe size for all four groups combined. Assign mentally designated serial numbers to indicate the numerical sequence of each item in the universe. An example of the computation is shown in table II.8.

Table II.8

Selecting a Single Systematic Sample from Four Groups of Records

Group	Serial number Beginning	Serial number Ending	No. of records	Mentally designated serial number Beginning	Mentally designated serial number Ending
	(1)	(2)	(3)[a]	(4)[b]	(5)[c]
A	14,542	17,921	3,380	1	3,380
B	19,055	19,988	934	3,381	4,314
C	22,001	23,021	1,021	4,315	5,335
D	25,500	26,401	902	5,336	6,237
			6,237		

[a]Entry in column 3 = 1 plus entry in column 2 minus entry in column 1.
[b]Entry in column 4 = 1 plus entry on preceding line of column 5.
[c]Entry in column 5 = entry in column 3 plus entry on preceding line of column 5.

Next, divide the universe size by the required sample size, dropping decimals, to obtain the sampling interval. Suppose that from the universe of 6,237 records, a sample of 210 records is wanted. The sampling interval will be 6,237 divided by 210, or 29.7, or 29.

From a table of random digits, select a two-digit random number between 01 and 29. Assume the number is 03. This means that the third record will be the first sample item. The serial number of the first sample item is determined by subtracting 1 from the first number in group A and adding the random number 03 to the difference. The number of the first record selected for the sample will be 14,544 (14,542 - 1 + 3 = 14,544).

Determine the serial numbers of the second and successive sample items by adding the sampling interval, 29, to the serial number of each record previously selected. Thus, records with serial numbers 14,544, 14,573, 14,602, 14,631, etc., are selected from group A. Then determine the serial numbers of the last record to be selected from group A and the first record from group B by following the steps described below:

1. Subtract the serial number of the first record selected from the group from the highest serial number assigned to that group to obtain the balance of the group. For example:

Highest serial number assigned to group A	17,921
Less serial number of first record selected from group A	14,544
Balance of the group	3,377

2. Divide the balance of the group by the sampling interval and subtract the remainder from the balance of the group to obtain the group difference--that is, the difference between the serial numbers of the first and last records selected from the group. For example:

Balance of group A (3,377) divided by sampling interval (29) equals 116 with a remainder of 13	
Balance of the group	3,377
Less remainder	13
Group difference	3,364

3. Add the group difference to the serial number of the first record selected from the group to obtain the serial number of the last record to be selected from the group. For example:

Serial number of first sample record from group A	14,544
Plus group difference	3,364
Serial number of last sample record from group A	17,908

4. Subtract the remainder, obtained in step 2 above, from the sampling interval to determine the sequence of the first record to be selected from the next group. For example:

Sampling interval	29
Less remainder from group A	13
Sequence of first sample record to be selected from group B	16

5. Subtract 1 from the lowest serial number assigned to the next group and add the sequence obtained in step 4 to determine the serial number of the first sample record to be selected from the next group. For example:

Lowest serial number in group B	19,055
Less 1	1
Difference	19,054
Plus sequence of first sample record	16
Serial number of first sample record to be selected from group B	19,070

To obtain the serial numbers of the sample records in group B and the serial number of the first sample record to be selected from group C, repeat the procedures described above. For example:

Highest number assigned to group B	19,988
Less serial number of first record selected from group B	19,070
Balance of the group	918

Balance of the group (918) divided by sampling interval (29) equals 31 with a remainder of 19

Balance of the group	918
Less remainder	19
Group difference	899

Serial number of first sample record from group B	19,070
Plus group difference	899
Serial number of last sample record from group B	19,969

Sampling interval	29
Less remainder	19
Sequence of first sample record to be selected from group C	10

Lowest serial number in group C	22,001
Less 1	1
Difference	22,000
Plus sequence of first sample record	10
Serial number of first sample record to be selected from group C	22,010

Use this procedure when going from one group to another, continuing it through all groups until the entire sample has been selected. The actual sample size will be 215 because the calculated interval of 29.7 was rounded downward to 29.

Sampling by measurement

If the sampling units are equal-width documents such as punch cards, index cards, or sheets of paper, a quick systematic sample can be obtained by measurement. This method should be used with caution if there is any likelihood that the findings will be controversial, because it is impossible to document the sampling unit that was actually closest to the ruler's mark.

Simply measure the total length of the file of records and divide this by the desired sample size to obtain a sampling interval in inches or fractions of inches. For example, assume that you want to select a sample of 160 cards from two file drawers, each measuring 2-1/2 feet in length. The steps are as follows:

1. Convert the total length of the two file drawers, or 2-1/2 ft. + 2 1/2 ft = 5 ft, into inches:
   5 x 12 in. = 60 in.

2. To obtain the required sampling interval, divide the required sample size into the total length of the files:
   60/160 in. = 3/8 in.

3. Select the sample by laying a ruler on top of the file and selecting the cards that are opposite each 3/8-inch mark on the ruler. To obtain a random start, place the end of the ruler at a randomly selected card between the beginning of the file and a point that is 3/8 inch along the file.

Sometimes the sampling interval obtained by dividing the sample size into the total length of the universe will not coincide with one of the fractional parts of an inch marked on a ruler. If this happens, round the sampling interval downward to coincide with the closest marking. For example, if the quotient obtained by dividing the required sample size into the length of the universe were 0.65 inch, the sampling interval should be 5/8 inch (0.625 inch). If the quotient were 1.4 inches, the sampling interval should be 1-3/8 inches (1.375 inches). This sample selection method is much easier to use if a measuring instrument in centimeters and millimeters is available.

When the sample is to be drawn from a large number of items, selection by measuring, for all practical purposes, is equivalent to systematic selection by counting. However, selection by measurement cannot be used if the sampling units are of varying widths, because the thicker items will have a greater probability of being selected.

A similar problem arises in systematic selection by counting if the same sampling unit has several cards, folders, and the like or appears more than once on a list. When this occurs, all the cards, lists, and so on for the sampling unit must be considered a single sampling unit and counted as such; otherwise, these sampling units will have a greater probability of being selected than those listed once or having only one folder.

If it is not possible or efficient to combine several lists before sampling, alternative procedures may be used. One is to sample from each of the lists or sources available. For all selected units, check each list to see on how many of the lists they are found. Then subsample inversely to the number of times found; in other words, take a sample of half of the items on two lists, one third of the items on three lists, and so on. If it is impossible to determine how many lists a unit is on until the data are gathered, the data will have to be weighted to develop unbiased estimates.[1]

Expanding a systematic sample

Sometimes it may be necessary to increase the size of a systematic sample. The simplest method is to select a quantity of additional random starting points between 1 and the original sampling interval that will result in a total sample size approximately equal to the specified sample size.

Another method is, in effect, to redefine the universe by excluding the items selected in the original systematic sample. Using the procedures described above, select from the redefined universe a supplementary systematic sample that is approximately equal in size to the number of additional sampling units required. To make sure that none of the sampling units selected for the original sample are counted when locating the random starting point and the additional sampling units, identify them by check marks, paper clips, or the like. After the supplementary sample has been selected, combine it with the original sample and consider it to be a single sample in which each unit's probability of being selected equals the final sample size divided by the universe (n/N).

Assume that the evaluators have selected a preliminary sample of 30 items from a universe of 9,000 items by starting with the 2nd sampling unit (selected randomly) and taking every 300th item thereafter. Identify the sample items by check marks in the list from which they were selected. If after analyzing the results of the preliminary sample, you decide that the final sample should

[1]We are grateful to Seymour Sudman of the University of Illinois, Champaign, Ill., for pointing this out.

have 400 items, follow these steps to obtain the 370 additional items:

1. Redefine the universe to exclude items selected for the preliminary sample and subtract the preliminary sample size, 30, from the universe size, 9,000, to obtain the size of the redefined universe, 8,970 items.

2. Divide the number of items required for the supplementary sample, 370, into the redefined universe size, 8,970, to obtain the sampling interval for the supplementary sample, 24.

3. Select a random number between 1 and 24 to determine the starting point for the systematic sample from the redefined universe. Whatever the starting point may be, do not count, but merely skip over, original universe items (2, 302, 602, etc.) in selecting the supplementary sample.

If it is not practical to identify on the list the items selected for the preliminary sample, a third method can be used to expand a systematic sample. Subtract the number of items in the original sample from the universe size and divide the difference by the number of items required for the supplementary sample to obtain the sampling interval. Use this interval to select a systematic sample from the original universe, proceeding as if the original sample had not been selected. The second sample is combined with the original sample, and items from the second sample that duplicate items in the preliminary sample are eliminated.

Assume the same situation that existed in the previous example, except that when actually going through the universe, you are unable to identify readily the items selected in the preliminary sample. Follow the procedures below:

1. Calculate the sampling interval for the supplementary sample as 1 in 24, exactly as in the previous example.

2. Select a systematic sample, using the sampling interval of 1 in 24, from the entire original universe of 9,000 items, as if the preliminary sample had not been selected. The second sample will contain about 375 items.

3. Compare the items in the supplementary and preliminary samples and eliminate from the supplementary sample the items that are duplicates of items selected in the preliminary sample. (About 1.3 percent of the items in the supplementary sample will be eliminated.)

