1972

Confidence intervals for quantiles in stratified random sampling

John Sigmund Meyer
Iowa State University

73-9464

MEYER, John Sigmund, 1937-
CONFIDENCE INTERVALS FOR QUANTILES IN STRATIFIED
RANDOM SAMPLING.

Iowa State University, Ph.D., 1972
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

Confidence intervals for quantiles

in stratified random sampling

by

John Sigmund Meyer

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1972

TABLE OF CONTENTS

I. INTRODUCTION AND REVIEW OF LITERATURE

Most of the theory of sampling from finite populations pertains to point estimation of desired parameters such as the finite population mean and variance. Researchers and practitioners have generally been reluctant to make any assumptions about the distributions of the relevant random variables. This may be due to the very wide variety of finite populations met in practice, and the inherent lack of "smoothness" of many finite populations. Thus, the literature contains very little discussion about the formation of confidence intervals, tests of significance, etc., when sampling is from a finite population. Most of this (limited) literature pertains to simple random sampling and the use of normal approximations for the sampling distributions of the estimators (and, often, approximation of the distribution of a test statistic by a t-distribution). (For a general treatment, see Cochran [1963, Sections 2.7 and 2.13].) In no case will the confidence coefficient associated with such (approximate) confidence intervals be known exactly, and the validity of the normal approximations may be questioned (at least) for many of the extremely skewed finite populations encountered in practice--especially if sample sizes are small. When one employs a sample design more complex than simple random sampling, the validity of using normal (and "t") approximations for the sampling distributions of the estimators has received even less attention than for the case of simple random sampling. This is well illustrated by Cochran [1963, Section 5.4] in his discussion about confidence intervals for the finite population mean when stratified simple random sampling is used, and his lack of such discussion for sample designs such as single- and multi-stage cluster sampling.

It is clear from the considerations noted above that even for simple random sampling, the development of confidence intervals having known confidence coefficients requires an alternative approach. For the more complex sample designs used in practice, theoretical investigations of the use of normal (and "t") approximations appear to be quite formidable, and the validity of approximate confidence coefficients would, again, depend heavily on the type of finite population being sampled. Further, the validity of these approximations may depend on more assumptions (e.g., the random variable Y has a normal distribution in each stratum) than one would ordinarily wish to make.

In this thesis we suggest confidence interval procedures for any specified quantile having the property that one may determine the exact associated confidence coefficient for any finite population. Although quantiles (e.g., the finite population median) are of great interest to many practitioners (because, for example, of the highly skewed distributions encountered in applications), estimation of such parameters has not received much attention in the sample survey literature. Perhaps this is due to the difficulty of determining the properties of appropriate point estimators when sampling is from a finite population. However, it will be shown that if either simple random or stratified simple random sampling is employed, it is feasible to determine a confidence interval for any quantile with known confidence coefficient. It is apparent that the procedures to be described can be extended to other sample designs. These confidence interval methods are very simple to apply since the upper and lower confidence limits are either given by a pair of order statistics or are derived from the sample cumulative distribution function (C.D.F.). Given the sample design and confidence interval procedure, it is straightforward to tabulate, using an electronic computer, the possible confidence intervals and associated confidence coefficients. Given this tabulation, the simplicity of the confidence interval procedure should make its use attractive in places where sophisticated personnel and machines are not available, and for applications where preliminary estimates of specified parameters are required very quickly. For two of the confidence interval procedures, a computer program to evaluate the exact confidence coefficients has been written for use of the UNIVAC 1108 computer at the University of Wisconsin Computing Center or the IBM 1130 computer at Cornell College.

Wilks [1962, Section 11.4] has suggested a confidence interval procedure if simple random sampling from a finite population is used. His results, some amplification of these results, brief tables, and some generalizations, including joint confidence intervals and expected lengths of confidence intervals are discussed in Chapter II. In Chapter III, stratified simple random sampling is considered for the case of $L = 2$ strata and the three confidence interval procedures are introduced. In each case an expression for the exact confidence coefficient is derived, and (where appropriate) an approximation is suggested. Brief tables are included, and the three techniques are compared, both theoretically and by Monte Carlo methods.

A parallel development for two of the methods is given in Chapter IV for $L = 3$ strata; these results suggest how extensions to four or more strata can be made. When L strata are formed by utilizing the known distribution of a concomitant variable, X (which is closely related to the variable of interest, Y), considering only two or three strata may not be restrictive. Given the stratification and a general knowledge of the relation between X and Y, it may be reasonable to assert, a priori, that the (t/N)-th quantile of the finite population (e.g., the finite population median) is a variate value from one of only two or three of the L strata. Then, the methods described in Chapters III and IV may be used.

In Chapter V, several extensions and applications of the preceding chapters are explored. These include tolerance regions, the "best" finite population problem, and an extension of the previous work to cluster sampling.

In order to get the main ideas contained in this thesis, the reader may first wish to examine Sections A through C of Chapter II, and then turn to Sections A through D of Chapter III, to follow the three proposed methods.

The problem we are dealing with was given by Thompson [1936, pp. 122-128], in which he gives an equation for the probability of coverage of the median of a finite population by two selected symmetric order statistics of a simple random sample of size n drawn from the population.

Savur [1937, pp. 564-576] derives what he calls a "new test of significance", using the median. In his paper he first gives justification for considering the median, he then suggests the symbol प for it, and finally gives symmetric confidence intervals for the median and, from these, gets his tests of significance. He restricts his work strictly to a continuous C.D.F.

Nair [1940, pp. 551-558] takes the work of Thompson and Savur, and tabulates, for a continuous distribution, the probability that given symmetric order statistics from the sample will cover the median. He assumes an infinite population, and also points out a slight discrepancy in Savur's work.

Wilks [1962, p. 333] gives a brief description of confidence intervals for quantiles in finite populations, again, in the manner of Thompson, but generalizing to any pair of order statistics. We illustrate the work of Thompson and Wilks in Chapter II.

McCarthy [1965, pp. 772-783] discusses non-parametric methods for confidence intervals from stratified populations. In his work he derives a lower bound for the confidence coefficient, if proportional allocation is used. He assumes a continuous C.D.F. in his work. We consider his results in Chapter III, Section B.7.

Woodruff [1952, pp. 635-646] proposes a method using the sample C.D.F. to obtain a confidence interval for quantiles, utilizing general sampling plans. He also assumes a continuous C.D.F., and gives approximate confidence coefficients for his intervals. We describe his work in more detail in Chapter III, Section C.4.

Finally, it may be noted that Loynes [1966, pp. 497-512] investigates various aspects of point estimation of population quantiles when random and stratified random sampling are employed. For random sampling, distribution-free estimation procedures are obtained, and the admissible estimators are identified. Throughout, the population is assumed to be infinite.

II. THE NON-STRATIFIED SITUATION

In this chapter we investigate confidence intervals for the t-th ordered value in a finite population, when simple random sampling is used.

A. Definitions and Statement of Problem

Let Π_N be a population of N elements, whose elements have distinct Y-values associated with them. These Y-values can be simply ordered as

$$Y_{(1)} < Y_{(2)} < \cdots < Y_{(N)} . \tag{2.1}$$

Let t be a fixed integer in the range $1 \leq t \leq N$. We may then regard $Y_{(t)}$ as the (t/N)-th quantile of the population Π_N. In general, we define the λ-th quantile of our population to be $Y_{(N\lambda)}$ if $N\lambda$ is an integer, and to be $Y_{([N\lambda]+1)}$ otherwise, where $[\cdot]$ denotes the greatest integer function.

A simple random sample of size n is drawn without replacement from Π_N. We denote the values associated with the sample elements, after ordering, by

$$y_{(1)} < y_{(2)} < \cdots < y_{(n)} . \tag{2.2}$$

We wish to consider two-sided confidence intervals for $Y_{(t)}$ of the form $[y_{(k)}, y_{(r)}]$, where $1 \leq k < r \leq n$ and $k \leq t$. We first turn our attention to the computation of the confidence coefficients of these observable random intervals:

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \; . \tag{2.3}$$

B. Calculation of the Confidence Coefficient

In computing the confidence coefficient, we first note that

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} = P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} \leq Y_{(t-1)}\} \; . \tag{2.4}$$

This follows from the facts that

$$\begin{aligned} \{y_{(k)} \leq Y_{(t)}\} &= \{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \cup \{y_{(r)} < Y_{(t)}\} \\ &= \{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \cup \{y_{(r)} \leq Y_{(t-1)}\} \end{aligned} \tag{2.5}$$

and the latter events are disjoint.

Therefore, to evaluate (2.4), it is sufficient to arrive at an expression for $P\{y_{(k)} \leq Y_{(t)}\}$.

1. <u>The first approach</u>

Our first approach uses a technique suggested in Hogg and Craig [1970, p.352]. If we define the event

$\{A_i\}$ = {exactly i sample elements have values less than or equal to $Y_{(t)}$}

we then have

{At least k sample elements have values less than or equal to $Y_{(t)}$}

$$= \bigcup_{i=k}^{n} \{A_i\} \; . \tag{2.6}$$

Hence,

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{i=\max[k,t+n-N]}^{\min[t,n]} P\{A_i\}$$

$$= \sum_{i=k}^{n} \binom{t}{i}\binom{N-t}{n-i} \Big/ \binom{N}{n} \; . \tag{2.7}$$

By virtue of the definition

$$\binom{N}{k} = 0 \quad \text{if} \quad k < 0 \quad \text{or} \quad k > N \tag{2.8}$$

we can eliminate the maximum and minimum expressions in the limits of the summation.

Equations (2.4) and (2.7) lead to the following theorem.

Theorem 2.1: Using the notation in Section A,

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$$

$$= \left[\binom{t-1}{r-1}\binom{N-t}{n-r} + \sum_{i=k}^{r-1} \binom{t}{i}\binom{N-t}{n-i}\right] \Big/ \binom{N}{n} \; . \tag{2.9}$$

Proof: Using (2.4) and (2.7) we have

$$\binom{N}{n} P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$$

$$= \sum_{i=k}^{n} \binom{t}{i}\binom{N-t}{n-i} - \sum_{i=r}^{n} \binom{t-1}{i}\binom{N-t+1}{n-i}$$

$$= \sum_{i=k}^{r-1} \binom{t}{i}\binom{N-t}{n-i} + \sum_{i=r}^{n} \binom{t}{i}\binom{N-t}{n-i} - \sum_{i=r}^{n} \binom{t-1}{i}\binom{N-t+1}{n-i} \; .$$

Using the readily derivable combinatorial identity

$$\sum_{i=r}^{n} \binom{t}{i}\binom{N-t}{n-i} - \sum_{i=r}^{n} \binom{t-1}{i}\binom{N-t+1}{n-i} = \binom{t-1}{r-1}\binom{N-t}{n-r} \qquad (2.10)$$

on the last two terms yields the desired result.

An alternative method of looking at Theorem 2.1 is as follows:

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} = \sum_{i=k}^{r-1} P\{A_i\} + P\{A_r \cap \{y_{(r)} = Y_{(t)}\}\} . \qquad (2.11)$$

Writing the above expressions in their hypergeometric forms yields the conclusion.

2. The second approach

Our second approach to the problem of evaluating $P\{y_{(k)} \leq Y_{(t)}\}$ is suggested by Wilks [1962, p. 333]. The event $\{y_{(k)} \leq Y_{(t)}\}$ is written as the union of the disjoint events

$$\{y_{(k)} \leq Y_{(t)}\} = \{y_{(k)} = Y_{(t)}\} \cup \{y_{(k)} = Y_{(t-1)}\} \cdots \{y_{(k)} = Y_{(k)}\} . \qquad (2.12)$$

We then have

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{i=0}^{t-k} P\{y_{(k)} = Y_{(t-i)}\}$$

$$= \sum_{i=\max[0,n-k+t-N]}^{t-k} \binom{t-i-1}{k-1}\binom{N-t+i}{n-k} \Big/ \binom{N}{n} . \qquad (2.13)$$

This yields the confidence coefficient

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$$

$$= \left[\sum_{i=0}^{t-k} \binom{t-i-1}{k-1}\binom{N-t+i}{n-k} - \sum_{i=0}^{t-r-1} \binom{t-i-2}{r-1}\binom{N-t+i+1}{n-r} \right] \Big/ \binom{N}{n} \quad . \tag{2.14}$$

The equality of (2.9) and (2.14) can be demonstrated algebraically.

3. One-sided confidence intervals

It may, on occasion, be impractical to utilize a two-sided confidence interval. This would be especially true if we were attempting to estimate a very low or high population quantile. In such circumstances, one-sided intervals of the form $(-\infty, y_{(r)}]$ or $[y_{(k)}, \infty)$ could be much more practical.

In the case of forming a one-sided confidence interval for $Y_{(t)}$, t small, $t > 1$, the confidence coefficient associated with $(-\infty, y_{(r)}]$ is given by

$$\begin{aligned} P\{-\infty < Y_{(t)} \leq y_{(r)}\} &= 1 - P\{y_{(r)} < Y_{(t)}\} \\ &= 1 - P\{y_{(r)} \leq Y_{(t-1)}\} \ . \end{aligned} \tag{2.15}$$

Similarly, for t relatively large,

$$P\{y_{(k)} \leq Y_{(t)} < \infty\} = P\{y_{(k)} \leq Y_{(t)}\} \ . \tag{2.16}$$

C. Symmetric Confidence Intervals for the Median

As a special case of the above derivations, we consider confidence intervals of the form $[y_{(k)}, y_{(n-k+1)}]$ for the median of an odd-sized population $(N = 2m - 1)$ with distinct variate values. Such confidence intervals are called symmetric and were studied by Thompson [1936, p. 122]. Our reason for looking at an odd-sized population is that the median is well-defined, since $Y_{median} = Y_{(m)} = Y_{\left(\frac{N+1}{2}\right)}$.

Theorem 2.2: If N is odd, and if $m = (N + 1)/2$,

$$P\{y_{(k)} \le Y_{(m)} \le y_{(n-k+1)}\} = 2 \sum_{i=k}^{n} \binom{m}{i}\binom{m-1}{n-i} \Big/ \binom{N}{n} - 1 \; . \tag{2.17}$$

Proof: From Theorem 2.1, we have

$$\binom{N}{n} P\{y_{(k)} \le Y_{(m)} \le y_{(n-k+1)}\} = \binom{m-1}{n-k}\binom{m-1}{k-1} + \sum_{i=k}^{n-k} \binom{m}{i}\binom{m-1}{n-i} \; .$$

Applying identity (2.10), this becomes

$$\sum_{i=n-k+1}^{n} \binom{m}{i}\binom{m-1}{n-i} - \sum_{i=n-k+1}^{n} \binom{m-1}{i}\binom{m}{n-i} + \sum_{i=k}^{n-k} \binom{m}{i}\binom{m-1}{n-i}$$

$$= \sum_{i=k}^{n} \binom{m}{i}\binom{m-1}{n-i} - \binom{N}{n} + \sum_{i=0}^{n-k} \binom{m-1}{i}\binom{m}{n-i} \; .$$

Changing the index on the last summation, we have

$$\sum_{i=0}^{n-k} \binom{m-1}{i}\binom{m}{n-i} = \sum_{j=k}^{n} \binom{m-1}{n-j}\binom{m}{j}$$

and the result follows directly.

Thompson [1936, pp. 126-128] investigates the case of using symmetric confidence intervals for the median. He defines a function

$$\psi(r,s,r',s') = \sum_{\alpha=0}^{\min[r',s]} \binom{r+r'+1}{r+1+\alpha}\binom{s+s'+1}{s-\alpha} \Big/ \binom{r+s+r'+s'+2}{r+s+1} . \tag{2.18}$$

Then he demonstrates that

$$P\{y_{(k)} \leq Y_{(t)}\} = \psi(k-1,n-k,t-k,\ k+N-t-n-1)\ , \tag{2.19}$$

which reduces to equation (2.7), and

$$P\{y_{(k)} \leq Y_{(m)} \leq y_{(n-k+1)}\} = 1 - 2\psi(k-1,n-k,m-k,k+N-m-n-1)\ . \tag{2.20}$$

This latter term is in error; the correct result is the negative of (2.20).

For completeness we consider the case where our population is of size $N = 2m$. In this situation, we define $Y_{median} = (Y_{(m)} + Y_{(m+1)})/2$; hence $Y_{(m)} < Y_{median} < Y_{(m+1)}$. Since "median" is not integral, Theorem 2.2 does not apply.

However,

$$\begin{aligned} P\{y_{(k)} &\leq Y_{med} \leq y_{(n-k+1)}\} \\ &= P\{y_{(k)} \leq Y_{med}\} - P\{y_{(n-k+1)} < Y_{med}\} \\ &= P\{y_{(k)} \leq Y_{(m)}\} - P\{y_{(n-k+1)} \leq Y_{(m)}\} \\ &= \left[\sum_{i=k}^{n} \binom{m}{i}\binom{m}{n-i} - \sum_{i=n-k+1}^{n} \binom{m}{i}\binom{m}{n-i}\right] \Big/ \binom{N}{n} \\ &= \sum_{i=k}^{n-k} \binom{m}{i}\binom{m}{n-i} \Big/ \binom{N}{n}\ . \end{aligned} \tag{2.21}$$

D. Non-distinct Values in the Population

We now examine the situation where the elements in our finite population do not necessarily have distinct Y-values.

Theorem 2.3: Let Π_N be defined as in Section A of this chapter. For fixed n, k, r, and t, we denote $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$, as given in Theorem 2.1, by $1 - \alpha_{k,r}^{n,t,N}$.

Let Π_N^* be a population of size N in which some of the associated Y^*-values may be equal. Order the population values as

$$Y_{(1)}^* \leq Y_{(2)}^* \leq \cdots \leq Y_{(N)}^* .$$

A simple random sample of size n is drawn without replacement from Π_N^* and ordered as

$$y_{(1)}^* \leq y_{(2)}^* \leq \cdots \leq y_{(n)}^* .$$

Then

$$P\{y_{(k)}^* \leq Y_{(t)}^* \leq y_{(r)}^*\} \geq 1 - \alpha_{k,r}^{n,t,N} . \qquad (2.22)$$

Proof: Let $P^*\{\cdot\}$ indicate probabilities associated with Π_N^*, and $P\{\cdot\}$ probabilities associated with Π_N. We show

$$P^*\{y_{(k)}^* \leq Y_{(t)}^* \leq y_{(r)}^*\} \geq P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} .$$

We first note that, if $Y_{(t)}^* < Y_{(t+1)}^*$, then

$$P^*\{y_{(k)}^* \leq Y_{(t)}^*\} = P\{y_{(k)} \leq Y_{(t)}\} .$$

If $Y_{(t)}^* = Y_{(t+1)}^*$, then

$$P^*\{y_{(k)}^* \leq Y_{(t)}^*\} > P\{y_{(k)} \leq Y_{(t)}\} .$$

Similarly, if $Y_{(t-1)}^* < Y_{(t)}^*$,

$$P^*\{y_{(r)}^* < Y_{(t)}^*\} = P\{y_{(r)} < Y_{(t)}\} ,$$

and if $Y_{(t-1)}^* = Y_{(t)}^*$,

$$P^*\{y_{(r)}^* < Y_{(t)}^*\} < P\{y_{(r)} < Y_{(t)}\} .$$

Therefore,

$$\begin{aligned} P^*\{y_{(k)}^* \leq Y_{(t)}^* \leq y_{(r)}^*\} &= P^*\{y_{(k)}^* \leq Y_{(t)}^*\} - P^*\{y_{(r)}^* < Y_{(t)}^*\} \\ &\geq P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} < Y_{(t)}\} \\ &= P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \\ &= 1 - \alpha_{k,r}^{n,t,N} . \end{aligned}$$

E. Systematic Sampling

Since systematic sampling has certain practical advantages (Cochran [1963, Chapter 8] and Sedransk [1969, p. 39]), we show in this section that the confidence coefficient associated with $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ is the same if the sample is a simple random sample from Π_N, or if the sample is a systematic sample, where the population is in "random" order (Sedransk [1969, p. 40]).

Theorem 2.4: Using the notation of this chapter,

$$P_{SRS}\{y_{(k)} \le Y_{(t)} \le y_{(r)}\} = P_{SYS}\{y_{(k)} \le Y_{(t)} \le y_{(r)}\} , \tag{2.23}$$

where, by $P_{SRS}\{\cdot\}$ we mean the probability when simple random sampling is employed, and by $P_{SYS}\{\cdot\}$ we mean the sample is drawn by systematic sampling from a population in "random" order.

Proof: It is sufficient to prove that a particular sample is equally likely to be drawn under either sampling design.

Let $(e_{i_1}, e_{i_2}, \ldots, e_{i_n}) = \underline{e}$ be any n specified elements, for $i_j \in (1, 2, \ldots, N)$, $i_j \neq i_k$. Under simple random sampling,

$$P\{\underline{e}\} = \binom{N}{n}^{-1} .$$

For systematic sampling,

$$\begin{aligned} P\{\underline{e}\} &= \sum_{i=1}^{k} P\{\underline{e} \mid \text{starting point of sample is } i\}\, P\{i\} \\ &= \sum_{i=1}^{k} \binom{N}{n}^{-1} \frac{1}{k} \\ &= \binom{N}{n}^{-1} , \end{aligned}$$

since $P\{\underline{e} \mid \text{starting point of sampling is } i\} = (N-n)!\, n!/N!$ by counting.

F. Sampling from a Continuous C.D.F.

We now consider the situation where our sample is drawn from a population with continuous cumulative distribution function $F(x)$.

More specifically, let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ represent the values of a random sample of size n from a population with continuous C.D.F. $F(x)$. It is well known (Walsh [1962b, pp. 137-138], Wilks [1962, pp. 329-331], Thompson [1936, pp. 122-128], Savur [1937, pp. 564-576], Nair [1940, pp. 551-558], David [1970, p. 14]) that

$$P\{x_{(k)} \leq X_p \leq x_{(r)}\} = I_p(k,n-k+1) - I_p(r,n-r+1)$$

$$= \sum_{i=k}^{r-1} \binom{n}{i} p^i (1-p)^{n-i} \tag{2.24}$$

where X_p denotes the p-th quantile of the population, defined as $\int_{-\infty}^{X_p} f(x)\, dx = p$, and $I_p(k, n-k+1)$ is Karl Pearson's Incomplete Beta function

$$I_y(v_1,v_2) = \int_0^y \frac{\Gamma(v_1+v_2)}{\Gamma(v_1)\,\Gamma(v_2)} x^{v_1-1} (1-x)^{v_2-1}\, dx \qquad 0 < y < 1 . \tag{2.25}$$

David [1970, p. 14] gives an approximation for the symmetric confidence interval for the median with confidence coefficient $1 - \alpha$. His technique, using the normal approximation to the binomial, is to count off $\pm 1/2\sqrt{n}\; u_\alpha$ observations from the sample median and rounding to the nearest integer, where u_α is the upper $\alpha/2$ significance point of the standard normal distribution.

David also cites other work related to confidence intervals when a continuous C.D.F. is assumed.

G. Sampling from a Discrete C.D.F.

In the case where our distribution function is discrete, we have the following.

Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ denote an ordered random sample of size n on a variate X which may take on the values $0, 1, 2, \ldots$ with probabilities $p(0), p(1), p(2), \ldots$, respectively, where $p(\alpha) \geq 0$ and $\sum_{\alpha} p(\alpha) = 1$. When X takes on only a finite number of values (say $0, 1, \ldots, M$) let $p(M+\alpha) = 0$, $\alpha = 1, 2, \ldots$.

Let $P(x) = \sum_{\alpha=0}^{[x]} p(\alpha)$ be the distribution function of X. Let p be a fixed real number such that $0 < p < 1$. Define β to be that integer such that

$$P(\beta - 1) < p \leq P(\beta) \ .$$

Now,

$$P\{x_{(k)} \leq \beta \leq x_{(r)}\}$$

$$= P\{x_{(k)} \leq \beta\} - P\{x_{(r)} < \beta\}$$

$$= P\{x_{(k)} \leq \beta\} - P\{x_{(r)} \leq \beta-1\} \ .$$

Khatri [1963, p. 168] gives

$$P\{x_{(k)} \leq \beta\} = k \binom{n}{k} \int_0^{P(\beta)} \omega^{k-1} (1-\omega)^{n-k} \, d\omega = I_{P(\beta)}(k, n-k+1) \ .$$

Therefore,

$$P\{x_{(k)} \leq \beta \leq x_{(r)}\} = I_{P(\beta)}(k, n-k+1) - I_{P(\beta-1)}(r, n-r+1) \ . \qquad (2.26)$$

This term is greater than or equal to

$$I_p(k,n-k+1) - I_p(r,n-r+1) ,$$

the corresponding result for the continuous case.

Furthermore,

$$\begin{aligned}
P\{x_{(k)} < x_p < x_{(r)}\} \\
&= P\{x_{(k)} \leq \beta-1\} - P\{x_{(r)} \leq \beta\} \\
&= I_{P(\beta-1)}(k,n-k+1) - I_{P(\beta)}(r,n-r+1) \\
&\leq I_p(k,n-k+1) - I_p(r,n-r+1) . \qquad (2.27)
\end{aligned}$$

This result, with a different proof, is given in a theorem by Scheffe and Tukey [1945, pp. 187-192] and is also noted by David [1970, p. 14] and Noether [1967, p. 39].

Finally, we note by Khatri [1963, p. 170],

$$\begin{aligned}
E[y_{(r)} - y_{(k)}] &= \sum_{x=0}^{\infty \text{ or } M-1} \sum_{m=1}^{r-k} \binom{n}{r-m} [P(x)]^{r-m}[1 - P(x)]^{n-r+m} \\
&= \sum_{p=k}^{r-1} \sum_{x=0}^{\infty \text{ or } M-1} \binom{n}{p} [P(x)]^{p} [1 - P(x)]^{n-p} , \qquad (2.28)
\end{aligned}$$

a result analogous to the x_p approach of K. Pearson [1901, p. 391].

H. Tables and Charts for $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$

No direct tabulations for $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ with respect to simple random sampling from Π_N are, in general, available.

However, for the continuous case, Nair [1940, pp. 556-557] tabulates the smallest symmetric confidence intervals for the median, where by smallest he means $[(n - k + 1) - k]$ is minimized. To be precise, Nair tabulates, for samples of size $n = 6\ (1)\ 81$, k and $(n - k + 1)$ such that k is maximized, subject to

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \geq 1 - \alpha ,$$

where $1 - \alpha$ takes on the values .95 and .99.

Chung and DeLury [1950], in their book of charts, are concerned with the following problem: Given x defectives in a sample of size n from Π_N, what are the confidence intervals for k, the number of defectives in Π_N. They use population sizes of $N = 500$, 2500, and 10,000, and confidence coefficients of $(1 - \alpha) = .90$, .95, and .99. Their charts are based on "equal tail" probabilities, and due to the nature of the charts, and the large population sizes, only approximate answers can be expected. In addition, an error made in the construction of the charts, as explained in an errata sheet, makes the charts more difficult to use. However, for the limited cases $N = 500$, 2500, and 10,000, it is possible to utilize the Chung and DeLury charts to find k and r such that

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \geq 1 - \alpha ,$$

where $1 - \alpha = .90$, .95, and .99.

In view of the absence of published tables, in Tables 1, 2, and 3 we give a limited tabulation of the confidence coefficients

$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ for populations of sizes 39, 99, and 199, samples of size 10, and two values of t, corresponding to approximately the 25-th percentile and the median. The tables were computed using the UNIVAC 1108 computer at the University of Wisconsin Computer Center.

Table 4 tabulates the confidence coefficients under the assumptions of a continuous population being sampled. For aid in comparison with the first three tables, the sample size chosen was 10 and the 25-th and 50-th percentiles were picked. A more complete table for the median of a continuous population can be found in MacKinnon [1964, pp. 937-947].

It is of interest to compare Tables 1-3. We see that the confidence coefficient decreases as N increases for the usual type of interval--approximately symmetrical for the median, skewed to the left for the 25-th percentile.

Comparing these tables with Table 4, we note that a similar comment holds and that, for the usual types of intervals, the continuous C.D.F. gives a lower bound to the confidence coefficient, and that it estimates the confidence coefficient quite well in the case of $N = 199$. In general, for n/N small, the continuous C.D.F. would yield a reasonable approximation for the confidence coefficients, and has the advantage in a slight ease of tabulation, having only to tabulate for k, r, n, and t.