4. Combine the two samples to obtain an unduplicated sample of about 400 items, which is the sample size required.

REDUCING THE SIZE OF A SAMPLE

Sometimes it is necessary to reduce the size of, or "thin out," a sample already selected. Some possible reasons for this follow.

1. If the sample was drawn by systematic selection with a random start, the universe size may be larger than originally estimated and, as a result, the sample may be larger than required.

2. You may discover that it takes longer than anticipated to examine a sample item and examining all the items selected would take more time than justified by the review's objective.

3. You may have only a single opportunity to draw the sample. To ensure that the final sample will have sufficient items, you may select many more items than it was estimated the final sample would require. Draw a smaller sample, or subsample, from this large sample for the preliminary sample. You may also be able to use an additional subsample of the large sample as a supplementary sample to achieve the final required sample size.

Basically, thinning a sample is drawing a randomly selected subsample of items from a sample that has already been selected. All items in the original sample must have equal opportunity of being selected in the subsample. If they do not, the subsample will not be representative of the universe.

One of the simplest methods of thinning a sample is to use systematic selection with a random start to select either the items to be eliminated or those to be retained. This method will work regardless of the procedure used to select the sample and, in general, regardless of the sequence of the sample items, unless the evaluators suspect that the sequence may cause the characteristic being measured to recur at regular intervals. This can sometimes be detected by inspecting a list of the sample items and the corresponding values. If the situation is suspected, use random number sampling to thin the sample.

If the sample items have consecutively assigned identification numbers, the sample can be thinned by using sampling that is based on randomly selected combinations of terminal digits (discussed in chapter 6).

If random number sampling was used to draw the original sample and the random numbers are still in sequence in the order of selection or can be rearranged into that order, random numbers can be eliminated by beginning with the last one selected and working back until enough random numbers have been eliminated to reduce the sample to the required size. Or start with the first random number selected and, working forward, count out a quantity of random numbers equal to the required sample size. If the random numbers are arranged in any sequence other than the order of selection, this procedure cannot be used, unless they are randomized by some random procedure that can be documented. The numbers must be in some random order, although not necessarily in the order of selection.

Random number sampling can also be used to thin a sample. This procedure will work regardless of how the original sample was selected or how the sample items are sequenced. Simply use one of the procedures in this appendix to select either the items to be retained in the sample or those to be eliminated. Selection of the random numbers will be easier if the sample items have been renumbered consecutively from 1 through the last item in the sample.

DETAILS ON STRATIFIED AND CLUSTER SAMPLING

This appendix discusses the computation of estimates, sampling errors, and sample sizes and the allocation of a sample among strata when stratified sampling is used. It briefly describes how to construct stratum boundaries and determine the optimum number of strata. It concludes with a brief description of a two-stage, cluster-sampling problem.

TYPES OF STRATIFIED SAMPLING

As we noted in chapter 2, in stratified sampling the sample size in each stratum may be proportional to the total number of sampling units in the stratum ("proportional allocation"), or it may be disproportional. Examples of both kinds of allocation are given in this appendix.

In proportional allocation, the proportional relationship between the stratum sample size and the total sample size is the same as that between the stratum universe size and the total universe size. That is, the sampling fraction (sample size divided by universe size) is the same in all strata.

The advantages of proportional allocation over other allocation methods are (1) the formula for allocating the sample to the strata is simple, (2) the formulas for computing estimates are simple, and (3) proportional allocation is intuitively more familiar to those who use the final results, which may prevent them from making gross errors if they attempt to manipulate the sample results arithmetically.

In disproportional allocation, there are three methods of allocating the sample to the strata. The judgmental method is simply based on the evaluator's "desire" to meet a specific objective, such as doing a 100-percent audit of all high-value transactions and auditing a sample of the remaining transactions or doing a 100-percent audit of the more error-prone transactions, if they can be identified, and auditing a sample of the others.

The other methods of disproportional allocation are known as "Neyman allocation" and "optimum allocation." The Neyman method allocates the sample to each stratum in proportion to the product of the stratum universe and the stratum standard deviation, divided by the sum over all strata of the products of the stratum universes and standard deviations. The standard deviation can be estimated from a preliminary sample or from a prior audit or study. The advantage of the Neyman method is that the sampling error is minimized for a given sample size.

Optimum allocation allocates the sample to strata by taking into account the differences in (1) universe sizes, (2) standard deviations, and (3) the costs of collecting data among the various strata. Optimum allocation minimizes the sampling error for a specific total cost of data collection or, conversely, minimizes the total cost of data collection for a specified precision (tolerable error). A discussion of optimum allocation is beyond the scope of this paper.

STRATIFIED SAMPLING FOR VARIABLES

We will first discuss stratified sampling for variables. To illustrate proportional, judgmental, and Neyman sample allocation methods and the procedures for computing estimates, sampling errors, and sample sizes, consider the following example.

While reviewing shipping costs at a supply depot, the evaluators suspect that air freight forwarding, the shipping method used, is less economical than direct air carrier. By calculating the costs of several recent shipments from direct air carrier rate schedules, they find that in each case the direct air shipping cost is less than the amount paid to the air freight forwarder. The evaluators decide to estimate, using statistical sampling, the total savings that would have resulted had direct air carriers been used instead of air freight forwarders. The confidence level is set at 95 percent.

The universe is defined as all air freight forwarder shipments during the past 3 months. From the depot's file of shipping documents, the information required to compute the cost of shipments by direct air carrier rate schedules is copied onto 3 x 5 cards. This procedure results in 250 cards, each representing a single shipment. Using their judgment, the evaluators classify the cards into three groups, based on air freight forwarder shipping costs:

Air freight forwarder shipping costs	Number of shipments
Less than $100	150
$100 to $499	75
$500 or more	25
Total	250

The resulting savings on each shipment are shown in tables III.1 and III.2 on the next page. (Note that in real life, the savings would not be known until the shipping costs by direct air carrier had been computed. Calculating shipping costs is a very complicated, time-consuming procedure; this is why sampling was necessary.)

We will assume that the evaluators decide to take a preliminary sample of 50 items to estimate the total savings and the sampling error and to determine the final sample size. In the illustrations of the three allocation methods, we will use the true standard deviations calculated from the universe data.

Table III.1

Savings from Using Direct Air Shipment, Instead of Air Freight Forwarder, Classified by Air Freight Forwarder Shipping Costs of Less than $100[a]

Shipment	Saving	Shipment	Saving	Shipment	Saving	Shipment	Saving
1	$21	39	$38	77	$22	115	$15
2	27	40	26	78	40	116	39
3	33	41	13	79	38	117	37
4	44	42	31	80	36	118	28
5	11	43	47	81	21	119	30
6	52	44	51	82	22	120	46
7	23	45	29	83	34	121	36
8	32	46	37	84	42	122	33
9	43	47	29	85	39	123	22
10	39	48	25	86	45	124	20
11	23	49	17	87	34	125	33
12	26	50	18	88	15	126	23
13	19	51	29	89	17	127	45
14	24	52	18	90	21	128	26
15	39	53	25	91	38	129	50
16	22	54	28	92	28	130	0
17	35	55	22	93	24	131	28
18	35	56	35	94	42	132	25
19	39	57	33	95	34	133	17
20	34	58	24	96	28	134	10
21	13	59	15	97	26	135	33
22	19	60	31	98	30	136	44
23	4	61	6	99	23	137	6
24	30	62	19	100	34	138	24
25	31	63	16	101	37	139	22
26	16	64	29	102	37	140	36
27	22	65	41	103	38	141	24
28	13	66	24	104	30	142	55
29	46	67	26	105	42	143	25
30	37	68	18	106	34	144	26
31	47	69	30	107	41	145	33
32	37	70	27	108	37	146	43
33	15	71	29	109	45	147	23
34	27	72	42	110	44	148	15
35	10	73	40	111	42	149	30
36	20	74	18	112	23	150	35
37	35	75	31	113	30		
38	33	76	17	114	36		

[a]In real life, savings would not be known until after the calculation of a shipment's cost, a complicated and time-consuming procedure.

Table III.2

Savings from Using Direct Air Shipment, Instead of Air Freight Forwarder, Classified by Air Freight Forwarder Shipping Costs of More than $100[a]

$100 to $499				$500 or more	
Shipment	Saving	Shipment	Saving	Shipment	Saving
1	$ 32	39	$142	1	$431
2	62	40	170	2	500
3	190	41	108	3	502
4	140	42	113	4	320
5	96	43	139	5	259
6	99	44	121	6	457
7	78	45	143	7	304
8	130	46	147	8	276
9	66	47	232	9	404
10	75	48	192	10	270
11	48	49	182	11	255
12	160	50	182	12	373
13	110	51	71	13	252
14	145	52	98	14	348
15	159	53	63	15	336
16	200	54	132	16	264
17	109	55	65	17	321
18	100	56	57	18	360
19	153	57	128	19	375
20	127	58	140	20	251
21	45	59	141	21	285
22	90	60	113	22	210
23	157	61	149	23	445
24	92	62	201	24	288
25	155	63	112	25	462
26	167	64	188		
27	125	65	164		
28	59	66	94		
29	78	67	127		
30	78	68	156		
31	154	69	64		
32	158	70	198		
33	199	71	121		
34	96	72	208		
35	83	73	0		
36	105	74	134		
37	61	75	121		
38	138				

[a]In real life, savings would not be known until after the calculation of a shipment's cost, a complicated and time-consuming procedure.

In a real situation, the standard deviation would have to be computed from the sample data for each stratum, using the procedures described in chapter 3.

Using proportional allocation

First, assume that the evaluators use proportional allocation. The formulas for allocating the sample to the strata are shown below (and a work sheet for making the computations appears in table III.3). In the formulas, N equals the total universe size, 250; n equals the total sample size, 50; N_h equals the stratum universe size; W_h equals the stratum weight; and n_h equals the stratum sample size. Sample shipments and computed savings appear in table III.4.

Table III.3

The Proportional Allocation of Samples to Strata

Stratum: shipping cost	Universe size	Weight	Sample size
(h)	(N_h)	(W_h)	(n_h)
Less than $100	150	0.6	30
$100 to 499	75	0.3	15
$500 or more	25	0.1	5
Total	250	1.0	50

Table III.4

Proportional Allocation: Sample Shipments and the Computation of Totals

Less than $100		$100 to $499		$500 or more	
Shipment	Saving	Shipment	Saving	Shipment	Saving
6	$ 52	3	$ 190	1	$ 431
8	32	7	78	6	457
11	23	22	90	13	252
12	26	29	78	22	210
13	19	33	199	23	445
20	34	37	61		
25	31	38	138		
36	20	46	147		
39	38	49	182		
43	47	56	57		
44	51	58	140		
51	29	61	149		
53	25	64	188		
72	42	69	64		
74	18	75	121		
77	22				
85	39				
88	15				
93	24				
94	42				
100	34				
104	30				
109	45				
115	15				
127	45				
129	50				
132	25				
134	10				
143	25				
145	33				
	$941		$1,882		$1,795

Total = $4,618

$$W_h = \frac{N_h}{N} \text{ and } n_h = nW_h$$

With proportional allocation, the overall stratified mean, $\bar{y}_{st}$, can be computed by dividing the total of all the values in the sample by the total sample size:

$$\bar{y}_{st} = \frac{4{,}618}{50}$$
$$\bar{y}_{st} = 92.36$$

This method of computing the stratified mean works only with proportional allocation. If anything happens during the sample selection or data collection processes to distort the proportionality, such as failure to obtain the data for some of the sample items, the formulas for disproportional allocation must be used to compute the stratified mean.

The estimated total savings ($\hat{y}_{st}$) is \$23,090, computed as follows:

$$\hat{Y}_{st} = N\bar{y}_{st}$$
$$\hat{Y}_{st} = (250)(92.35)$$
$$\hat{Y}_{st} = 23{,}090$$

In our example, the formula for computing the sampling error of the estimated total ($E_{\hat{Y}_{st}}$) is

$$E_{\hat{Y}_{st}} = t\sqrt{\sum_{h=1}^{3} \frac{N_h^2(N_h - n_h)}{N_h} \cdot \frac{S_h^2}{n_h}}$$

A work sheet for computing the sampling error of the total appears in table III.5 on the next page.[1] This can be simplified to

$$E_{\hat{Y}_{st}} = t\sqrt{\sum_{h=1}^{3} \frac{N_h(N_h - n_h)S_h^2}{n_h}}$$

where $E_{\hat{Y}_{st}}$ equals the sampling error of the estimated total, and S_h equals the standard deviation for stratum h. The computations follow.

---

[1]There are specific formulas for computing sampling errors for proportional allocation and for Neyman allocation. In this paper, we show the general formula, which works with any type of allocation, because true Neyman allocation and, often, true proportional allocation are difficult to achieve.

Table III.5

Work Sheet for Computing Sampling Error of Total with Proportional Allocation

Stratum: shipping cost (h)	Universe size (N_h)	Sample size (n_h)	Standard deviation (S_h)	N_h-n_h	$N_h(N_h-n_h)$	S_h^2	$\frac{S_h^2}{n_h}$	$\frac{N_h(N_h-n_h)S_h^2}{n_h}$
Less than $100	150	30	10.54	120	18,000	111.1	3.703	66,654
$100 to $499	75	15	47.17	60	4,500	2,225	148.3	667,350
$500 or more	25	5	83.77	20	500	7,017	1,403	701,500
	250							1,435,504

$$E_{\hat{Y}_{st}} = 1.96\sqrt{1,435,504}$$

$$E_{\hat{Y}_{st}} = 2,348$$

If the sampling error of the stratified mean is needed, use

$$E_{\bar{y}_{st}} = \frac{E_{\hat{Y}_{st}}}{N}$$

Using judgmental allocation

To illustrate judgmental allocation, we can assume that the evaluators decide to compute savings for all shipments of $500 or more, for 15 sample shipments from the $100 to $499 stratum, and for 10 sample shipments from the less-than-$100 stratum. The results obtained are shown in table III.6. The procedure for

Table III.6

An Example of the Judgmental Allocation of Shipping Costs

	Less than $100		$100 to $500		$500 or more	
	Shipment	Saving	Shipment	Saving	Shipment	Saving
	35	$ 10	2	$ 62	1	$ 431
	57	33	5	96	2	500
	62	19	10	75	3	502
	80	36	20	127	4	320
	81	21	23	157	5	259
	113	30	28	59	6	457
	117	37	30	78	7	304
	126	23	31	154	8	276
	129	50	37	61	9	404
	135	33	38	138	10	270
			43	139	11	255
			45	143	12	373
			61	149	13	252
			66	94	14	348
			67	127	15	336
					16	264
					17	321
					18	360
					19	375
					20	251
					21	285
					22	210
					23	445
					24	288
					25	462
Total		$292		$1,659		$8,548
Mean		$29.20		$110.60		$341.92

Table III.7

Work Sheet for Computing Estimated Total Savings with Judgmental Allocation

Stratum: shipping cost	Universe size	Estimated mean	Estimated total
(h)	(N_h)	($\bar{y}_h$)	($\hat{Y}_h$)
Less than $100	150	$ 29.20	$ 4,380
$100 to $499	75	110.60	8,295
$500 or more	25	341.92	8,548[a]
	250		$21,223

[a]This amount could also be obtained from the sample total, since 100-percent sampling was used in this stratum.

Table III.8

Judgmental Allocation Work Sheet for Computing Sampling Error of Total

Stratum: shipping cost	Universe size	Sample size	Standard deviation					
(h)	(N_h)	(n_h)	(S_h)	N_h-n_h	$N_h(N_h-n_h)$	S_h^2	$\frac{S_h^2}{n_h}$	$\frac{N_h(N_h-n_h)S_h^2}{n_h}$
Less than $100	150	10	10.54	140	21,000	111.1	11.11	233,310
$100 to $499	75	15	47.17	60	4,500	2,225	148.33	667,485
$500 or more	25	25	83.77	0	a	a	a	0
	250							900,795

[a]Computations unnecessary because this stratum does not contribute to sampling error.

computing the estimated total savings is shown in table III.7. The estimated total savings is $21,223, and the stratified mean ($\bar{y}_{st}$) equals $21,223 divided by 250, or $84.89.

$$\hat{Y}_h = N_h \bar{y}_h \quad \text{and} \quad \hat{Y}_{st} = \sum_{h=1}^{3} \hat{Y}_h$$

The work sheet for computing the sampling error of the total is shown in table III.8. Using the formula for computing the sampling error of the estimated total, we find

$$E_{\hat{Y}_{st}} = 1.96\sqrt{900{,}795}$$

$$E_{\hat{Y}_{st}} = 1{,}860$$

The sampling error of the stratified mean equals $1,860 divided by 250, or $7.44.

- Using Neyman allocation

The third allocation method we illustrate is Neyman allocation. The work sheet in table III.9 is on the next page, and the formulas are given below.

Table III.9

Work Sheet for Allocating Samples to Strata with Neyman Allocation in Sampling for Variables

Stratum: shipping cost	Universe size	Standard deviation		Weight	Sample size
(h)	(N_h)	(S_h)	N_hS_h	(w_h)	(n_h)
Less than $100	150	10.54	1,581	0.2192	11
$100 to $499	75	47.17	3,538	0.4905	24
$500 or more	25	83.77	2,094	0.2903	15
	250		7,213	1.0000	50

$$w_h = \frac{N_h S_h}{\sum_{h=1}^{3} N_h S_h} \quad \text{and} \quad n_h = n w_h$$

It should be noted that this example is not very realistic and is given solely for purposes of illustration. In a real-life situation, the evaluators would not know what the stratum standard deviations of the savings were unless they had taken a small preliminary sample. In many real-life situations, this would not be possible. Later in this appendix, we discuss practical methods for overcoming this problem.

Independent random samples of the required size are selected from each stratum; the results are shown in tables III.10 and III.11. From table III.11, we see that the estimated total saving is $21,223; thus, the stratified mean ($\bar{y}_{st}$) is $21,223 divided by 250, or $84.89. The work sheet for computing the sampling error of the total is shown in table III.12. Using the formulas,

$$\hat{Y}_h = N_h \bar{y}_h \quad \text{and} \quad \hat{Y}_{st} = \sum_{h=1}^{3} \hat{Y}_h$$

we find

$$E_{\hat{Y}_{st}} = (1.96)\sqrt{682{,}151}$$

$$E_{\hat{Y}_{st}} = 1{,}619$$

The sampling error of the stratified mean is obtained by dividing $1,619 by 250, or $6.48.