Table 1. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ N = 39, n = 10

k	t	r = 1	2	3	4	5	6	7	8	9	10
1	10	.0158	.2183	.5497	.8187	.9364	.9646	.9683	.9685	.9685	.9685
	20	.0001	.0052	.0390	.1572	.3935	.6771	.8839	.9742	.9968	.9997
2	10		.0608	.3922	.6612	.7789	.8071	.8107	.8110	.8110	.8110
	20		.0023	.0361	.1543	.3906	.6742	.8810	.9713	.9939	.9968
3	10			.0884	.3574	.4751	.5033	.5069	.5071	.5072	.5072
	20			.0136	.1317	.3680	.6516	.8584	.9487	.9713	.9742
4	10				.0628	.1804	.2087	.2123	.2125	.2125	.2125
	20				.0413	.2777	.5612	.7680	.8584	.8810	.8839
5	10					.0235	.0518	.0554	.0556	.0556	.0556
	20					.0709	.3545	.5612	.6516	.6742	.6771
6	10						.0047	.0083	.0086	.0086	.0086
	20						.0709	.2777	.3680	.3906	.3935
7	10							.0005	.0007	.0007	.0007
	20							.0414	.1317	.1543	.1572
8	10								.0000	.0000	.0000
	20								.0136	.0361	.0390
9	10									.0000	.0000
	20									.0023	.0052

Table 2. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ N = 99, n = 10

k	t	r: 1	2	3	4	5	6	7	8	9	10
1	25	.0071	.2006	.4995	.7573	.8947	.9417	.9522	.9537	.9539	.9539
	50	.0001	.0080	.0485	.1667	.3828	.6421	.8488	.9569	.9923	.9989
2	25		.0232	.3221	.5800	.7173	.7644	.7749	.7764	.7765	.7765
	50		.0014	.0419	.1601	.3762	.6355	.8422	.9503	.9858	.9924
3	25			.0319	.2897	.4271	.4742	.4847	.4862	.4863	.4863
	50			.0065	.1246	.3407	.6001	.8068	.9148	.9503	.9569
4	25				.0241	.1614	.2085	.2190	.2205	.2206	.2206
	50				.0165	.2326	.4920	.6987	.8068	.8422	.8488
5	25					.0110	.0581	.0686	.0701	.0702	.0702
	50					.0259	.2853	.4920	.6001	.6355	.6421
6	25						.0031	.0136	.0151	.0153	.0153
	50						.0259	.2326	.3407	.3762	.3828
7	25							.0006	.0021	.0022	.0022
	50							.0165	.1246	.1601	.1667
8	25								.0000	.0002	.0002
	50								.0065	.0419	.0485
9	25									.0000	.0000
	50									.0014	.0080

Table 3. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ N = 199, n = 10

		r									
k	t	1	2	3	4	5	6	7	8	9	10
1	50	.0037	.1942	.4841	.7381	.8802	.9331	.9463	.9486	.9488	.9488
	100	.0000	.0089	.0512	.1690	.3793	.6318	.8378	.9507	.9904	.9985
2	50		.0114	.3013	.5554	.6974	.7503	.7636	.7658	.7660	.7660
	100		.0008	.0431	.1608	.3712	.6237	.8297	.9426	.9823	.9904
3	50			.0155	.2695	.4115	.4644	.4777	.4799	.4802	.4802
	100			.0034	.1211	.3315	.5840	.7900	.9028	.9426	.9507
4	50				.0119	.1539	.2068	.2201	.2223	.2225	.2225
	100				.0082	.2186	.4711	.6771	.7900	.8297	.8378
5	50					.0057	.0586	.0719	.0741	.0743	.0743
	100					.0126	.2651	.4711	.5840	.6237	.6318
6	50						.0018	.0150	.0173	.0175	.0175
	100						.0126	.2186	.3315	.3712	.3793
7	50							.0004	.0026	.0028	.0028
	100							.0082	.1211	.1609	.1690
8	50								.0000	.0003	.0003
	100								.0034	.0431	.0512
9	50									.0000	.0000
	100									.0008	.0089

Table 4. $P\{y_{(k)} \leq \xi_p < y_{(r)}\}$ Continuous C.D.F., n = 10

k	t(%)	r = 1	2	3	4	5	6	7	8	9	10
1	25		.1877	.4693	.7196	.8656	.9240	.9402	.9433	.9437	.9437
	50		.0098	.0537	.1709	.3760	.6221	.8271	.9443	.9883	.9980
2	25			.2816	.5318	.6778	.7362	.7525	.7556	.7559	.7560
	50			.0439	.1611	.3662	.6123	.8174	.9346	.9785	.9883
3	25				.2503	.3963	.4547	.4709	.4740	.4744	.4744
	50				.1172	.3223	.5684	.7734	.8906	.9346	.9443
4	25					.1460	.2044	.2206	.2237	.2241	.2241
	50					.2051	.4512	.6562	.7734	.8174	.8271
5	25						.0584	.0746	.0777	.0781	.0781
	50						.2461	.4512	.5684	.6123	.6221
6	25							.0162	.0193	.0197	.0197
	50							.2051	.3223	.3662	.3760
7	25								.0031	.0035	.0035
	50								.1172	.1611	.1709
8	25									.0004	.0004
	50									.0439	.0537
9	25										.0000
	50										.0098

I. Best Confidence Intervals

Up to this point we have been interested in the confidence coefficient $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$, where k and r have been predetermined. We may look at the problem from a different standpoint: find those subscripts $\{k,r\}$ which satisfy

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \geq 1 - \alpha \tag{2.29}$$

where $(1 - \alpha)$ has been fixed. This is Nair's [1940, pp. 551-558] approach which we discussed earlier. This leads to the difficulty that there may be many pairs (k,r) satisfying (2.29), and the problem then becomes one of picking the "best" of these pairs.

We will adopt the following criterion for the "best" confidence interval.

Definition: The best confidence interval of the form $[y_{(k)}, y_{(r)}]$ of level at least $(1 - \alpha)$ is that interval which satisfies (2.29) and has shortest expected length. If there are several intervals which have equal shortest expected length, choose the interval with largest confidence coefficient.

Since, in every case for $k < r$,

$$E(y_{(k)}) \leq E(y_{(r)}) \ , \tag{2.30}$$

we have the result that, if there are two confidence intervals, I_1 and I_2, which satisfy (2.29), and if $I_1 \subset I_2$, then the expected length of I_1 is less than the expected length of I_2. By our definition of best confidence interval, we can eliminate I_2 from further consideration. In the remainder of our work in this section, we will assume such "initial eliminations" have been performed.

Pratt [1961, p. 549] suggests another "natural measure" in place of expected length for testing the desirability of a confidence interval procedure. He suggests "a natural measure of the extent to which the confidence interval procedure includes a particular false value is the probability of including that particular value. To "average" this over all the values, one might simply integrate it over all the false values. This gives an apparently different measure of the "average extent" of the false values included". However, Pratt goes on to prove that the two measures are equal. Hence we will concern ourselves only with expected length.

In order to find the best confidence interval, we must find the expected lengths of the confidence intervals. To do this, we must make an assumption about the Y-values in our population.

Let $F(x;\theta)$ be a continuous cumulative distribution function. We can consider Π_N to be a random sample of size N drawn from an infinite "super-population" with distribution function $F(x;\theta)$. (Wilks [1962, p. 195]). Note that Π_N will have elements with distinct Y-values with probability one. This technique of using a "super-population" is suggested in Cochran [1963, pp. 214-216] in deriving results in systematic sampling. As a result of the assumption, we cannot prove results which apply to any single finite population--that is, to any specific values $Y_1, Y_2, \ldots, Y_N$,

--but our results apply to the average of all finite populations which can be drawn from the infinite population.

Using this assumption about the "super-population", we are able to bypass Π_N when working with $E(y_{(r)} - y_{(k)})$ since the distribution of $\{y_{(1)}, \ldots, y_{(n)}\}$, the order statistics of a random sample of size n drawn from a population with C.D.F. $F(x;\theta)$, is the same as the distribution of $\{z_{(1)}, \ldots, z_{(n)}\}$, the order statistics of a simple random sample of size n drawn from Π_N, where Π_N is a simple random sample of size N drawn from a population with C.D.F. $F(x;\theta)$.

As an example of how a change in the "super-population" assumption may change the interval to be chosen, consider a 70% confidence interval for $Y_{(50)}$ when $N = 199$ and $n = 10$. We see from Table 3 that $[y_{(1)}, y_{(4)}]$ and $[y_{(2)}, y_{(6)}]$ both yield intervals with confidence coefficient $>$ 70%. We then have the following (Hastings, et al., [1947, p. 417]):

Super-population	$E(y_{(4)} - y_{(1)})$	$E(y_{(6)} - y_{(2)})$
Uniform $(-\sqrt{3}, \sqrt{3})$	.945	1.260
Normal (0,1)	1.163	1.124

Hence, under the uniform assumption, we would use $[y_{(1)}, y_{(4)}]$, and under the normal assumption, $[y_{(2)}, y_{(6)}]$.

For a discussion of the expected values of order statistics, reference can be made to David [1970].

J. Joint Confidence Intervals

In this section we consider the construction and confidence coefficient for a joint confidence region for $Y_{(t)}$ and $Y_{(t')}$, $(t < t')$ of the form

$$([y_{(k)}, y_{(r)}], [y_{(k')}, y_{(r')}]), \quad k < k', \; r < r', \; k < r, \; k' < r' .$$

In Figure 1 we represent this region graphically.

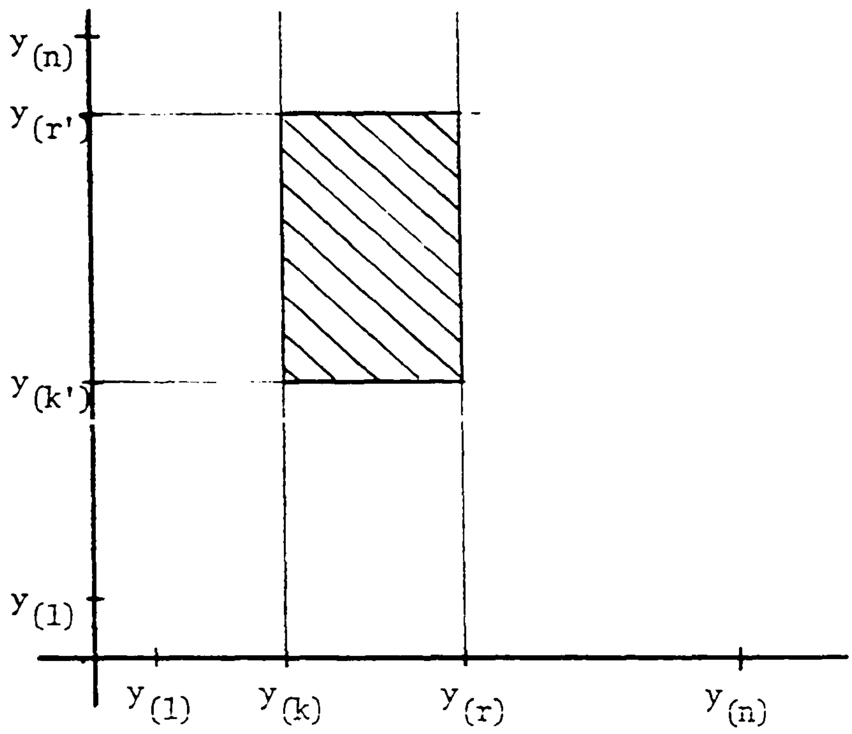

Figure 1. Joint confidence region

The confidence coefficient for this region can be calculated as follows:

$$P\{y_{(k)} \le Y_{(t)} \le y_{(r)} \cap y_{(k')} \le Y_{(t')} \le y_{(r')}\}$$

$$= P\{y_{(k)} \le Y_{(t)} \cap y_{(k')} \le Y_{(t')}\}$$

$$- P\{y_{(r)} < Y_{(t)} \cap y_{(r')} < Y_{(t')}\}$$

$$= P\{y_{(k)} \le Y_{(t)} \cap y_{(k')} \le Y_{(t')}\} - P\{y_{(r)} < Y_{(t)}\}$$

$$- P\{y_{(r')} < Y_{(t')}\} + P\{y_{(r)} < Y_{(t)} \cap y_{(r')} < Y_{(t')}\}$$

$$= P\{A\} - P\{B\} - P\{C\} + P\{D\} \ . \tag{2.31}$$

We first turn to the calculation of $P\{A\}$. Let $(k + i)$ be the exact number of observations in the sample with values less than or equal to $Y_{(t)}$, and let $(k' + j)$ be the exact number of observations in the sample less than or equal to $Y_{(t')}$. We then have

$$P\{A\} = \sum_{j=\max\left\{\begin{matrix}0\\ t'+n-N-k'\end{matrix}\right\}}^{\min\left\{\begin{matrix}t'-k'\\ n-k'\end{matrix}\right\}} \ \sum_{i=\max\left\{\begin{matrix}0\\ k'-k-(t'-t)+j\end{matrix}\right\}}^{\min\left\{\begin{matrix}t-k\\ k'-k+j\end{matrix}\right\}} \binom{t}{k+i}\binom{t'-t}{(k'+j)-(k+i)} \times \binom{N-t'}{n-(k'+j)} \Bigg/ \binom{N}{n} \tag{2.32}$$

$$P\{B\} = \sum_{i=r}^{\min\left\{\begin{matrix}n\\ t-1\end{matrix}\right\}} \binom{t-1}{i}\binom{N-t+1}{n-i} \Bigg/ \binom{N}{n} \tag{2.33}$$

$$P\{C\} = \sum_{i=r'}^{\min\left\{\begin{smallmatrix} n \\ t'-1 \end{smallmatrix}\right\}} \binom{t'-1}{i}\binom{N-t'+1}{n-i} \Big/ \binom{N}{n} \tag{2.34}$$

$$P\{D\} = \sum_{j=0}^{t'-k'-1} \sum_{i=0}^{t-k-1} \binom{t-1}{r+i}\binom{t'-t}{(r'+j)-(r+i)}\binom{N-t'+1}{n-(r'+j)} \Big/ \binom{N}{n} . \tag{2.35}$$

Clearly, the confidence region given above can be used to form confidence intervals for parameters such as $Y_{(t')} - Y_{(t)}$. Further, if $t' - t$ is not too large, one may prefer to have an "outer confidence interval" for $Y_{(t)}$ and $Y_{(t')}$. The confidence coefficient associated with the "outer confidence interval" given by $[y_{(k)} \leq Y_{(t)} < Y_{(t')} \leq y_{(r)}]$ may be determined by using the same approach employed in the beginning of this chapter.

III. CONFIDENCE INTERVALS WITH TWO STRATA

In this chapter we investigate confidence intervals for the t-th ordered value in a finite population, when the population has been stratified into two strata. Three distinct methods are proposed. The first method, called the "combined method" takes samples from both strata, combines and orders the sample values, and then uses two of the combined sample values for the endpoints of the confidence interval. The second method, called the "C.D.F. method", employs the empirical cumulative distribution function to find the endpoints of the desired interval. The third method, called the "separate method", uses one value from each of the strata samples to form the interval. Exact confidence coefficient formulas are derived for each method. Comparisons of the three techniques are given, both by theoretical work and Monte Carlo studies. Brief tables are also given for the Combined and Separate methods.

A. Definitions and Notations

Let Π_N be a population of N elements whose elements have distinct Y-values associated with them. These values can be simply-ordered as

$$Y_{(1)} < Y_{(2)} < \cdots < Y_{(N)} . \tag{3.1}$$

Assume that Π_N has been divided into two strata, with Stratum I containing N_1 elements with values

$$Y_{1(1)} < Y_{1(2)} < \cdots < Y_{1(N_1)} \tag{3.2}$$

and Stratum II containing N_2 elements

$$Y_{2(1)} < Y_{2(2)} < \cdots < Y_{2(N_2)} . \tag{3.3}$$

Of course, $N_1 + N_2 = N$.

A simple random sample of size n_1 is drawn from Stratum I. The ordered observations in the sample are denoted by

$$y_{1(1)} < y_{1(2)} < \cdots < y_{1(n_1)} . \tag{3.4}$$

Similarly, a simple random sample of size n_2 is drawn from Stratum II, yielding

$$y_{2(1)} < y_{2(2)} < \cdots < y_{2(n_2)} . \tag{3.5}$$

Combining and ordering the two samples yields the combined sample

$$y_{(1)} < y_{(2)} < \cdots < y_{(n)} , \tag{3.6}$$

where $n = n_1 + n_2$.

B. The Combined Method

Turning to the first of our three methods, we are interested in the confidence coefficients of confidence intervals for $Y_{(t)}$ of the form $[y_{(k)}, y_{(r)}]$.