With Neyman allocation, the sample size for a certain stratum may be equal to or greater than the stratum universe size. For example, if a sample of 100 were taken, the sample in the highest stratum would be 29 shipments. Since this stratum has only 25 shipments, the remaining 4 shipments would have to be allocated to the two other strata.

Table III.10

Neyman Allocation: Sample Shipments and the Computation of Means

Less than $100 Shipment	Less than $100 Saving	$100 to $499 Shipment	$100 to $499 Saving	$500 or more Shipment	$500 or more Saving
1	$ 21	2	$ 62	2	$ 500
52	18	3	190	6	457
61	6	4	140	7	304
90	21	6	99	8	276
97	26	7	78	9	404
104	30	9	66	11	255
110	44	10	75	12	373
111	42	12	160	13	252
113	30	13	110	15	336
124	20	16	200	16	264
127	45	21	45	17	321
		27	125	18	360
		28	59	19	375
		33	199	21	285
		43	139	23	445
		50	182		
		51	71		
		53	63		
		55	65		
		56	57		
		64	188		
		69	64		
		71	121		
		74	134		
Total	$303		$2,692		$5,207
Mean	$27.55		$112.17		$347.13

Table III.11

Neyman Allocation: Computation of Estimated Total Savings

Stratum: shipping cost	Universe size	Estimated mean	Estimated total
(h)	(N_h)	($\bar{y}_h$)	($\hat{Y}_h$)
Less than $100	150	$ 27.55	$ 4,132
$100 to $499	75	112.17	8,413
$500 or more	25	347.13	8,678
	250		$21,223

Table III.12

Neyman Allocation Work Sheet for Computing Sampling Frror of Total

Stratum: shipping cost	Universe size	Sample size	Standard deviation					
(h)	(N_h)	(n_h)	(S_h)	N_h-n_h	$N_h(N_h-n_h)$	S_h^2	$\frac{S_h^2}{n_h}$	$\frac{N_h(N_h-n_h)S_h^2}{n_h}$
Less than $100	150	11	10.54	139	20,850	111.1	10.10	210,585
$100 to $499	75	24	47.17	51	3,825	2,225	92.71	354,616
$500 or more	25	15	83.77	10	250	7,017	467.80	116,950
	250							682,151

A comparison of the sampling errors

To summarize, our sampling errors are $2,348, $1,860, and $1,619 for proportional, judgmental, and Neyman allocation, respectively. As can be seen, Neyman allocation gave the smallest sampling error. Judgmental allocation provided the next best estimate of the sampling error, because the stratum with the largest variation was sampled 100 percent. However, a comparison with Neyman allocation reveals that the sample in the top stratum (shipments of $500 or more) was too large and that the sample from the middle stratum was too small. Proportional allocation gave the largest sampling error because half the sample was drawn from the stratum with the least variation. Judgmental allocation will not always give a better result than proportional allocation. If 60 percent of the sample had been allocated to the bottom stratum (shipments less than $100), the sampling error would have been greater than that obtained with proportional allocation.

It is interesting to see what would have happened if a simple random sample of 50 shipments had been taken from the entire universe without regard to stratification. The standard deviation for the unstratified universe is $101.59. The sampling error of the total would be $6,297, which is roughly 2.7 times the result obtained with proportional allocation.

Computing sample sizes

With stratified sampling, just as with simple random sampling, the evaluators can compute the overall sample size required. We will illustrate the computation of sample size with proportional and Neyman allocation. The formulas are slightly more complicated than with simple random sampling, but the work can be greatly simplified if it is done on a work sheet.

Before computing the sample size, the evaluators must specify the confidence level and the required precision. They will also need an estimate of the standard deviation obtained either from a preliminary sample or from a prior review. (When sampling for attributes, some reasonable "guesstimate" of the value of the proportion of interest allows the calculation of a standard deviation.)

Assume that the evaluators, after seeing the results of the preliminary sample of air freight forwarder shipments, decide that the sampling error of the total should be $1,500 at the 95-percent confidence level. In proportional allocation, the required sample size is computed by using the following formulas and the work sheet in table III.13. The formula to compute the desired sampling error of the stratified mean (E) is as follows. Note that the second term in the denominator is the finite population correction.

$$E = \frac{\text{desired sampling error of estimated total}}{N}$$

$$E = \frac{1500}{250}$$

$$E = 6$$

$$n = \frac{\sum_{h=1}^{3} W_h S_h^2}{\left(\frac{E}{t}\right)^2 + \frac{1}{N}\sum_{h=1}^{3} W_h S_h^2}$$

Table III.13

Work Sheet for Computing Sample Size

Stratum: shipping cost	Universe size	Standard deviation	Weight		
(h)	(N_h)	(S_h)	(W_h)	W_hS_h	$W_hS_h^2$
Less than $100	150	10.54	0.6	6.324	66.65
$100 to $499	75	47.17	0.3	14.151	667.50
$500 or more	25	83.77	0.1	8.377	701.74
	250		1.0	28.852	1,435.89

Using 2 for the t factor for 95-percent confidence, we compute:

$$n = \frac{1{,}436}{\left(\frac{6}{2}\right)^2 + \frac{1{,}436}{250}}$$

$$n = 97.40 \quad \text{or} \quad 98 \quad \text{(rounding up)}$$

Since our preliminary sample contained 50 shipments, 48 additional shipments would have to be sampled. The additional items would be allocated to the strata in proportion to the stratum universe sizes.

The formula for computing the required sample size using Neyman allocation is

$$n = \frac{\left(\sum_{h=1}^{3} W_h S_h\right)^2}{\left(\frac{E}{t}\right)^2 + \frac{1}{N}\sum_{h=1}^{3} W_h S_h^2}$$

$$n = \frac{(28.85)^2}{\left(\frac{6}{2}\right)^2 + \frac{1{,}436}{250}}$$

$$n = 56.45 \quad \text{or} \quad 57$$

E will be the same, \$6, as for proportional allocation, and the value for

$$\sum W_h S_h$$

can be obtained from the "total" line of the work sheet used for proportional allocation.

Thus, when we use Neyman allocation, only 7 shipments in addition to those in the preliminary sample would have to be reviewed. (The additional shipments would have to be allocated to the strata by using the procedures described in the section above on Neyman allocation.)

STRATIFIED SAMPLING FOR ATTRIBUTES

Stratified sampling can also be used when sampling for attributes. Compared with simple random sampling, stratified sampling may slightly reduce the sampling error. It also allows the development of separate estimates for individual strata, if this is necessary, provided that the sample sizes in the strata are sufficiently large. However, the disadvantage of stratifying when sampling for attributes is that the increased precision is usually not worth the additional work required.

To illustrate stratified sampling for attributes, we can assume that the evaluators are reviewing civilian payroll records at three military bases. Since the records are separately maintained at each location, the universe is stratified by location. The evaluators decide to select independent random samples of 100 payroll records at each base (judgmental allocation), with the results shown in table III.14. The work sheets for computing the overall stratified percentage of payroll records with errors and the estimated number of records with errors are shown in table III.15. The formulas are

$$p_h = \frac{a_h}{n_h}$$

$$\hat{A}_h = N_h p_h$$

To compute the overall stratified estimate of the number of errors in all three strata ($\hat{A}_{st}$), the formula is

$$\hat{A}_{st} = \sum_{h=1}^{h=3} \hat{A}_h$$

The stratified estimate of the number of payroll records with errors is 650 (the total for the last column of the work sheet).

Table III.14

The Results of a Review of Civilian Payroll Records at Three Military Bases

Location	Payroll records: In universe	Sampled	With errors
(h)	(N_h)	(n_h)	(a_h)
1	1,100	100	45
2	1,500	100	5
3	400	100	20
Total	3,000	300	70

Table III.15

The Computation of the Estimated Number and Percentage of Payroll Records with Errors

Location	Universe size	Sample size	Number of errors in sample	Proportion with errors	Estimated number of errors
(h)	(N_h)	(n_h)	(a_h)	(p_h)	($\hat{A}_h$)
1	1,100	100	45	0.45	495
2	1,500	100	5	0.05	75
3	400	100	20	0.20	80
	3,000	300	70		650

The overall estimated percentage of payroll records with errors (p_{st}) is computed as

$$p_{st} = \frac{\hat{A}_{st}}{N}$$

$$p_{st} = \frac{650}{3{,}000}$$

$$p_{st} = 0.2167 \quad \text{or} \quad 21.67\%$$

The formulas to compute sampling errors are shown below, and the work sheet appears in table III.16.

Table III.16

Work Sheet for Computing Payroll Sampling Error

Location	Universe size	Sample size	Proportion with errors			(q_h)			
(h)	(N_h)	(n_h)	(p_h)	N_h-n_h	$N_h(N_h-n_h)$	$1-p_h$	p_hq_h	$\frac{p_hq_h}{n_h}$	$\frac{N_h(N_h-n_h)p_hq_h}{n_h}$
1	1,100	100	0.45	1,000	1,100,000	0.55	0.2475	0.002475	2,722.5
2	1,500	100	0.05	1,400	2,100,000	0.95	0.0475	0.000475	997.5
3	400	100	0.20	300	120,000	0.80	0.1600	0.001600	192.0
	3,000								3,912.0

The formula to compute the sampling error of the estimated number of payroll records with errors ($E_{\hat{A}_{st}}$) is

$$E_{\hat{A}_{st}} = t\sqrt{\sum_{h=1}^{h=3} \frac{N_h(N_h - n_h)p_h q_h}{n_h}}$$

$$E_{\hat{A}_{st}} = (1.96)\sqrt{3{,}912.0}$$

$$E_{\hat{A}_{st}} = 122.59 \quad \text{or} \quad 123$$

Thus, the sampling error of the estimated number of payroll records with errors is 123 at the 95-percent confidence level.

The sampling error of the overall estimated percentage of payroll records with errors ($E_{p_{st}}$) is

$$E_{p_{st}} = \frac{E_{\hat{A}_{st}}}{N}$$

$$E_{p_{st}} = \frac{122.59}{3{,}000}$$

$$E_{p_{st}} = 0.040863 \quad \text{or} \quad 4.09\%$$

Thus, the sampling error of the stratified percentage is 4.09 percent at the 95-percent confidence level.

Note that the only difference between this computation and the computation of the sampling error for stratified variables is that $\sqrt{p_h q_h}$ has been substituted for S_h (see tables III.5 and III.16).

The same formulas and the same work sheet would be used if the sampling errors were computed with proportional allocation and Neyman allocation. If proportional allocation were used, the computation of the estimated percentage and number of payroll records with errors would have been much simpler. The estimated percentage is simply the total number of errors found divided by the total sample size:

$$p_{st} = \frac{a}{n}$$

The estimated number of errors is simply the estimated percentage of errors multiplied by the universe size: $\hat{A}_{st} = Np_{st}$. These formulas can be used only with proportional allocation.

- <u>Computing sample sizes</u>

Assume that the evaluators would like to know what sample sizes would be required, using proportional allocation and Neyman

Table III.17

Work Sheet for Computing Payroll Sample Size

Location (h)	Universe size (N_h)	p_hq_h	W_h	$\sqrt{p_hq_h}$	$W_h\sqrt{p_hq_h}$	$W_hp_hq_h$
1	1,100	0.2475	0.3667	0.4975	0.1824	0.09076
2	1,500	0.0475	0.5000	0.2179	0.1090	0.02375
3	400	0.1600	0.1333	0.4000	0.0533	0.02133
	3,000		1.0000		0.3447	0.13584

allocation, to reduce the sampling error of the stratified percentage (E) to 2 percent at the 95-percent confidence level. The computations are shown in table III.17. (Let W_h equal N_h/N.) For proportional allocation, the formula is

$$n = \frac{\sum_{h=1}^{3} W_h p_h q_h}{\left(\frac{E}{t}\right)^2 + \frac{1}{N}\sum_{h=1}^{3} W_h p_h q_h}$$

Using a t factor of 2 for 95-percent confidence, we compute:

$$n = \frac{0.1358}{\left(\frac{0.02}{2}\right)^2 + \frac{0.1358}{3,000}}$$

$$n = 934.8 \quad \text{or} \quad 935$$

The sample would be allocated to the strata in proportion to the values for W_h. This would yield the following sample sizes: for location 1, 343; location 2, 467; location 3, 125 (total = 935). The sample for location 1 was rounded downward.

For Neyman allocation, the formula to compute the sample size is

$$n = \frac{\left(\sum_{h=1}^{3} W_h \sqrt{p_h q_h}\right)^2}{\left(\frac{E}{t}\right)^2 + \frac{1}{N}\sum_{h=1}^{3} W_h p_h q_h}$$

$$n = \frac{(0.3447)^2}{\left(\frac{0.02}{2}\right)^2 + \frac{0.1358}{3,000}}$$

$$n = 817.6 \quad \text{or} \quad 818$$

Table III.18

Work Sheet for Allocating Samples to Strata with Neyman Allocation in Sampling for Attributes

Stratum (h)	$W_h \sqrt{p_h q_h}$	w_h	n_h
1	0.1824	0.5292	433
2	0.1090	0.3162	259
3	0.0533	0.1546	126
Total	0.3447	1.0000	818

In the formula for allocating the sample to the strata, $n_h = nw_h$ and w_h equals

$$\frac{W_h \sqrt{p_h q_h}}{\sum W_h \sqrt{p_h q_h}}$$

The allocation is shown in table III.18. Thus, with Neyman allocation, the specified precision can be obtained by using 117 fewer payroll records than with proportional allocation.

After allocating the final sample, we would compute the estimates and sampling errors.

OTHER TOPICS ON STRATIFICATION

Stratification has some other advantages not mentioned above. One is that, by careful stratification, the evaluators can maximize the dollars protected; that is, they can review the maximum dollar amounts of transactions, documents, or accounts with a given sample size. Another advantage is that, by careful stratification, the evaluators can maximize the number of errors discovered and corrected; that is, they can include error-prone items in one stratum and relatively error-free items in another. They can then sample more heavily from the error-prone stratum. Last, but not least, stratification permits the development of estimates for the individual strata, if such estimates are needed.

Some practicalities of stratification

A word should be said about the realities of stratification in most applications. Textbook illustrations usually assume that (1) the strata were designed by the sampler to increase precision, (2) the stratification is based on the variable being estimated, and (3) the standard deviations are known or can be computed for the variable being estimated. In real life, however, the strata are often defined by the review objectives or the physical location or arrangement of the universe.

The standard deviations for the variable being estimated are not known and must be computed from preliminary samples taken in each stratum. The stratum boundaries and sample sizes (both overall and within strata) must be calculated from some variable other than the variable being estimated, because this is the only variable available. For example, consider the direct air shipment versus air freight forwarder problem. The stratum boundaries were based on air freight forwarder shipping costs, but in real life, shipping costs, not savings, might have to be used to compute standard deviations for optimum allocation. The basis for this is the belief that the variance of the variable being estimated is highly correlated with the variance of the variable used to set the stratum boundaries, calculate the total sample size, and allocate the total sample to the strata.

Sometimes, from a practical point of view, it is just as efficient to use general rules of thumb to determine the total sample size and to allocate the sample to the individual strata on the proportion of the individual stratum total to the grand total. For example, assume that the air freight forwarder shipping costs were

Stratum	Shipping costs	Percent of total
Less than $100	$13,200	24
$100 to $499	25,300	46
$500 or more	16,500	30
Totals	$55,000	100

If we were allocating a sample of 50 items on this basis, we would draw 12 items from the first stratum (24 percent of 50), 23 items from the second (46 percent of 50), and 15 items from the third (30 percent of 50).

This allocation method assumes that the stratum standard deviations of the savings are roughly proportional to the stratum means of the shipping costs. In practice, this method often works out fairly close to the results obtained by using Neyman allocation. If the results are not as precise as required, it may be necessary to increase the sample size in one or more of the strata.

If the air freight forwarder shipping costs were not known for each stratum, another possibility would be to assume that the mean shipping cost per stratum equals the stratum midpoint. (The "$500 or more" stratum presents a problem because it is "open ended"; however, we can often make a reasonable assumption about midpoints for open-ended strata.) Here, for example, assume midpoints of $50 for stratum 1, $300 for stratum 2, and $600 for stratum 3. The allocation of a sample of 50 to the various strata is shown in table III.19 on the next page.

Table III.19

Work Sheet for Allocating Samples to Strata with Assumed Stratum Mean

Stratum: shipping cost	Universe size (N_h)	Assumed midpoint ($\bar{x}_h$)	$N_h\bar{x}_h$	w_h[a]	Sample size (n_h)[b]
Less than $100	150	50	7,500	0.167	8
$100 to $500	75	300	22,500	0.500	25
$500 or more	25	600	15,000	0.333	17
	250		45,000	1.000	50

[a]w_h = percent of total.
[b]n_h = nw_h.

As can be seen, this method gives a different allocation of the sample from that obtained by using Neyman allocation, but it is very close and far better than the results obtained by using proportional allocation. The advantage of this method is that it does not require prior information about the variable being estimated.

Guidelines on constructing strata

A word about the construction of strata is also in order. The evaluators may well ask: How many strata should we have? Where should we set the stratum boundaries? A body of mathematical theory has been developed on how to determine the optimum number of strata and how to set the stratum boundaries, but a discussion of this theory is beyond the scope of this paper. However, some general rules of thumb can be given.

In sampling for variables, when the evaluator frequently uses stratification to minimize the sampling error, six strata are usually sufficient. If the number of strata is increased beyond six, the reduction in sampling error is usually not worth the extra work required.

As for setting the stratum boundaries, if the universe is listed in ascending or descending order of value, the boundary locations usually become obvious when the evaluators scan the list. If the universe is so large that it is not possible to list every item, the evaluators may list a sample of the items, say 5 or 10 percent, sorted in order of value, to set the boundaries. Another possibility is to base a frequency distribution of the items on dollar amounts. For example, consider the frequency distribution based on transaction dollar amounts shown in table III.20.