We note here that if ties are permitted among our population Y-values, the proof given in Chapter II that the confidence coefficient is at least as great as when our population has all distinct Y-values holds without change for the stratified situation.

1. Derivation of the general formula

Since we are considering distinct population values, we have, as in Chapter II,

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} = P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} \leq Y_{(t-1)}\} \ . \qquad (3.7)$$

Let A_i be the event "exactly i observations in the combined sample have values less than or equal to $Y_{(t)}$";

let A^{ℓ}_{ij} be the event "exactly i observations in the combined sample have values less than or equal to $Y_{(t)}$ and $Y_{(t)} = Y_{\ell(j)}$, $\ell = 1,2$";

let $A^{\ell}_{i|j}$ be the event "exactly i observations in the combined sample have values less than or equal to $Y_{(t)}$, given that $Y_{(t)} = Y_{\ell(j)}$, $\ell = 1,2$";

and let B^{ℓ}_{j} be the event "$Y_{(t)} = Y_{\ell(j)}$, $\ell = 1,2$".

We then have

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{i=\max[k,t-(N-n)]}^{\min[t,n]} P\{A_i\}$$

$$= \sum_{i} \sum_{j=\max[1,t-N_2]}^{\min[t,N_1]} P\{A^1_{ij}\} + \sum_{i} \sum_{j^*=\max[1,t-N_1]}^{\min[t,N_2]} P\{A^2_{ij^*}\}$$

$$= \sum_{i} \sum_{j} P\{A^1_{i|j}\}P\{B^1_j\} + \sum_{i} \sum_{j^*} P\{A^2_{i|j^*}\}P\{B^2_{j^*}\} \ . \qquad (3.8)$$

We first turn to the evaluation of $P\{A^1_{i|j}\}$. Let m represent the number of elements in the sample from the first stratum with values less than or equal to $Y_{(t)}$; that is, values less than or equal to $Y_{1(j)}$. For a fixed m, the number of cases favorable to $A^1_{i|j}$ can be obtained directly:

$$\binom{j}{m}\binom{N_1 - j}{n_1 - m}\binom{t - j}{i - m}\binom{N_2 - (t - j)}{n_2 - (i - m)} , \tag{3.9}$$

where the range on m is given by

$$\max\begin{Bmatrix} 0 \\ i - n_2 \\ j + n_1 - N_1 \\ i + j - t \end{Bmatrix} \leq m \leq \min\begin{Bmatrix} i \\ j \\ n_1 \\ N_2 - (t-j) - n_2 + i \end{Bmatrix} . \tag{3.10}$$

Adding (3.9) over all possible values of m, as given in (3.10), and dividing by $\binom{N_1}{n_1}\binom{N_2}{n_2}$ yields

$$P\{A^1_{i|j}\} = \frac{\sum_m \binom{j}{m}\binom{N_1-j}{n_1-m}\binom{t-j}{i-m}\binom{N_2-(t-j)}{n_2-(i-m)}}{\binom{N_1}{n_1}\binom{N_2}{n_2}} . \tag{3.11}$$

Similarly,

$$P\{A^2_{i|j^*}\} = \frac{\sum_{m^*} \binom{j^*}{m^*}\binom{N_2-j^*}{n_2-m^*}\binom{t-j^*}{i-m^*}\binom{N_1-(t-j^*)}{n_1-(i-m^*)}}{\binom{N_1}{n_1}\binom{N_2}{n_2}} \tag{3.12}$$

where the range on m^* is given by

$$\max\left\{\begin{matrix} 0 \\ i - n_1 \\ j^* + n_2 - N_2 \\ i + j^* - t \end{matrix}\right\} \le m^* \le \min\left\{\begin{matrix} i \\ j^* \\ n_2 \\ N_1 - (t - j^*) - n_1 + i \end{matrix}\right\} . \quad (3.13)$$

At this point, it remains to arrive at expressions for $P\{B_j^1\}$ and $P\{B_{j^*}^2\}$. In order to do this, we must make an assumption concerning the stratification. To illustrate the possibilities, two different assumptions are considered in this chapter. The first we call "random stratification" and the second "$Y_{1(s)} < Y_{2(u)}$". It is intended that each of these assumptions represents a type of stratification found in practical applications. However, for a particular finite population, other postulates may be more appropriate and these, of course, should be used.

2. Random stratification

Following the notation of Section A, we will say that the stratification is random if Stratum I is equally likely to be any one of the $\binom{N}{N_1}$ possible subsets of size N_1 of Π_N.

This assumption of "random stratification" may be appropriate where strata are formed primarily for administrative convenience. For example, administrative districts may be used as separate strata, even though there is little difference among the strata in the distribution of the Y-values.

In order to arrive at an expression for $P\{B_j^1\}$ we note that the event B_j^1 can be thought of as putting into Stratum I $(j - 1)$ units out of the first $(t - 1)$ population units, then putting the t-th population unit into the j-th position in Stratum I, and finally putting $(N_1 - j)$ units into Stratum I out of the last $(N - t)$ population units. This can be done in $\binom{t-1}{j-1}\binom{1}{1}\binom{N-t}{N_1-j}$ ways.

Hence, under random stratification, we have

$$P\{B_j^1\} = \begin{cases} \binom{t-1}{j-1}\binom{N-t}{N_1-j} \Big/ \binom{N}{N_1} & \text{if } \max\begin{Bmatrix}1 \\ t-N_2\end{Bmatrix} \le j \le \min\begin{Bmatrix}t \\ N_1\end{Bmatrix} \\ 0 & \text{otherwise} . \end{cases} \tag{3.14}$$

Similarly,

$$P\{B_{j^*}^2\} = \begin{cases} \binom{t-1}{j^*-1}\binom{N-t}{N_2-j^*} \Big/ \binom{N}{N_2} & \text{if } \max\begin{Bmatrix}1 \\ t-N_1\end{Bmatrix} \le j^* \le \min\begin{Bmatrix}t \\ N_2\end{Bmatrix} \\ 0 & \text{otherwise} . \end{cases} \tag{3.15}$$

Combining (3.8), (3.11), (3.12), (3.14), and (3.15) yields

$$\begin{aligned} P\{y_{(k)} \le Y_{(t)}\} = {} & \left(\frac{N!}{n_1!n_2!(N_1-n_1)!(N_2-n_2)!}\right)^{-1} \\ & \times \sum_i \left[\sum_j \binom{t-1}{j-1}\binom{N-t}{N_1-j} \sum_m \binom{j}{m}\binom{t-j}{i-m}\binom{N_1-j}{n_1-m}\binom{N_2-(t-j)}{n_2-(i-m)} \right. \\ & \left. + \sum_{j^*} \binom{t-1}{j^*-1}\binom{N-t}{N_2-j^*} \sum_{m^*} \binom{j^*}{m^*}\binom{t-j^*}{i-m^*}\binom{N_2-j^*}{n_2-m^*}\binom{N_1-(t-j^*)}{n_1-(i-m^*)}\right] \end{aligned} \tag{3.16}$$

where the limits on the summation of i and j are given in (3.8) on m in (3.10), and m^* in (3.13).

It can be shown algebraically that (3.16) reduces to (2.7). That is, the confidence coefficient associated with $[y_{(k)}, y_{(r)}]$ is the same if either (1) simple random sampling or (2) stratified simple random sampling with "random" stratification is assumed. In the following theorem we prove this result holds for any number of strata. It may be noted that this equivalence does not necessarily hold for the other types of confidence intervals discussed later in this chapter.

Theorem 3.1: Let Π_N be a population of N distinct units $\{u_1, u_2, \ldots, u_N\}$. Assume Π_N has been divided into L strata of sizes N_i, $i = 1, \ldots, L$, in such a way that any one of the $\binom{N}{N_1, N_2, \ldots, N_L}$ possible stratifications is equally likely to occur. From each stratum a simple random sample of size n_i, $i = 1, \ldots, L$, is drawn. Then if $S = [(u_{i_1}, u_{i_2}, \ldots, u_{i_n}),\ i_j \in \{1, \ldots, N\},\ i_j \neq i_k,\ j \neq k]$ is any fixed sample, the probability of obtaining it is $\binom{N}{n}^{-1}$.

Proof: P{obtaining S}

$$= P\{n_i \text{ elements of } S \text{ are in Stratum } i,\ i=1,\ldots, L\}$$

$$\times P\{\text{drawing those } n_i \text{ elements, } i=1,\ldots, L\}$$

$$= \frac{\binom{n}{n_1, n_2, \ldots, n_L}\binom{N-n}{N_1-n_1, N_2-n_2, \ldots, N_L-n_L}}{\binom{N}{N_1, N_2, \ldots, N_L}} \frac{1}{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_L}{n_L}}$$

$$= \binom{N}{n}^{-1} . \tag{3.17}$$

In the previous chapter we showed that systematic sampling with the elements in "random" order was equivalent to simple random sampling. Hence, in this chapter, if we apply systematic sampling to each strata, under the assumption of "random" stratification, it would be equivalent to simple random sampling from each strata, and we conclude that this would be equivalent to simple random sampling without stratification.

3. Ordered stratification

In many surveys the stratification variable, X, is closely related to the variable under study, Y. For example, X may denote the value of Y at some previous time. Then, the usual stratification consists of placing those units with the smallest values of X in Stratum I, those with the next smallest values of X in Stratum II, etc. Knowledge about the relation between Y and X, and the method of stratification may enable the investigator to assert that, for example, $Y_{1(s)} < Y_{2(u)}$. For instance, in some situations it may be reasonable to assume that the median among the variate values in Stratum I is less than the median among the variate values in Stratum II. Note that although in the ensuing analysis the only restriction made is $Y_{1(s)} < Y_{2(u)}$, other restrictions of the same type may be added.

4. $Y_{1(s)} < Y_{2(u)} < Y_{1(s+1)}$

For simplicity, it is assumed at first that $Y_{1(s)} < Y_{2(u)} < Y_{1(s+1)}$ where $1 \leq s \leq N_1$, $1 \leq u \leq N_2$, and $Y_{1(N_1+1)} = +\infty$. Note that this specification includes the possibility of asserting that $Y_{1(N_1)} < Y_{2(1)}$; such an assertion may be a reasonable approximation where, for example, there are only a small number of very "large" units and these are all included in a single stratum.

In order to compute $P\{y_{(k)} \leq Y_{(t)}\}$ under these specifications, we note that equations (3.8), (3.11), and (3.12) still hold. It therefore remains to compute $P\{B_j^1\}$ and $P\{B_{j*}^2\}$. It may be of assistance to consider Figure 2 which represents the population in both the combined and stratified forms under our specification.

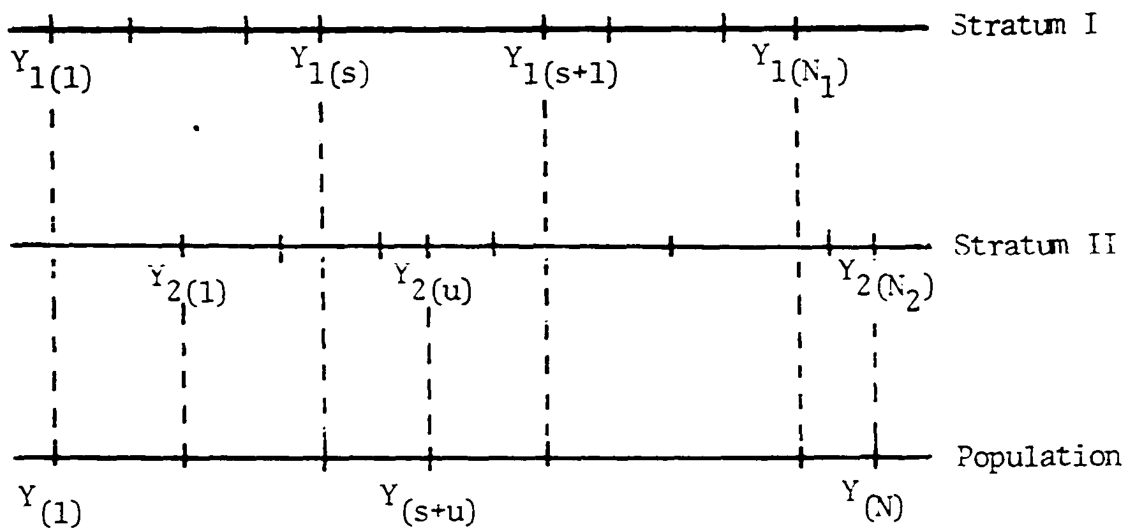

Figure 2. Stratification under $Y_{1(s)} < Y_{2(u)} < Y_{1(s+1)}$

We break up our analysis into three cases:

Case I: $t < s+u$

Case II: $t = s+u$

Case III: $t > s+u$,

and arrive at expressions for $P\{B_j^1\}$ and $P\{B_{j*}^2\}$ by counting techniques.

The total number of possibilities for distributing the ordered units among the two strata, given $Y_{1(s)} < Y_{2(u)} < Y_{1(s+1)}$, is

$$\binom{s+(u-1)}{s}\binom{1}{0}\binom{N-(s+u)}{N_1-s} \tag{3.18}$$

Case I: $t < s + u$. There are $\binom{t-1}{j-1}\binom{s+(u-1)-t}{s-j}\binom{N-(s+u)}{N_1-s}$ ways to stratify favorable to $Y_{(t)} = Y_{1(j)}$. Therefore

$$P\{B_j^1\} = \begin{cases} \dfrac{\binom{t-1}{j-1}\binom{s+(u-1)-t}{s-j}}{\binom{s+(u-1)}{s}} & \text{if } \max\begin{Bmatrix} 1 \\ t-(u-1)\end{Bmatrix} \leq j \leq \min\begin{Bmatrix} s \\ t \end{Bmatrix} \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

Also, there are $\binom{t-1}{j^*-1}\binom{s+(u-1)-t}{u-1-j^*}\binom{N-(s+u)}{N_2-u}$ ways to stratify favorable to $Y_{(t)} = Y_{2(j^*)}$. Therefore

$$P\{B_{j*}^2\} = \begin{cases} \dfrac{\binom{t-1}{j^*-1}\binom{s+(u-1)-t}{(u-1)-j^*}}{\binom{s+(u-1)}{s}} & \text{if } \max\begin{Bmatrix} 1 \\ t-s\end{Bmatrix} \leq j^* \leq \min\begin{Bmatrix} t \\ u-1 \end{Bmatrix} \\ 0 & \text{otherwise} \end{cases} \tag{3.20}$$

Case II: $t = s + u$. In this case, we necessarily have $Y_{(t)} = Y_{(s+u)} = Y_{2(u)}$. Hence

$$P\{B_j^1\} = 0 \quad \text{for all } j \tag{3.21}$$

and

$$P\{B_{j*}^2\} = \begin{cases} 1 & \text{if } j* = u \\ 0 & \text{otherwise} \end{cases} . \tag{3.22}$$

Case III. $t > s + u$. There are $\binom{s+(u-1)}{s}\binom{t-(s+u)-1}{j-s-1}\binom{N-t}{N_1-j}$ ways to stratify favorable to $Y_{(t)} = Y_{1(j)}$. Hence

$$P\{B_j^1\} = \begin{cases} \dfrac{\binom{t-(s+u)-1}{j-s-1}\binom{N-t}{N_1-j}}{\binom{N-(s+u)}{N_1-s}} & \text{if } \max\begin{Bmatrix} s+1 \\ t-N_2 \end{Bmatrix} \le j \le \min\begin{Bmatrix} N_1 \\ t-u \end{Bmatrix} \\ 0 & \text{otherwise} \end{cases} . \tag{3.23}$$

Finally, there are $\binom{s+(u-1)}{u-1}\binom{t-(s+u)-1}{j*-u-1}\binom{N-t}{N_2-j*}$ ways to stratify favorable to $Y_{(t)} = Y_{2(j*)}$. Therefore

$$P\{B_{j*}^2\} = \begin{cases} \dfrac{\binom{t-(s+u)-1}{j*-u-1}\binom{N-t}{N_2-j*}}{\binom{N-(s+u)}{N_2-u}} & \text{if } \max\begin{Bmatrix} u+1 \\ t-N_1 \end{Bmatrix} \le j* \le \min\begin{Bmatrix} N_2 \\ t-s \end{Bmatrix} \\ 0 & \text{otherwise} \end{cases} . \tag{3.24}$$

It is easily shown that, for each of the cases,

$$\sum_j P\{B_j^1\} + \sum_{j*} P\{B_{j*}^2\} = 1 \quad . \tag{3.25}$$

Combining (3.8), (3.11), (3.12), and the expressions (3.19) - (3.24), we have $P\{y_{(k)} \le Y_{(t)}\}$ for the following cases:

Case I. $t < s + u$.

$$\begin{aligned} P\{y_{(k)} \le Y_{(t)}\} = \Bigg[& \sum_{i=\max[k,t-(N-n)]}^{\min[t,n]} \; \sum_{j=\max[1,t-(u-1)]}^{\min[s,t]} \; \sum_m \binom{j}{m}\binom{N_1-j}{n_1-m} \\ & \times \binom{t-j}{i-m}\binom{N_2-(t-j)}{n_2-(i-m)}\binom{t-1}{j-1}\binom{s+(u-1)-t}{s-j} \\ & + \sum_i \; \sum_{j*=\max[1,t-s]}^{\min[(u-1),t]} \; \sum_{m*} \binom{j*}{m*}\binom{N_2-j*}{n_2-m*} \\ & \times \binom{t-j*}{i-m*}\binom{N_1-(t-j*)}{n_1-(i-m*)}\binom{t-1}{j*-1}\binom{s+(u-1)-t}{(u-1)-j*} \Bigg] \\ & \times \left[\binom{N_1}{n_1}\binom{N_2}{n_2}\binom{s+(u-1)}{s}\right]^{-1} \end{aligned} \tag{3.26}$$

where the limits of the summation of m and $m*$ are given by (3.10) and (3.13) respectively.

Case II. $t = s + u$.

$$P\{y_{(k)} \le Y_{(t)}\} = \sum_i \sum_{m*} \frac{\binom{u}{m*}\binom{N_2-u}{n_2-m*}\binom{s}{i-m*}\binom{N_1-s}{n_1-(i-m*)}}{\binom{N_1}{n_1}\binom{N_2}{n_2}} \tag{3.27}$$

where the range of the summation of m^* is

$$\max\begin{Bmatrix} 0 \\ i - n_1 \\ n_2 + u - N_2 \\ i - s \end{Bmatrix} \le m^* \le \min\begin{Bmatrix} i \\ u \\ n_2 \\ N_1 - s - n_1 - i \end{Bmatrix} . \tag{3.28}$$

Case III. $t > s + u$.

$$P\{y_{(k)} \le Y_{(t)}\} = \Bigg[\sum_i \sum_{j=\max[s+1,t-N_2]}^{\min[N_1,t-u]} \sum_m \binom{j}{m}\binom{N_1-j}{n_1-m}\binom{t-j}{i-m}\binom{N_2-(t-j)}{n_2-(i-m)}$$

$$\times \binom{t-(s+u)-1}{j-s-1}\binom{N-t}{N_1-j}$$

$$+ \sum_i \sum_{j^*=\max[u+1,t-N_1]}^{\min[N_2,t-s]} \sum_{m^*} \binom{j^*}{m^*}\binom{N_2-j^*}{n_2-m^*}\binom{t-j^*}{i-m^*}$$

$$\times \binom{N_1-(t-j^*)}{n_1-(i-m^*)}\binom{t-(s+u)-1}{j^*-u-1}\binom{N-t}{N_2-j^*}\Bigg]$$

$$\times \left[\binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N-(s+u)}{N_1-s}\right]^{-1} \tag{3.29}$$

where again the ranges on summation of m and m^* are given by (3.10) and (3.13).

5. $\underline{Y_{1(s)} < Y_{2(u)}}$

It would appear that the assumption $Y_{1(s)} < Y_{2(u)} < Y_{1(s+1)}$ is rather unrealistic and unduly restrictive. Hence we turn to the more realistic and general assumption that $Y_{1(s)} < Y_{2(u)}$, where $1 \le s \le N_1$ and $1 \le u \le N_2$.

As in the previous section, equations (3.8), (3.11), and (3.12) are still valid. It remains to compute $P\{B_j^1\}$ and $P\{B_{j*}^2\}$ under our new assumptions.

Let d be an integer in the range $0 \le d \le N_1 - s$ such that $Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}$. (If $s + d + 1 = N_1 + 1$, let $Y_{1(s+d+1)} = +\infty$). We then have

$$P\{B_j^1\} = \sum_{d=0}^{N_1-s} P\{B_j^1 \cap Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}\}$$

$$= \sum_{d=0}^{N_1-s} P\{B_j^1 \mid Y_{1(s+d+1)} < Y_{2(u)} < Y_{1(s+d+1)}\}$$

$$\times \; P\{Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}\} \; . \tag{3.30}$$

Letting

$$P\{C_{j|d}^1\} = P\{B_j^1 \mid Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}\} \tag{3.31}$$

and

$$P\{D_d\} = P\{Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}\} \tag{3.32}$$

we then have

$$P\{B_j^1\} = \sum_d P\{C_{j|d}^1\} P\{D_d\} . \tag{3.33}$$

All of these probabilities are, of course, conditional on $Y_{1(s)} < Y_{2(u)}$.

Evaluation of $P\{D_d\}$. Consider N_1, N_2, s, u, and d fixed. There are then

$$\binom{s+d+(u-1)}{s+d}\binom{N-(s+d+u)}{N_1-(s+d)} \tag{3.34}$$

ways to stratify so that $Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}$. It follows that the total number of possible arrangements such that $Y_{1(s)} < Y_{2(u)}$ is

$$\sum_{d=0}^{N_1-s} \binom{s+d+(u-1)}{s+d}\binom{N-(s+d+u)}{N_1-(s+d)} . \tag{3.35}$$

Considering each of these arrangements (stratifications) equally likely, we then have

$$P\{D_d\} = \frac{\binom{s+d+(u-1)}{s+d}\binom{N-(s+d+u)}{N_1-(s+d)}}{\sum_{d=0}^{N_1-s} \binom{s+d+(u-1)}{s+d}\binom{N-(s+d+u)}{N_1-(s+d)}} . \tag{3.36}$$

Evaluation of $P\{C_{j|d}^1\}$. As in the previous section, we break our work up into three cases.

Case I: $t < s + d + u$. From the previous section,

$$P\{C^1_{j|d}\} = \begin{cases} \dfrac{\binom{t-1}{j-1}\binom{s+d+(u-1)-t}{s+d-j}}{\binom{s+d+(u-1)}{s+d}} & \text{if } \max\begin{Bmatrix} 1 \\ t-(u-1) \end{Bmatrix} \le j \le \min\begin{Bmatrix} t \\ s+d \end{Bmatrix} \\ 0 & \text{otherwise} . \end{cases} \tag{3.37}$$

Case II: $t = s + d + u$. As before

$$P\{C^1_{j|d}\} = 0 \quad \text{for all } j . \tag{3.38}$$

Case III: $t > s + d + u$.

$$P\{C^1_{j|d}\} = \begin{cases} \dfrac{\binom{t-(s+d+u)-1}{j-(s+d)-1}\binom{N-t}{N_1-j}}{\binom{N-(s+d+u)}{N_1-(s+d)}} & \text{if } \max\begin{Bmatrix} s+d+1 \\ t-N_2 \end{Bmatrix} \le j \le \min\begin{Bmatrix} t-u \\ N_1 \end{Bmatrix} \\ 0 & \text{otherwise} . \end{cases} \tag{3.39}$$

Evaluation of $P\{B^1_j\}$. We define an indicator function α by

$$\alpha^t_z = \begin{cases} 1 & \text{if } t \le z \\ 0 & \text{otherwise} . \end{cases} \tag{3.40}$$

Putting (3.36) - (3.40) into (3.33), we have

$$
P\{B_j^1\} = \begin{cases} \left\{ \binom{t-1}{j-1} \sum_d \left[\binom{s+d+(u-1)-t}{s+d-j} \binom{N-(s+d+(u-1))-1}{N_1-(s+d)} \alpha^t_{s+d+(u-1)} \right] \right. \\ \quad \left. + \binom{N-t}{N_1-j} \sum_d \left[\binom{s+d+(u-1)}{s+d} \binom{t-(s+d+(u-1))-2}{j-(s+d)-1} \left(1-\alpha^t_{s+d+u}\right) \right] \right\} \\ \quad \times \left[\sum_d \binom{s+d+(u-1)}{s+d} \binom{N-(s+d+u)}{N_1-(s+d)} \right]^{-1} \\ \qquad \text{if } \min\left\{ \begin{matrix} \max\left\{ \begin{matrix} 1 \\ 1-(u-1) \end{matrix} \right\} \\ \max\left\{ \begin{matrix} s+1 \\ t-N_2 \end{matrix} \right\} \end{matrix} \right\} \le j \le \max\left\{ \begin{matrix} \min\left\{ \begin{matrix} N_1 \\ t \end{matrix} \right\} \\ \min\left\{ \begin{matrix} t-u \\ N_1 \end{matrix} \right\} \end{matrix} \right\}, \\ 0 \qquad \text{otherwise} . \end{cases} \tag{3.41}
$$

<u>Evaluation of $P\{B^2_{j*}\}$</u>. Turning now to Stratum II, we have, corresponding to (3.33),

$$
P\{B^2_{j*}\} = \sum_d P\{C^2_{j*|d}\} \, P\{D_d\} . \tag{3.42}
$$

$P\{D_d\}$ is as in equation (3.36). The evaluations of $P\{C^2_{j*|d}\}$ are very similar to equations (3.37) - (3.39). Putting these into (3.42), and recalling the definition of the Kronecker delta:

$$
\delta_{a,b} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases} \tag{3.43}
$$

we have

$$P\{B^2_{j*}\} = \begin{cases} \left\{ \binom{t-1}{j^*-1} \left[\sum_d \binom{s+d+(u-1)-t}{(u-1)-j^*} \binom{N-(s+d+(u-1))-1}{N_1-(s+d)} \alpha^t_{s+d+(u-1)} \right] \right. \\ \quad + \binom{t-1}{t-u} \binom{N-t}{N_1-(t-u)} \delta_{j^*,u} \\ \quad + \left. \binom{N-t}{N_2-j^*} \left[\sum_d \binom{s+d+(u-1)}{s+d} \binom{t-(s+d+(u-1))}{j^*-u-1} \left(1-\alpha^t_{s+d+u}\right) \right] \right\} \\ \quad \times \left[\sum_d \binom{s+d+(u-1)}{s+d} \binom{N-(s+d+u)}{N_1-(s+d)} \right]^{-1} \\ \qquad \text{for } \min\left\{ \begin{matrix} \max\left\{ \begin{matrix} 1 \\ t-N_1 \end{matrix} \right\} \\ \max\left\{ \begin{matrix} u+1 \\ t-N_1 \end{matrix} \right\} \end{matrix} \right\} \le j^* \le \max\left\{ \begin{matrix} \min\left\{ \begin{matrix} t \\ u-1 \end{matrix} \right\} \\ \min\left\{ \begin{matrix} N_2 \\ t-s \end{matrix} \right\} \end{matrix} \right\} \\ 0 \qquad \text{otherwise} . \end{cases} \tag{3.44}$$

Thus, under the assumption that $Y_{1(s)} < Y_{2(u)}$, one may determine $P\{y_{(k)} \le Y_{(t)}\}$ by using (3.8), (3.11), and (3.12) together with (3.41) and (3.44).

6. Tables for $P\{y_{(k)} \le Y_{(t)} \le y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$

A FORTRAN program was written for use on the UNIVAC 1108 computer at the University of Wisconsin Computer Center to evaluate $P\{y_{(k)} \le Y_{(t)} \le y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$. The variable parameters are N_1, N_2, n_1, n_2, t, s, and u. The results were printed in tabular form, giving the above probabilities for $1 \leq k \leq r \leq n$. Tables 5 through 9 present some typical results. In all cases presented, $N_1 = N_2 = 10$, and the values assumed for the (s,u) parameters are (1,1), (1,5), (1,10), (5,1), (5,5), (5,10), (10,1), (10,5), and (10,10). Tables 5 and 6 are for $n_1 = n_2 = 3$ (proportional allocation), and $t = 4$ and 8, respectively. Tables 7, 8, and 9 deal with non-proportional allocation. In Tables 7 and 8, $n_1 = 2$, $n_2 = 4$, and $t = 4$ and 8, respectively. Table 9 has $n_1 = 4$, $n_2 = 2$ and $t = 8$. For comparative purposes, in each table we have underlined the entry for $(s,u) = (1,10)$, which is essentially equivalent to "random" stratification.

Some interesting patterns emerge as we study the tables. For proportional allocation (Tables 5 and 6) the probability of coverage remains essentially constant for $(s,u) = (1,1)$, (1,5), (1,10), (5,5), (5,10), and (10,10) and this probability is equal to that of "random" stratification. In Table 5, for $t = 4$, we see a slight improvement in coverage probabilities in the other (s,u) entries for $(k,r) = (1,r)$, $r = 3, 4, 5, 6$. That is, as the strata become "more ordered", those confidence intervals which include $y_{(1)}$ have a higher probability of coverage. Also, for all other intervals, we have no gains, and some losses in probability over the probabilities for "random" stratification. In Table 6, for proportional allocation and $t = 8$, we see improvements in coverage probability, some substantial, for $(k,r) = (1,r)$, $(2,r)$, and $(3,r)$ as the strata become "more ordered".

In Table 7 (non-proportional allocation), we have no improvement in the probability of coverage for fixed (k,r) and every pair (s,u)

differing from "random" stratification. Hence, this would indicate that, for improvement, a different allocation of sample sizes should be used. In Table 8, we have gains for those intervals which include $y_{(1)}$. As the $[y_{(k)}, y_{(r)}]$ intervals "shift to the right", (e.g., from $[y_{(1)}, y_{(4)}]$ to $[y_{(2)}, y_{(5)}]$, the coverage probabilities drop. However, in Table 9, the probabilities improve as the intervals "shift to the right". This indicates that, if we seek improvement of coverage probabilities by using the "combined" method, and an assumption of the type "$Y_{1(s)} < Y_{2(u)}$", the sampling fractions and intervals used must be selected carefully.