After examining the frequency distribution, the evaluators may decide to set stratum boundaries at less than $10, $10 to $19, $20 to $49, $50 to $99, $100 to $199, $200 to $499, $500 to $999, and $1,000 or more. This gives a set of strata in which the upper stratum boundary is about twice the lower, except in the lowest and highest strata. Another possibility is to divide

Table III.20

Number and Dollar Amounts of Transactions

Amount of transaction	Number of transactions	Dollar amount	Cumulative amount
Less than $10	4,063	$ 30,879	$ 30,879
$10 to $19	3,323	61,190	92,069
$20 to $29	3,063	70,151	162,220
$30 to $39	2,544	95,909	258,129
$40 to $49	1,424	67,926	326,055
$50 to $59	839	46,145	372,200
$60 to $69	593	39,434	411,634
$70 to $79	397	30,768	442,402
$80 to $89	352	30,976	473,378
$90 to $99	274	26,770	500,148
$100 to $199	194	34,338	534,486
$200 to $299	183	47,214	581,700
$300 to $399	119	39,746	621,446
$400 to $499	61	27,023	648,469
$500 to $599	41	22,427	670,896
$600 to $699	29	18,879	689,775
$700 to $799	20	15,080	704,855
$800 to $899	23	19,040	723,895
$900 to $999	9	8,757	732,652
$1,000 or more	10	14,800	747,452
	17,561	$747,452	

the overall total by the required number of strata to obtain the average dollar amount per stratum. Then the boundaries are set where the cumulative totals are closest to the product of the stratum numbers and the average dollar amount per stratum.

For example, suppose we wanted to have six strata. We divide the total dollar amount, $747,452, by 6 to obtain $124,575, or the average dollar amount per stratum. We then multiply this amount by each of the stratum numbers and obtain

Stratum number	Product	Stratum number	Product
1	$124,575	4	$498,300
2	249,150	5	622,875
3	373,725	6	747,450

Then we look at the cumulative amounts column in the frequency distribution, locate the amounts that are closest to the products, and set the strata boundaries there. Using this system, we obtain the following boundaries:

Cumulative amounts	Stratum boundary
$ 92,069	Less than $20
258,129	$20 to $39
372,200	$40 to $59
500,148	$60 to $99
621,446	$100 to $399
747,452	$400 or more

This method will make the total dollar amount in all strata approximately equal. If equal sample sizes are allocated to all strata, the result approximates Neyman allocation, provided that the stratum standard deviations of the variable being estimated are approximately proportional to the stratum means of the variable being used to set the stratum boundaries.

CLUSTER SAMPLING

Following is an example of a two-stage cluster sampling problem in which the clusters or primary sampling units are selected by simple random sampling.

While reviewing the procurement activity of a government-operated scientific laboratory, the evaluators decide to determine the dollar amount of prompt payment discounts that were lost on invoices paid during the past fiscal year, either because the invoices were not paid promptly or because the discount was not taken. The invoices paid during the fiscal year, together with their supporting documentation, are tied in 2,100 bundles, containing varying quantities of invoices. Thus, each bundle can be defined as a cluster.

Setting the confidence level at 95 percent, the evaluators decide to take a preliminary random sample of 40 clusters. The invoices on which discounts were offered can be identified easily by examining the terms of sale. However, calculating the actual amount of discounts lost involves (1) determining how long the discount period was and whether the invoice was paid within the discount period and (2) multiplying the invoice amount by the discount rate (percent) if the invoice was not paid within the discount period or was paid promptly but the discount was not taken.

The evaluators decide that if a sample bundle contains less than 10 invoices on which discounts were offered, they will calculate the discount lost for all such invoices. But to save work, if the bundle contains 10 or more invoices on which discounts were offered, they will calculate the discount lost for a random sample of 5 invoices. Random number sampling will be used to select the sample invoices. The results of the sample are shown in table III.21.

Our first step is to compute the total amount of discounts lost for each sample cluster (y_i) using the formula below. Let

$$\sum_{j=1}^{m_i} y_{ij}$$

represent the sum of all the sample invoices in sample cluster number i, where j varies from 1 to m_i, the size of the sample in cluster i.

$$y_i = \sum_{j=1}^{m_i} y_{ij}$$

Then we compute the average discount lost for each sample cluster ($\bar{y}_i$).

$$\bar{y}_i = \frac{y_i}{m_i}$$

Next, we compute the estimated total for each sample cluster ($\hat{Y}_i$), using $\hat{Y}_i = M_i \bar{y}_i$. (Note that for clusters in which the sample size is equal to the number of invoices in the cluster, the cluster total equals the sample total.) We then square each cluster total.

Our next step is to compute the estimated total discounts lost for all 2,100 bundles combined ($\hat{Y}_{c1}$). The formula for this

Table III.21

The Results of a Review of 40 Bundles of Requisitions

Sample bundle	Requisitions: No. in bundle	Requisitions: No. sampled	Amounts of discounts lost	Sample total	Sample mean	Cluster total	
(i)	(M_i)	(m_i)	(y_{ij})	(y_i)	($\bar{y}_i$)	($\hat{Y}_i$)	$\hat{Y}_i^2$
1	7	7	44, 0, 32, 17, 0, 0, 0	$ 93	$13.286	$ 93.00	$ 8,649
2	60	5	0, 0, 0, 37, 46	83	16.600	996.00	992,016
3	7	7	0, 0, 0, 0, 0, 50, 6	56	8.000	56.00	3,136
4	48	5	18, 0, 46, 0, 32	96	19.200	921.60	849,347
5	68	5	25, 22, 0, 0, 0	47	9.400	639.20	408,577
6	65	5	2, 0, 0, 0, 35	37	7.400	481.00	231,361
7	70	5	0, 0, 39, 0, 0	39	7.800	546.00	298,116
8	55	5	0, 30, 38, 15, 0	83	16.600	913.00	833,569
9	12	5	19, 0, 0, 0, 19	38	7.600	91.20	8,317
10	4	4	23, 0, 0, 0	23	5.750	23.00	529
11	38	5	29, 25, 0, 7, 0	61	12.200	463.60	214,925
12	9	9	0, 3, 1, 0, 37, 9, 0, 0, 0	50	5.556	50.00	2,500
13	12	5	4, 32, 0, 0, 0	36	7.200	86.40	7,465
14	70	5	14, 0, 0, 10, 0	24	4.800	336.00	112,896
15	8	8	0, 0, 30, 0, 20, 0, 0, 0	50	6.250	50.00	2,500
16	48	5	18, 0, 7, 0, 28	53	10.600	508.80	258,877
17	42	5	0, 0, 0, 0, 21	21	4.200	176.40	31,117
18	2	2	38, 6	44	22.000	44.00	1,936
19	5	5	0, 0, 47, 0, 0	47	9.400	47.00	2,209
20	3	3	31, 0, 43	74	24.667	74.00	5,476
21	65	5	10, 5, 5, 0, 18	38	7.600	494.00	244,036
22	5	5	15, 35, 44, 0, 0	94	18.800	94.00	8,836
23	15	5	0, 15, 0, 13, 41	69	13.800	207.00	42,849
24	7	7	50, 0, 37, 3, 33, 0, 0	123	17.571	123.00	15,129
25	6	6	42, 10, 17, 41, 0, 0	110	18.333	110.00	12,100
26	24	5	8, 15, 1, 0, 0	24	4.800	115.20	13,271
27	1	1	42	42	42.000	42.00	1,764
28	8	8	3, 17, 18, 29, 0, 38, 0, 0	105	13.125	105.00	11,025
29	9	9	0, 1, 21, 13, 0, 0, 43, 0, 0	78	8.667	78.00	6,084
30	5	5	22, 0, 0, 39, 50	111	22.200	111.00	12,321
31	36	5	0, 0, 0, 32, 0	32	6.400	230.40	53,084
32	3	3	42, 0, 0	42	14.000	42.00	1,764
33	4	4	0, 0, 0, 11	11	2.750	11.00	121
34	15	5	0, 16, 22, 0, 46	84	16.800	252.00	63,504
35	13	5	21, 38, 0, 0, 0	59	11.800	153.40	23,532
36	72	5	0, 0, 29, 0, 0	29	5.800	417.60	174,390
37	33	5	0, 8, 0, 0, 0	8	1.600	52.80	2,788
38	85	5	42, 0, 7, 15, 0	64	12.800	1,088.00	1,183,744
39	5	5	23, 0, 16, 24, 0	63	12.600	63.00	3,969
40	18	5	18, 0, 0, 0, 41	59	11.800	212.40	45,114
	1,062	208				$10,598.00	$6,192,942

computation is

$$\hat{Y}_{cl} = \frac{N}{n}\sum^{n}\hat{Y}_i$$

$$\hat{Y}_{cl} = \frac{2,100}{40}(10,598)$$

$$\hat{Y}_{cl} = 556,395$$

Thus, the estimated total discount lost is $556,395.