7. Bounds and approximations for the combined method

a. McCarthy's conjecture. McCarthy [1965, pp. 772-783] has considered the type of procedure we have just discussed. He assumes an arbitrary number, L, of strata, a continuous C.D.F. in each stratum, and proportional sample size allocation. He proves that any pair of symmetric order statistics from the combined stratified random sample of size n provides a confidence interval for the population median, whose confidence coefficient is not less than the confidence coefficient associated with the interval determined by the corresponding order statistics in a random sample of n observations drawn from the entire population. He proves (by counterexample) that the result necessarily holds only when proportional allocation is employed, and also notes that his main result can easily be extended to any other quantile of the population. Our approach differs from that of McCarthy in that (1) we give the probability

Table 5. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$ -- Combined Method

$N_1 = N_2 = 10,\ n_1 = n_2 = 3,\ t = 4$

		r = 3			r = 4			r = 5			r = 6		
k	s	u= 1	5	10	1	5	10	1	5	10	1	5	10
1	1	.776	.775	.776	.793	.793	.793	.793	.793	.793	.793	.793	.793
	5	.825	.776	.776	.833	.793	.793	.833	.793	.793	.833	.793	.793
	10	.825	.784	.776	.833	.800	.793	.833	.800	.793	.833	.800	.793
2	1	.325	.325	.325	.343	.343	.343	.343	.343	.343	.343	.343	.343
	5	.325	.325	.325	.333	.343	.343	.333	.343	.343	.333	.343	.343
	10	.325	.325	.325	.333	.341	.343	.333	.341	.343	.333	.341	.343
3	1	.043	.044	.043	.061	.061	.061	.061	.061	.061	.061	.061	.061
	5	.025	.043	.043	.033	.061	.061	.033	.061	.061	.033	.061	.061
	10	.025	.041	.043	.033	.057	.061	.033	.057	.061	.033	.057	.061
4	1				.003	.003	.003	.003	.003	.003	.003	.003	.003
	5				.000	.003	.003	.000	.003	.003	.000	.003	.003
	10				.000	.002	.003	.000	.002	.003	.000	.002	.003

Table 6. $P\{y_{(k)} \le Y_{(t)} \le y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$ -- Combined Method

$N_1 = N_2 = 10,\ n_1 = n_2 = 3,\ t = 8$

		r = 3			r = 4			r = 5			r = 6		
k	s	u= 1	5	10	1	5	10	1	5	10	1	5	10
1	1	.640	.640	.640	.899	.898	.898	.969	.969	.969	.976	.976	.976
	5	.663	.640	.640	.930	.899	.898	.982	.969	.969	.985	.976	.976
	10	.708	.653	.640	1.000	.915	.898	1.000	.978	.969	1.000	.982	.976
2	1	.477	.477	.477	.735	.734	.735	.805	.805	.805	.813	.812	.813
	5	.514	.477	.477	.780	.735	.735	.833	.805	.805	.835	.812	.812
	10	.642	.500	.477	.933	.762	.735	.933	.824	.805	.933	.829	.812
3	1	.119	.119	.119	.377	.377	.377	.448	.448	.448	.455	.455	.455
	5	.132	.119	.119	.399	.377	.377	.451	.448	.448	.453	.455	.455
	10	.175	.124	.119	.467	.387	.377	.467	.449	.448	.467	.454	.455
4	1				.060	.060	.060	.130	.130	.130	.137	.137	.137
	5				.061	.060	.060	.114	.130	.130	.116	.137	.137
	10				.000	.055	.060	.000	.117	.130	.000	.122	.137

Table 7. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$ -- Combined Method

$N_1 = N_2 = 10,\ n_1 = 2,\ n_2 = 4,\ t = 4$

		r = 3			r = 4			r = 5			r = 6		
k	s	u= 1	5	10	1	5	10	1	5	10	1	5	10
1	1	.752	.774	.776	.764	.791	.793	.764	.791	.793	.764	.791	.793
	5	.667	.750	.776	.667	.764	.793	.667	.764	.793	.667	.764	.793
	10	.667	.725	.770	.667	.734	.786	.667	.734	.786	.667	.734	.786
2	1	.291	.322	.352	.302	.339	.342	.302	.339	.342	.302	.339	.342
	5	.133	.289	.325	.133	.302	.342	.133	.302	.342	.133	.302	.342
	10	.133	.250	.316	.133	.260	.333	.133	.260	.333	.133	.260	.333
3	1	.036	.043	.043	.047	.060	.061	.047	.060	.061	.047	.060	.061
	5	.000	.034	.043	.000	.047	.061	.000	.047	.061	.000	.047	.061
	10	.000	.024	.041	.000	.033	.058	.000	.033	.058	.000	.033	.058
4	1				.002	.003	.003	.002	.003	.003	.002	.003	.003
	5				.000	.002	.003	.000	.002	.003	.000	.002	.003
	10				.000	.001	.003	.000	.001	.003	.000	.001	.003

Table 8. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$ -- Combined Method

$N_1 = N_2 = 10$, $n_1 = 2$, $n_2 = 4$, $t = 8$

		r = 3			r = 4			r = 5			r = 6		
k	s u=	1	5	10	1	5	10	1	5	10	1	5	10
1	1	.663	.643	.640	.906	.899	.898	.967	.969	.969	.972	.976	.976
	5	.800	.686	.641	.946	.912	.899	.959	.963	.969	.959	.966	.976
	10	.978	.744	.652	.978	.929	.902	.978	.959	.967	.978	.961	.974
2	1	.486	.478	.477	.729	.734	.735	.790	.804	.805	.796	.811	.813
	5	.548	.487	.477	.694	.713	.735	.707	.764	.805	.707	.767	.812
	10	.622	.505	.480	.622	.690	.729	.622	.720	.795	.622	.722	.801
3	1	.120	.119	.119	.362	.376	.377	.424	.445	.448	.429	.452	.455
	5	.127	.108	.119	.273	.334	.377	.286	.385	.447	.286	.388	.454
	10	000	.095	.116	.000	.280	.366	.000	.310	.432	.000	.311	.438
4	1				.056	.059	.060	.117	.129	.130	.123	.136	.137
	5				.034	.045	.059	.046	.096	.130	.046	.099	.137
	10				.000	.028	.056	.000	.058	.121	.000	.059	.128

Table 9. $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \mid Y_{1(s)} < Y_{2(u)}\}$ -- Combined Method

$N_1 = N_2 = 10,\ n_1 = 4,\ n_2 = 2,\ t = 8$

		r = 3			r = 4			r = 5			r = 6		
k	s	u= 1	5	10	1	5	10	1	5	10	1	5	10
1	1	.617	.638	.640	.891	.898	.898	.971	.969	.969	.979	.976	.976
	5	.484	.594	.640	.856	.885	.898	.983	.975	.969	.995	.986	.976
	10	.333	.540	.623	.833	.871	.895	1.000	.981	.970	1.000	.993	.978
2	1	.467	.476	.477	.742	.736	.735	.821	.807	.805	.829	.814	.813
	5	.410	.466	.477	.782	.757	.735	.909	.847	.806	.921	.858	.813
	10	.333	.452	.474	.833	.784	.741	1.000	.894	.816	1.000	.906	.824
3	1	.118	.119	.119	.392	.379	.377	.472	.451	.448	.480	.458	.455
	5	.113	.130	.119	.485	.421	.378	.611	.511	.449	.624	.521	.456
	10	.200	.145	.122	.700	.476	.389	.867	.586	.464	.867	.598	.472
4	1				.063	.060	.060	.143	.131	.130	.151	.139	.137
	5				.086	.074	.060	.212	.164	.130	.225	.175	.138
	10				.167	.091	.063	.333	.200	.139	.333	.213	.147

of coverage (rather than a lower bound) for our confidence intervals, (2) we consider a finite number of elements in each stratum (i.e., a strictly finite population), and (3) our stratum sample sizes are arbitrary (i.e., we are not restricted to proportional allocation). We now show (using a counterexample) that McCarthy's main result does not necessarily hold if sampling is without replacement from a finite population. (This result can also be observed by examining Tables 5 and 6.)

Let Π_9 be a population of size 9, where all the elements have distinct values. A sample of size 3 is chosen without replacement. The probability that $[x_{(1)}, x_{(3)}]$ covers the median, $Y_{(5)}$, of Π_9 is, by (2.9), $\frac{19}{21}$.

Suppose the population is stratified into two strata of sizes $N_1 = 3$ and $N_2 = 6$, and we sample proportionally from each strata: $n_1 = 1$, $n_2 = 2$. There are seven types of stratification possible. In Table 10 we list these, along with the probability that $[y_{(1)}, y_{(3)}]$ covers $Y_{(5)}$, the probability of obtaining that particular stratification under the assumption of "random" stratification, and the product of these two probabilities. In the Type column, a "-" sign indicates a value below the population median, "0" the median, and a "+" sign a value above the median.

Table 10. Counter-example

Case	Type	$P_1\{y_{(1)} \leq Y_{(5)} \leq y_{(3)}\}$	$P_2\{\text{obtaining case}\}$	$P_1 \times P_2$
1	+++ ¦ ----0+	1	4/84	180/3780
2	0++ ¦ ----++	43/45	6/84	258/3780
3	-++ ¦ ---0++	40/45	24/84	960/3780
4	-0+ ¦ ---+++	39/45	16/84	624/3780
5	--+ ¦ --0+++	40/45	24/84	960/3780
6	--0 ¦ --++++	43/45	6/84	258/3780
7	--- ¦ -0++++	1	4/84	180/3780

In Cases 3 through 5, the probability of coverage is less than the 19/21 as found for the non-stratified situation.

Hence we have shown that using the combined sample order statistics may lead to a lower confidence coefficient than obtained sampling from the non-stratified population. Of course, in the case of random stratification, we know that the two expressions are equal. This is confirmed by our table, since the sum of the last column is 19/21.

b. Theoretical lower bounds. We now explore ways of approximating $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ for the two strata situation. As before, $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} = P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} \leq Y_{(t-1)}\}$. Furthermore,

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_j P\{y_{(k)} \leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\} \; P\{Y_{(t)} = Y_{1(j)}\}$$

$$+ \sum_{j^*} P\{y_{(k)} \leq Y_{(t)} \mid Y_{(t)} = Y_{2(j^*)}\} \; P\{Y_{(t)} = Y_{2(j^*)}\}. \tag{3.45}$$

Therefore,

$$P\{y_{(k)} \le Y_{(t)}\} \ge \min\begin{cases} P\{y_{(k)} \le Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\} \\ P\{y_{(k)} \le Y_{(t)} \mid Y_{(t)} = Y_{2(j^*)}\} \end{cases}, \tag{3.46}$$

and

$$P\{y_{(k)} \le Y_{(t)}\} \le \max\begin{cases} P\{y_{(k)} \le Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\} \\ P\{y_{(k)} \le Y_{(t)} \mid Y_{(t)} = Y_{2(j^*)}\} \end{cases}. \tag{3.47}$$

Using (3.11) and (3.12), we note that

$$P\{y_{(k)} \le Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\} = P\{y_{(k)} \le Y_{(t)} \mid Y_{(t)} = Y_{2(t-j)}\} \tag{3.48}$$

and so the minimum values and maximum values in (3.46) and 3.47) are identical; if the minimum (maximum) occurs when $j = p$, the minimum (or maximum) will occur when $j^* = t-p$. The only exception to this due to the fact that j (and j^*) cannot take on the value 0, whereas the possibility exists that j^* (and j) can assume the value t. The graphical work to follow demonstrates why we do not need to concern ourselves with this fact.

Unfortunately, it is extremely difficult to work with the convolutions of the hypergeometric distributions as they appear above. Hence we turn to the binomial approximation to the hypergeometric as given in Johnson and Kotz [1969, p. 148].

In this case,

$$P\{y_{(k)} \leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\}$$

$$\doteq \sum_{i=k}^{n} \sum_{\ell} \binom{n_1}{\ell} \left(\frac{j}{N_1}\right)^{\ell} \left(1 - \frac{j}{N_1}\right)^{n-\ell} \binom{n_2}{i-\ell} \left(\frac{t-j}{N_2}\right)^{i-\ell} \left(1 - \frac{t-j}{N_2}\right)^{n_2-(i-\ell)}$$

$$= P'\{y_{(k)} \leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\} \quad . \tag{3.49}$$

In Figure 3 we present $P\{y_{(k)} \leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\}$ for $k = 1, 2, \ldots, 10$, $N_1 = 20$, $N_2 = 80$, $n_1 = 2$, $n_2 = 8$ (i.e., proportional allocation), and $t = 50$, plotted as a function of j. Figure 4 shows the binomial approximation to this same data, and Figure 5 plots the same probability, with $n_1 = n_2 = 5$ (non-proportional allocation). Figure 5 also represents the binomial approximation.

A study of plots of these types would indicate that for proportional allocation, well-defined maxima and minima do exist. Below, we prove a theorem concerning them; this leads to approximate lower bounds for the confidence coefficient of the type suggested by McCarthy. This is followed with another approach which yields approximate lower bounds for both proportional and non-proportional allocation.

We now turn to finding the maximum and minimum values for (3.48).

Theorem 3.2: The maximum (minimum) of expression (3.49) occurs at $j = N_1(t/N)$ for $k \geq (nt/N + 1)$ $(\leq nt/N)$, assuming proportional allocation.

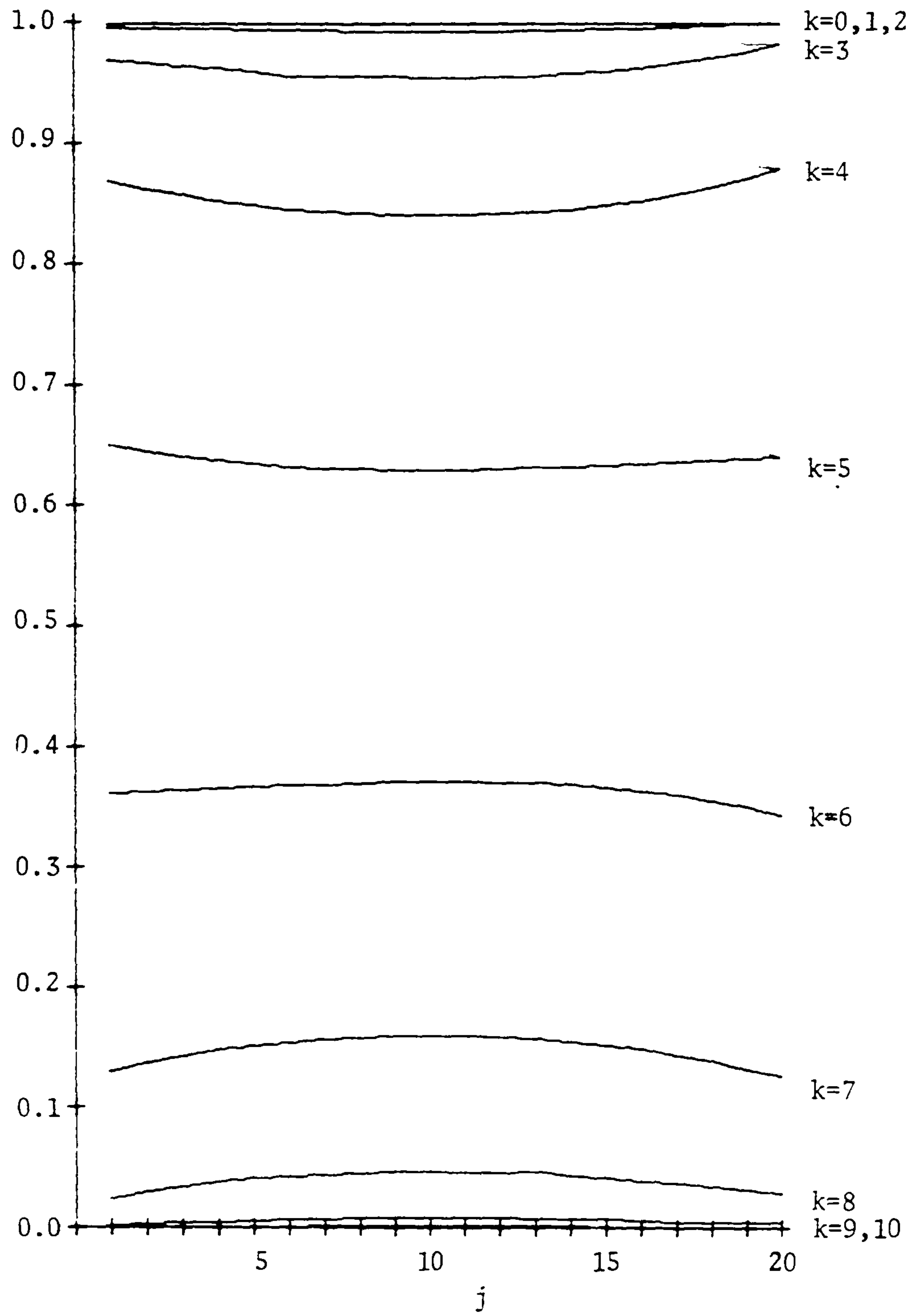

Figure 3. P{at least k observations $\leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}$}

$N_1 = 20$, $N_2 = 80$, $n_1 = 2$, $n_2 = 8$, $t = 50$

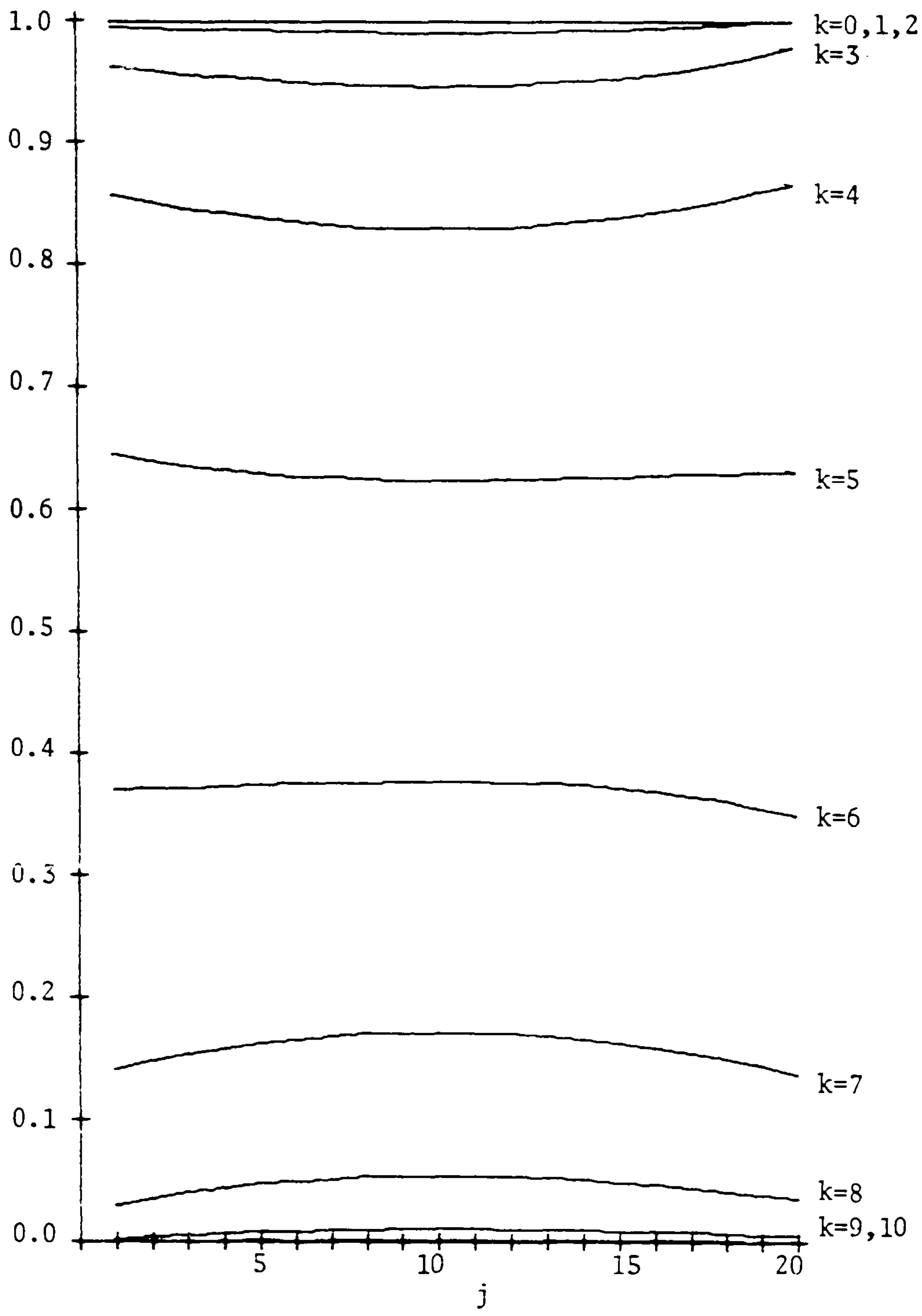

Figure 4. P{at least k observations $\leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}$}

$N_1 = 20$, $N_2 = 80$, $n_1 = 2$, $n_2 = 8$, $t = 50$ (Binomial approximation)

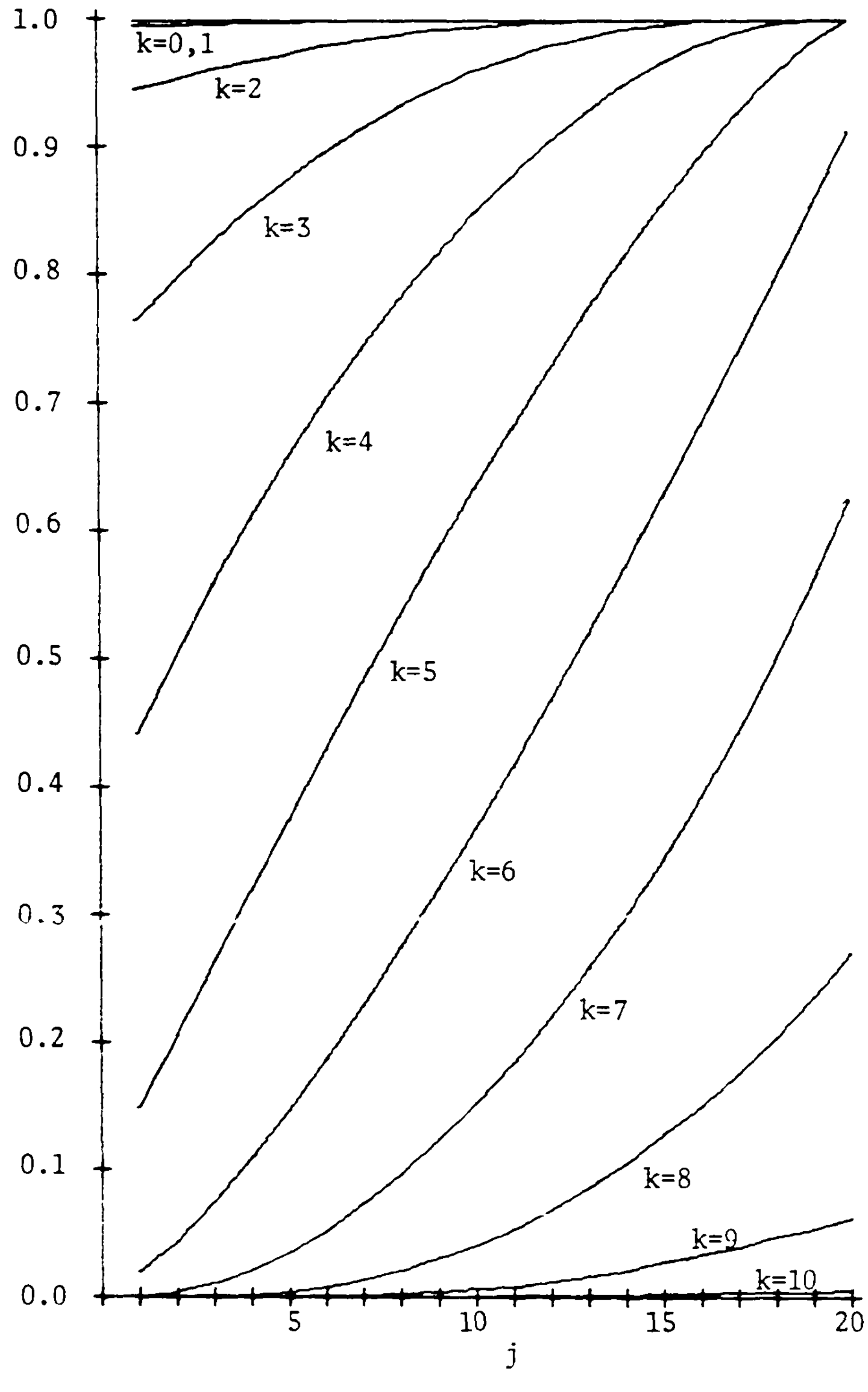

Figure 5. P{at least k observations $\leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}$}

$N_1 = 20$, $N_2 = 80$, $n_1 = n_2 = 5$, $t = 50$, (Binomial approximation)

Proof: Notationally,

$$\mathrm{Bi}(n,p;r) = \sum_{j=0}^{r} \binom{n}{j} p^j (1-p)^{n-j} \quad . \tag{3.50}$$

For $j = N_1(t/N)$, expression (3.49) becomes

$$\sum_{i=k}^{n} \sum_{\ell} \binom{n_1}{\ell} \left(\frac{t}{N}\right)^{\ell} \left(1-\frac{t}{N}\right)^{n_1-\ell} \binom{n_2}{i-\ell} \left(\frac{t}{N}\right)^{i-\ell} \left(1-\frac{t}{N}\right)^{n_2-(i-\ell)}$$

$$= \sum_{i=k}^{n} \left(\frac{t}{N}\right)^{i} \left(1-\frac{t}{N}\right)^{n_1+n_2-i} \sum_{\ell} \binom{n_1}{\ell} \binom{n_2}{i-\ell}$$

$$= \sum_{i=k}^{n} \binom{n}{i} \left(\frac{t}{N}\right)^{i} \left(1-\frac{t}{N}\right)^{n-i}$$

$$= 1 - \mathrm{Bi}(n,t/N;\ k-1) \quad . \tag{3.51}$$

Turning to Hoeffding's Theorem, as stated in Anderson and Samuels [1967, pp. 1-12]:

"Let $F(k)$ be the probability of not more than k successes in n independent trials where the i-th trial has probability p_i of success. Let $\lambda = p_1 + p_2 + \ldots + p_n$. Then

$$\mathrm{Bi}(n,\lambda/n;k) \begin{cases} \geq F(k) & \text{for } k \leq \lambda - 1 \\ \leq F(k) & \text{for } k \geq \lambda \end{cases} \quad .$$

Equality holds only if $p_1 = \ldots = p_n = \lambda/n$."

Applying this notation to the theorem we are proving, expression (3.49) becomes $1 - F(k-1)$; $p_1 = \ldots = p_{n_1} = j/N_1$; $p_{n_1+1} = \ldots = p_n = (t-j)/N_2$. Because of proportional allocation $\sum_{i=1}^{n} p_i = n_1 j/N_1 +$

$n_2(t-j)/N_2 = nt/N$. It then follows by Hoeffding's Theorem,

$$Bi(n,t/N;k-1) \begin{cases} \geq F(k-1) & \text{for } k-1 \leq (nt/N-1) \\ \leq F(k-1) & \text{for } k-1 \geq nt/N \quad . \end{cases} \tag{3.52}$$

Therefore,

$$1 - F(k-1) \begin{cases} \geq 1 - Bi(n,t/N;k-1) & \text{for } k \leq nt/N \\ \leq 1 - Bi(n,t/N;k-1) & \text{for } k \geq (nt/N)+1 \ , \end{cases} \tag{3.53}$$

with equality only if $j/N_1 = (t-j)/N_2$ so that $j = N_1t/N$, thus proving our theorem.

Hence,

$$\begin{aligned} P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} &= P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} \leq Y_{(t-1)}\} \\ &\geq \min_j P\{y_{(k)} \leq Y_{(t)} \mid Y_{(t)} = Y_{1(j)}\} \\ &\quad - \max_j P\{y_{(r)} \leq Y_{(t-1)} \mid Y_{(t-1)} = Y_{1(j)}\} \\ &\doteq 1 - Bi(n,t/N;k-1) - (1 - Bi(n,(t-1)/N;r-1)) \\ &= Bi(n,(t-1)/N;r-1) - Bi(n,t/N;k-1) \\ &\quad \text{provided } k \leq nt/N \text{ and } r \geq (n(t-1)/N)+1 \ . \end{aligned} \tag{3.54}$$

A second approach to this problem yields approximate lower bounds to the coverage probability for both the proportional and non-proportional allocation cases. The binomial approximation to the hypergeometric distribution is also used in this derivation.

Let

$$X_{ki} = \begin{cases} 1 & \text{if } y_{k(i)} \leq Y_{(t)} \\ 0 & \text{otherwise} \end{cases} \quad \begin{array}{l} k = 1, 2 \\ i = 1, \ldots, n_k \end{array} \qquad (3.55)$$

Let

$$Z = \sum_{i=1}^{n_1} X_{1i} + \sum_{i=1}^{n_2} X_{2i} \quad . \qquad (3.56)$$

Z is then the number of observations in the sample with associated Y-values less than or equal to $Y_{(t)}$.