The computation of the sampling error of the total ($E_{\hat{Y}_{cl}}$) is shown below. Let $S_{\hat{Y}}$ represent the standard deviation of the cluster totals. Then,

$$E_{\hat{Y}_{cl}} = \frac{NtS_{\hat{Y}}}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

Since the sample size is less than 5 percent of the universe, the finite population correction

$$\sqrt{\frac{N-n}{N}}$$

can be dropped and the formula simplified to

$$E_{\hat{Y}_{cl}} = \frac{NtS_{\hat{Y}}}{\sqrt{n}}$$

We compute the standard deviation of the cluster totals by the formulas below. Let

$$\sum^{n}\hat{Y}_i^2$$

equal the sum of the squares of the cluster totals and $\hat{\bar{Y}}$ the mean of the cluster totals (the average discount lost per cluster).

$$\hat{\bar{Y}} = \frac{1}{n}\sum^{n}\hat{Y}_i$$

$$\hat{\bar{Y}} = \frac{10,598}{40}$$

$$\hat{\bar{Y}} = 264.950$$

$$S_{\hat{Y}} = \sqrt{\frac{\sum^{n} \hat{Y}_i^2 - n\hat{\bar{Y}}^2}{n-1}}$$

$$S_{\hat{Y}} = \sqrt{\frac{6{,}192{,}942 - (40)(264.950)^2}{40-1}}$$

$$S_{\hat{y}} = 294.61$$

Substituting into the sampling error formula, we obtain

$$E_{\hat{y}_{cl}} = \frac{(2{,}100)(1.96)(294.61)}{\sqrt{40}}$$

$$E_{\hat{y}_{cl}} = 191{,}731$$

Thus, the sampling error of the total discounts lost is $191,731 at the 95-percent confidence level.

It should be noted that this computation underestimates the sampling error. When subsamples are taken within clusters, two types of variation contribute to the sampling error: the variation between cluster totals and the variation between items within clusters. In this computation of the sampling error, we have ignored the within-cluster variation. For many applications, this is possible if the number of clusters sampled is large (30 or more). However, the fact that the sampling error is underestimated should be recognized. In this particular example, if the sampling error were to be computed by taking into account both the between-cluster variation and the within-cluster variation, it would have been increased by 0.6 percent.

Another method could have been used to estimate the dollar amount of discounts lost. This is the ratio-to-size estimate, which uses the technique of ratio estimation discussed in chapter 4. To use this method, we would have to know the total number of invoices paid during the fiscal year.

In this paper, we have not gone into mathematical methods for computing sample sizes with cluster sampling; however, a few general comments are in order. According to a rule of thumb, we should try to have at least 30 clusters in the sample. A sample with as few as 20 clusters will sometimes give fairly precise results. However, for various reasons that need not be discussed here, estimates developed from a cluster sample are usually much less precise than estimates developed from a simple random sample consisting of the same number of items.

If we are using two-stage sampling and we can make a choice between sampling more items within the cluster, or reducing the number of items sampled within the cluster, and increasing the number of clusters in the sample, it is better to increase the number of clusters and reduce the number of items. This will practically always yield more precise estimates.

A SELECTED ANNOTATED BIBLIOGRAPHY

Arkin, H. Handbook of Sampling for Auditing And Accounting, 3rd ed. New York: McGraw-Hill, 1984.

Contains some lucid discussions of concepts that can round out understanding. Useful reference for statisticians specializing in sampling for auditing and accounting. Description of sample selection procedures especially useful.

Cochran, W. G. Sampling Techniques, 3rd ed. New York: Wiley, 1977.

A classic in the field. Emphasizes the theory of sampling, derivations of formulas, and proofs. Useful as a reference book.

Deming, W. E. Sample Design in Business Research. New York: Wiley, 1960.

Examples are not as appropriate for GAO work as some of the other texts, but provides useful background and reference for statisticians specializing in sampling. Discussions and presentation of material are clear and accurate although somewhat dated. Oriented mathematically.

Eichner, W. R. Statistical Auditing. Washington, D.C.: U.S. General Accounting Office, Office of Organization and Human Development, 1985.

A good, brief introduction to sampling as used in GAO, emphasizing concepts rather than calculations. Used in GAO's course in statistical auditing.

Guy, Donald M. An Introduction to Statistical Sampling in Auditing New York: Wiley, 1981.

An intermediate text, a middle ground between a brief introduction and a highly technical presentation. Primarily for auditors performing financial audits.

Kalton, G. Introduction to Survey Sampling. Sage University Series on Quantitative Applications in the Social Sciences. Beverly Hills, Calif.: Sage Publications, 1983.

An introductory booklet aimed at persons who need to have some background in sampling. Useful as a refresher or summary. Style is straightforward and practical. Covers most of the topics appropriate for GAO work.

Kish, L. Survey Sampling. New York: Wiley, 1965.

One of the classics in the field of survey sampling. Thoroughly covers many details with great emphasis on procedures. Presentation of material very technical and complex. Useful as a reference book.

Levy, P. S., and S. Lemeshow. Sampling for Health Professionals. Belmont, Calif.: Wadsworth, 1980.

An excellent book for individuals who have had a few graduate courses in research and evaluation methods, statistics, etc.

Could be very useful for persons in other empirical fields. Presentations on the use of structured instruments, report writing, and interpretation of results are particularly useful.

Scheaffer, R. L., W. Mendenhall, and L. Ott. Elementary Survey Sampling. Belmont, Calif.: Wadsworth, 1979.
An excellent practical reference book for many fields. Probably the best source of usable formulas. Direct style sets off groups of formulas well.

Slonim, M. J. Sampling in a Nutshell, rev. ed. New York: Simon and Schuster, 1960.
A good brief introduction emphasizing concepts rather than calculation. Covers most of the topics relevant to GAO work. Suitable for undergraduates and general readers.

Sudman, S. Applied Sampling. New York: Academic Press, 1976.
A good book for individuals with a few graduate courses in quantitative methods. Aimed at social researchers doing medium-sized studies. Direct, relatively clear style covers most of the topics applicable to GAO.

Williams, B. A Sampler on Sampling. New York: Wiley, 1977.
An excellent introductory book for nonmathematicians. Furthers understanding by demonstration rather than formal proofs. Covers most of the material necessary for GAO.

FOUR COMPUTER SOFTWARE PACKAGES

This appendix briefly describes SPSS-X, SAS, DYL-AUDIT, and IMSL, four computer software packages that have been found useful for sampling at GAO. They have been available for many years and are used by GAO staff on mainframe time-sharing systems such as that of the National Institutes of Health Computer Center. There are other such packages, so this list should not be considered comprehensive. Furthermore, the descriptions are not intended as evaluations of any of the packages.

In recent years, statistical software packages for use on microcomputers have proliferated. Literally scores of packages are available, and new ones are being introduced regularly. Microcomputer statistical software packages are currently being evaluated separately and, therefore, are not included in this appendix.

Both statistical packages used on mainframe computers and packages used on microcomputers are continually being updated and improved. Current information regarding their utility, capabilities, and accuracy is consequently difficult to maintain. It is recommended that, whenever these packages are used, assistance be requested from the appropriate Technical Assistance Group or Design, Methodology, and Technical Assistance Group.

SPSS-X

SPSS-X, the latest version of the Statistical Package for the Social Sciences, is used at GAO more than all other packages combined. It is available for many brands and sizes of computers.

The writers of the package have made a long and concerted effort to keep in touch with its users, and this shows in its high quality. The documentation is well written and clear. The language that the user writes in is very close to English and includes terminology commonly used by analysts in policy and oversight research applications. The case selection part of the language allows taking simple samples of a fixed size from a universe. It is relatively easy to draw other types of samples using a few lines of code.

The manuals describe the statistical procedures and explain why options should be chosen. They contain the procedures for doing almost all that is needed in GAO work. The ability to produce self-documenting data files greatly aids in referencing jobs and in reproducing analyses. The transformation of data--recoding, scoring, and so on--is very straightforward. Easy inclusion of sampling weights allows the calculation of estimates from complex sampling designs.

As with all the packages available today, SPSS-X does not handle the special case of finite populations well and does not compute accurate sampling errors for complex sample designs. However, it can prepare intermediate results for input to other programs. The output is very clearly organized and can be well labeled. The presentation of results is very flexible, and it is relatively easy to prepare the data so that they can be moved to other machines.

SAS

SAS, or Statistical Analysis System, is available mostly for IBM computers. It is oriented toward mathematical, rather than applied, statistics and analyses.

The written documentation is somewhat clear about the actual coding but does not deal adequately with the effects of choosing options or the reasons for using different options. The language that the user writes in is close to English, but the terminology is not commonly used by analysts in evaluation and policy analysis applications. The ability to produce data files with some self-documenting features aids in referencing jobs and in reproducing analyses. Data transformations are also available. SAS contains the procedures to do most of what is needed in GAO work.

As with all the packages available today, it does not handle the special case of finite populations well. However, it can prepare intermediate results for input to other programs. The output is not very clearly organized, but the presentation of results is somewhat flexible. It is relatively easy to prepare the data so that they can be moved to another IBM computer.

This package is recommended when analysts understand only this package and when the analysis does not call primarily for cross-tabulations. SAS is recommended when specialized econometric or operations research techniques are needed.

DYL-AUDIT

DYL-AUDIT, or DYL, which runs only on IBM computers, is primarily a data retrieval and report package. Some of GAO's Design, Methodology, and Technical Assistance Groups and regional Technical Assistance Groups use it. It is oriented more toward use by data processing specialists than by data analysis specialists. The documentation is reasonably well written and fairly clear. The language that the user writes in is very close to English and includes terminology commonly used by IBM data processors. The case selection part of the language allows taking many kinds of samples by using a few lines of code.

The manuals contain useful descriptions of the sampling procedures and explain why options should be chosen. This documentation can be useful even as background information in sampling. The self-documenting features are somewhat limited, and the data modification features are limited. DYL does have the ability to do grouped frequency counts. It also computes many kinds of subtotals but does not compute all marginal subtotals. It is very useful for extracting data already on an IBM computer. It is not typically used at GAO for calculating estimates after applying measurement techniques to the sample cases. The output is very clearly organized and includes work sheets that aid in data gathering. The presentation of results is very flexible, and it is relatively easy to prepare the data so that they can be moved to other machines.

DYL-AUDIT is recommended for extracting data when there is a very large amount of data and machine efficiency is a major consideration. Because jobs requiring the use of this package are very large, a specialist who knows DYL should be consulted.

IMSL

IMSL, or International Mathematical and Statistical Library, contains computational subroutines written in FORTRAN and has been tested by mathematical and statistical computation specialists. The writers of the library adhered to rigorous standards for computation and documentation. The written documentation is clearly organized and always gives detailed instruction on the input and output to the subroutines. The routines usually contain checks for many kinds of errors. IMSL is oriented toward high-level specialists who need to create programs for functions not included in the standard packages. Versions are available for many sizes and brands of computers. The manuals describe the procedures but do not explain why options should be chosen.

IMSL is recommended when new programs must to be written. Typically, it is used by statisticians who can write in FORTRAN.

REFERENCES

Arkin, H. Handbook of Sampling for Auditing and Accounting, 3rd ed. New York: McGraw-Hill, 1984.

Cochran, W. G. Sampling Techniques, 3rd ed. New York: Wiley, 1977.

Deming, W. E. Sample Design in Business Research. New York: Wiley, 1960.

Dixon, W. J., and F. J. Massey. Introduction to Statistical Analysis, 3rd. ed. New York: McGraw-Hill, 1969.

Gantz, D. T. Personal letter dated February 2, 1984.

GAO, Program Evaluation and Methodology Division. Designing Evaluations. Methodology Transfer Paper 4. Washington, D.C.: July 1984.

-----. Using Structured Interviewing Techniques. Methodology Transfer Paper 5. Washington, D.C.: July 1985.

Ijiri, Yuji, and Robert S. Kaplan. "The Four Objectives of Sampling in Auditing: Representative, Corrective, Protective, and Preventive." Management Accounting, 52:60 (December 1970), 42-44.

Kalton, G. Introduction to Survey Sampling. Sage University Series on Quantitative Applications in the Social Sciences. Beverly Hills, Calif.: Sage Publications, 1983.

Kish, L. Survey Sampling. New York: Wiley, 1965.

Kruskal, W., and F. Mosteller. "Representative Sampling." International Statistical Review, 47 (1979), 13-24, 111-27, 245-65; 48 (1980), 169-95.

Levy, P. S., and S. Lemeshow. Sampling for Health Professionals. Belmont, Calif.: Wadsworth, 1980.

Scheaffer, R. L., W. Mendenhall, and L. Ott. Elementary Survey Sampling. Belmont, Calif.: Wadsworth, 1979.

Slonim, M. J. Sampling in a Nutshell, rev. ed. New York: Simon and Schuster, 1960.

Snedecor, G. W., and W. G. Cochran. Statistical Methods, 7th ed. Ames, Iowa: The Iowa State University Press, 1980.

Sudman, S. Applied Sampling. New York: Academic Press, 1976.

Williams, B. A Sampler on Sampling. New York: Wiley, 1977.

GLOSSARY

Attribute. An inherent quality or characteristic that an item either has or does not have. The quality can be either a simple characteristic, such as being or not being a high school graduate, or a complex one, such as conforming or not conforming to specifications.

Bias. The existence of a factor that causes an estimate made on the basis of a sample to differ systematically from the universe parameter being estimated. Bias may originate from poor sample design, deficiencies in carrying out the sampling process, or an inherent characteristic of the estimating technique used.

Cluster sample. A simple random sample in which each sampling unit is a collection of elements.

Confidence coefficient, or confidence level. A measure (usually expressed as a percentage) of the degree of assurance that the estimate obtained from a sample differs from the universe parameter being estimated by less than the sampling error. In this document, we use the letter "t" to represent the confidence coefficient although, in theory, we should use "z" when the sample size is 30 or more and shows that we used the normal distribution and "t" when the sample size is less than 30 and shows that we used the student's t distribution.

Correlation. The interdependence between two sets of numbers; a relation between two quantities, such that when one changes, the other changes. Simultaneous increasing or decreasing is called "positive correlation"; one increasing and the other decreasing is called "negative correlation."

Domains of interest. Classes into which a universe may be subdivided so that separate estimates can be developed for each domain. This is different from stratification, because a domain of interest can extend across several strata and because the classification may be based on the sample data.

Finite population correction. A multiplier that makes adjustments for the sampling efficiency gained when sampling is without replacement and when the sample size is large (greater than 5 percent) with respect to the universe size. This multiplier reduces the sampling error for a given sample size or reduces the required sample size for a specified precision.

Frequency distribution. A table in which data are grouped into classes and the number of items that fall into each class are recorded.

Mean. The sum of all the values in a set of observations divided by the number of observations. Also known as "average" or "arithmetic mean," it indicates the typical value for a set of observations.

Optimum allocation. A method of allocating a sample to strata by taking into account not only the difference in strata universe sizes and standard deviations but also the differences in the costs of collecting data for the various strata.

Parameter. A measure such as mean, median, standard deviation, or proportion that defines or describes an attribute or a characteristic of a universe.

Probability. The ratio of the number of outcomes that will produce a specific event to the total number of possible outcomes, or the likelihood that specific events will occur, expressed as a proportion or percentage.

Probability sampling. The selection of a sample by some random method to obtain information or draw conclusions about a universe. Each possible sample from the universe, and thus each item in the universe, has a known (nonzero) probability of being selected.

Random decimal digits. A table of digits 0 through 9 arranged so that digits may be randomly selected according to any procedure, subject to the sole restriction that a digit's selection be influenced only by its location in the table. Its purpose is to permit the drawing of random samples.

Random number sampling. A sampling method in which combinations of random digits, within the range of the number of items in a universe, are selected from a table of random decimal digits until a given sample size is obtained. For example, if a sample of 60 items is required from a universe numbered 1 through 2000, then 60 combinations of digits between 0001 and 2000 are selected.

Random selection. A selection method that uses an acceptable table of random numbers in a standard manner. The method minimizes the influence of nonchance factors in selecting the sample items.

Regression. The line of average relationship between the dependent (or primary) variable and the independent (or auxiliary) variable.

Regression coefficient. A measure of change in a primary variable associated with a unit change in the auxiliary variable.

Sample. A portion of a universe that is examined or tested in order to obtain information or draw conclusions about the entire universe.

Sampling error, or precision. A measure of the expected difference between the value found in a probability sample and the value of the same characteristic that would have been found by examining the entire universe. Sampling errors are always stated at a specific confidence level.

Sampling frame. A means of access to a universe, usually a list of the sampling units contained in the universe. The list may be printed on paper, a magnetic tape file, a file of punch cards, or a physical file of such things as payroll records or accounts receivable.

Sampling units. The elements into which a universe is divided; they must cover the whole universe and not overlap, in the sense that each element in the universe belongs to one and only one unit.

Sampling with replacement. A sampling method in which each item selected for a sample is returned to the universe and can be selected again. In this method, the universe can be regarded as infinite.

Sampling without replacement. A sampling method in which an item selected for a sample is "used up": it is not returned to the universe and cannot be selected again. In this method, the universe can be regarded as finite.

Scientific sampling. See Probability sampling.

Standard deviation. A numerical measurement of the dispersion, or scatter, of a group of values about their mean. Also called "root-mean-square" deviation.

Statistic. A measure, such as a mean, proportion, or standard deviation derived from a sample and used as a basis for estimating the universe parameter.

Statistical estimate. A numerical value assigned to a universe parameter on the basis of evidence from a sample.

Statistical sampling. See Probability sampling.

Strata. Two or more mutually exclusive subdivisions of a universe defined in such a way that each sampling unit can belong to only one subdivision or stratum.

Systematic selection with a random start. A sampling method in which a given sample size is divided into the universe size in order to obtain a sampling interval. A random starting point between 1 and the sampling interval is obtained. This item is selected first; then every item whose number or location is equal to the previously selected item plus the sampling interval is selected, until the universe is used up.

Tolerable error. The specified precision of the maximum sampling error that will still permit the results to be useful.

Universe, or population. All the members of a group to be studied as defined by the evaluators; the total collection of individuals or items from which a sample is selected.

Variable. A characteristic having values that can be expressed numerically or quantitatively and that may vary from one observation to another. Examples are the dollar amount of error in a voucher, a quantity shipped, and the height of a person.

(973178)