Then

$$\begin{aligned} P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \mid Y_{(t)} = Y_{1(j)}\} \\ = P\{k \leq Z \leq r-1 \mid Y_{(t)} = Y_{1(j)}\} \\ + P\{y_{(r)} = Y_{1(j)} \mid Y_{(t)} = Y_{1(j)}\} \end{aligned} \qquad (3.57)$$

and

$$\begin{aligned} P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \mid Y_{(t)} = Y_{2(j^*)}\} \\ = P\{k \leq Z \leq r-1 \mid Y_{(t)} = Y_{2(j^*)}\} \\ + P\{y_{(r)} = Y_{2(j^*)} \mid Y_{(t)} = Y_{2(j^*)}\} \quad . \end{aligned} \qquad (3.58)$$

Therefore,

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$$

$$\geq \sum_{j} P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{1(j)}\}\, P\{Y_{(t)} = Y_{1(j)}\}$$

$$+ \sum_{j^*} P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{2(j^*)}\}\, P\{Y_{(t)} = Y_{2(j^*)}\}$$

$$\geq \min\left\{\begin{matrix} \min\limits_{j} P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{1(j)}\} \\ \min\limits_{j^*} P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{2(j^*)}\} \end{matrix}\right\} \quad . \tag{3.59}$$

Considering $P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{1(j)}\}$, we have

$$E(Z) = \frac{n_1}{N_1}\, j + \frac{n_2}{N_2}\,(t\text{-}j) \quad . \tag{3.60}$$

Switching to the binomial approximation and applying Hoeffding's Theorem as stated in McCarthy [1965, p. 776],

$$P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{1(j)}\} \geq \sum_{i=k}^{r-1} \binom{n}{i} p_1^i\, (1\text{-}p_1)^{n-i} \tag{3.61}$$

where

$$p_1 = \frac{1}{n}\left(\frac{n_1}{N_1}\, j + \frac{n_2}{N_2}\,(t\text{-}j)\right) \quad ,$$

$$0 \leq k \leq np_1 \leq r\text{-}1 \leq n \quad . \tag{3.62}$$

Similarly,

$$P\{k \leq Z \leq r\text{-}1 \mid Y_{(t)} = Y_{2(j^*)}\} \geq \sum_{i=k}^{r-1} \binom{n}{i} p_2^i\, (1\text{-}p_2)^{n-i} \tag{3.63}$$

where

$$p_2 = \frac{1}{n}\left(\frac{n_1}{N_1}(t-j^*) + \frac{n_2}{N_2}j^*\right)$$

$$0 \le k \le np_2 \le r-1 \le n \quad . \tag{3.64}$$

Thus,

$$P\{y_{(k)} \le Y_{(t)} \le y_{(r)}\}$$
$$\ge \min\left(\min_{j} \sum_{i=k}^{r-1} \binom{n}{i} p_1^i (1-p_1)^{n-i},\ \min_{j^*} \sum_{i=k}^{r-1} \binom{n}{i} p_2^i (1-p_2)^{n-i}\right). \tag{3.65}$$

In the case of proportional allocation,

$$\frac{n}{N} = \frac{n_i}{N_2} = \frac{n_2}{N_2}$$

and

$$p = p_1 = p_2 = \frac{t}{N} \quad . \tag{3.66}$$

Then

$$P\{y_{(k)} \le Y_{(t)} \le y_{(r)}\} \ge \sum_{i=k}^{r-1} \binom{n}{i} p^i (1-p)^{n-i} \tag{3.67}$$

provided

$$1 \le k \le \frac{nt}{N} \le r-1 \le n \quad .$$

For non-proportional allocation, we find the minimum value which $f(p_1) = \sum_{i=k}^{r-1} \binom{n}{i} p_1^i (1-p_1)^{n-i}$ can attain, recalling that p_1 is a function of j. Similarly, we find the minimum value of $f(p_2) = \sum_{i=k}^{r-1} \binom{n}{i} p_2^i (1-p_2)^{n-i}$, where p_2 is a function of j^*. Then $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$ is greater than or equal to the smaller of these two minima.

We now turn to finding the minima of $f(p_1)$ and $f(p_2)$. Let

$$f(p) = \sum_{i=k}^{r-1} \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \frac{n!}{(k-1)!(n-k)!} \int_0^p x^{k-1}(1-x)^{n-k}\, dx$$

$$- \frac{n!}{(r-1)!(n-r)!} \int_0^p x^{r-1} (1-x)^{n-r}\, dx \quad . \tag{3.68}$$

Then

$$f'_{(p)} = n \left[\binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} - \binom{n-1}{r-1} p^{r-1}(1-p)^{n-r} \right] \quad . \tag{3.69}$$

Setting (3.69) equal to zero and solving, the resulting polynomial has a root of multiplicity $(k-1)$ at zero, a root of multiplicity $(n-r)$ at one, so that, after removing these roots, we have the equation

$$\binom{n-1}{k-1} (1-p)^{r-k} - \binom{n-1}{r-1} p^{r-k} = 0 \quad . \tag{3.70}$$

Solving this equation for p, the only root in $(0,1)$ is

$$p = \frac{1}{1 + \left[\binom{n-1}{r-1} \Big/ \binom{n-1}{k-1} \right]^{1/(r-k)}} \quad . \tag{3.71}$$

Hence, since $f(p)$ is continuous on $[0,1]$, $f'(p)$ continuous on $(0,1)$, $f(0) = f(1) = 0$, and there exists exactly one p^* in $(0,1)$ such that $f'(p^*) = 0$, the minimum of $f(p)$ must occur "near" zero or one.

Recalling that p_1 is a function of j,

$$p_1 = \frac{1}{n}\left[j\left(\frac{n_1}{N_2} - \frac{n_2}{N_2}\right) + t\,\frac{n_2}{N_2}\right] ,$$

we have the following:

If $\frac{n_1}{N_1} > \frac{n_2}{N_2}$, the smallest value p can assume is when $j = \max\{1, t-N_2\}$. Call that value of p, "p min". Also, if $\frac{n_1}{N_1} > \frac{n_2}{N_2}$, the largest value p can assume is when $j = \min\{t, N_1\}$. Call that value of p, "p max".

Similarly, if $\frac{n_1}{N_1} < \frac{n_2}{N_2}$, the smallest value p can assume is when $j = \min\{t, N_1\}$ ("p min") and the largest value is when $j = \max\{1, t-N_2\}$ ("p max").

At this point, evaluate $f(p\ \min)$ and $f(p\ \max)$, and compare them to obtain $\min_j f(p_1)$.

Repeating the same procedure for Stratum II, with $p_2 = \frac{1}{n}\left(\frac{n_1}{N_1}t + \left(\frac{n_2}{N_2} - \frac{n_1}{N_1}\right)j^*\right)$, we are able to find $\min_{j^*} f(p_2)$.

Finally, comparing $\min_j f(p_1)$ and $\min_{j^*} f(p_2)$, the smaller of these will be an approximate j^* lower bound for $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$, under the assumption of expressions (3.62) and (3.64).

8. Sampling with replacement from a finite population

Let our population consist of $N = 2m - 1$ elements, having distinct associated Y-values. Obviously, $Y_{(m)}$ is the median. Using the notation of the first section, and using proportional sampling with replacement from each stratum, McCarthy [1965, pp. 776-777] derives a lower bound for $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\}$.

Let $W_1 = N_1/N$. Then W_1 and $1 - W_1$ are our strata weights. Let p_1 be the proportion of elements in Stratum I with associated values less than or equal to $Y_{(m)}$; p_2 the corresponding proportion in Stratum II. We then have $W_1p_1 + W_2p_2 = m/(2m-1)$ and $n_1p_1 + n_2p_2 = nm/(2m-1)$. If we let S_i = number of observations from the i-th stratum with values less than or equal to $Y_{(m)}$, and $S_1 + S_2 = S$, $E(S) = nm/(2m-1)$. Appealing to Hoeffding [1956, pp. 713-721], we then have

$$P\{k \leq S \leq r\} = P\{y_{(k)} \leq Y_{(m)} \leq y_{(r)}\}$$

$$\geq \sum_{i=k}^{r} \binom{n}{i} \left(\frac{m}{2m-1}\right)^{i} \left(\frac{m-1}{2m-1}\right)^{n-i} \quad . \tag{3.72}$$

In the case of an even-sized population,

$$P\{y_{(k)} < Y_{med} < y_{(r)}\} \geq \sum_{i=k}^{r} \binom{n}{i} \left(\frac{1}{2}\right)^{n} \quad . \tag{3.73}$$

This generalizes directly to more than two strata.

C. Confidence Intervals Derived from the Sample C.D.F.

1. Definition of the confidence interval

The confidence interval procedure described in Section B may not be satisfactory in all situations. Given a stratified simple random with, for example, $(n_1/N_1) \gg (n_2/N_2)$, a wide confidence interval of the form

$[y_{(k)}, y_{(r)}]$ might be necessary to achieve a desired confidence coefficient. (For example, with s large, u small, $N_1 = N_2$ and $Y_{(t)}$ the population median, both the upper and lower confidence limits would likely be variate values from Stratum I unless r were chosen to be very large.) Such possible difficulties may be eliminated by deriving a confidence interval from the sample C.D.F. defined by

$$\hat{F}(y) = \begin{cases} 0 & \text{if } y < y_{(1)} \\ j\dfrac{N_1}{n_1N} + (i-j)\dfrac{N_2}{n_2N} & \text{if } y_{(i)} \leq y < y_{(i+1)} \text{ and } y_{(i)} = y_{1(j)} \\ (i-j)\dfrac{N_1}{n_1N} + j\dfrac{N_2}{n_2N} & \text{if } y_{(i)} \leq y < y_{(i+1)} \text{ and } y_{(i)} = y_{2(j)} \\ 1 & \text{if } y \geq y_{(n_1+n_2)} \quad . \end{cases} \tag{3.74}$$

The graph of $\hat{F}(y)$ is illustrated in Figure 6. Note that the "jumps" corresponding to observations from Strata I and II are $j_1 = (N_1/n_1N)$ and $j_2 = (N_2/n_2N)$, respectively. If proportional sample size allocation is used, these "jumps" are equal.

To each pair (α, β) where $\alpha_0 < \alpha < \beta < 1$ and $\alpha_0 = \max\{(N_1/n_1N), (N_2/n_2N)\}$, there corresponds a unique pair of integers (k,r) where $1 \leq k \leq r \leq n$ such that

$$\begin{aligned} &\hat{F}(y_{(k)}) \leq \alpha < \hat{F}(y_{(k+1)}) \\ &\hat{F}(y_{(r-1)}) < \beta \leq \hat{F}(y_{(r)}) \end{aligned} \tag{3.75}$$

where $\hat{F}(y_{(0)}) = 0$, $\hat{F}(y_{(n+1)}) = 1$.

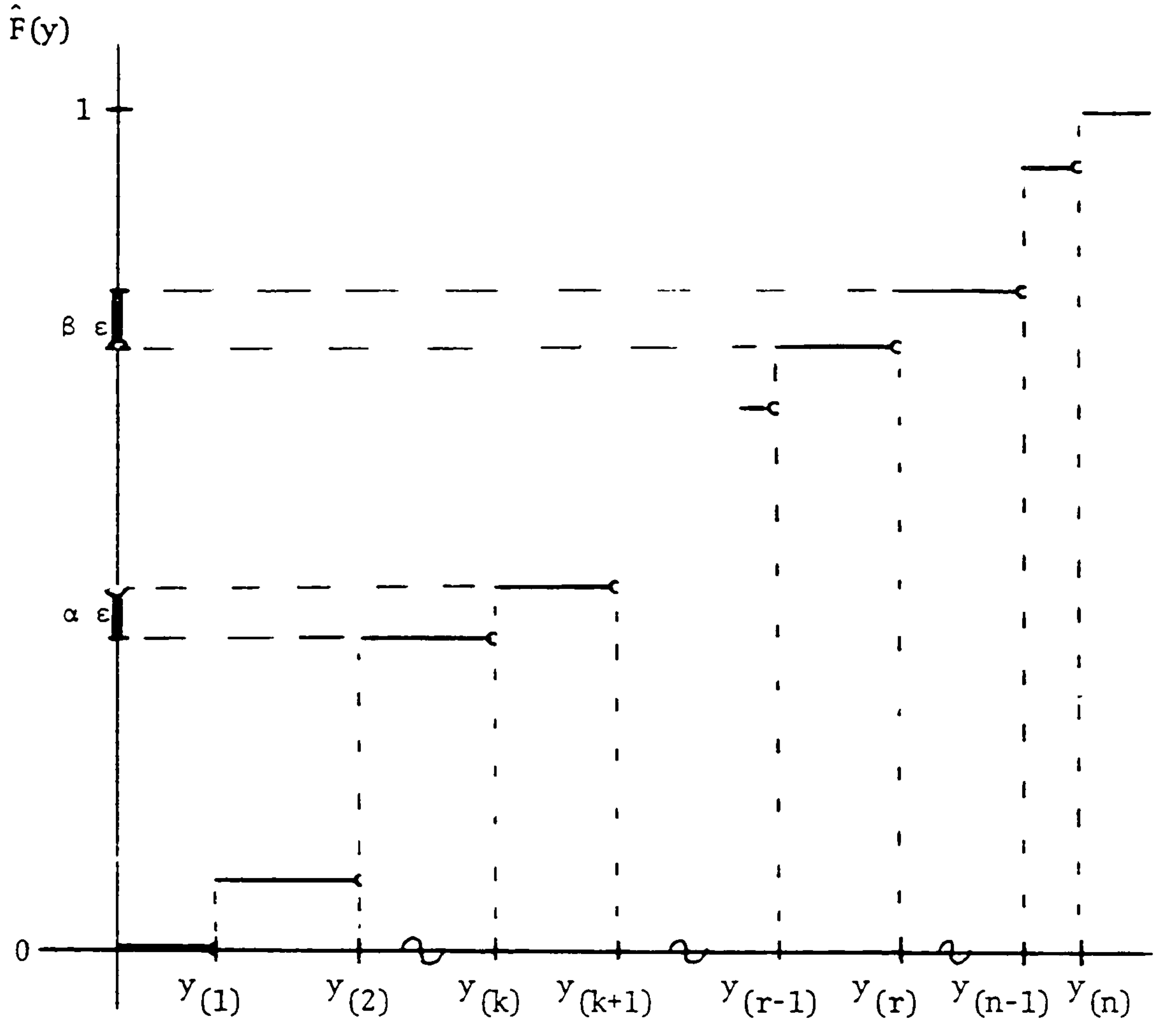

Figure 6. $\hat{F}(y)$. (C.D.F. method)

For given values of α and β, the confidence interval for $Y_{(t)}$ is given by $[y_{(k)}, Y_{(r)})$ where $y_{(k)}, y_{(r)}$ are defined by (3.75). Of course, the integers k and r may vary in repeated sampling from Π_N. This confidence interval procedure is illustrated in Figure 6. For <u>any</u> value of β in the shaded region on the Y-axis, the upper confidence limit is $y_{(r)}$; similarly for <u>any</u> value of α in the shaded region, the lower confidence limit is $y_{(k)}$.

In the case of proportional allocation, the jumps j_1 and j_2 will, of course, be equal, and in this case, the integers k and r will not vary for fixed α and β. Hence, for proportional allocation, the C.D.F. method is equivalent to the combined method.

2. A lower bound for the confidence coefficient

Before determining the exact confidence coefficient associated with the confidence interval defined in the previous subsection, a lower bound for the confidence coefficient is derived. This lower bound should provide a good approximation for the confidence coefficient.

$$\begin{aligned} P\{y_{(k)} \leq Y_{(t)} < y_{(r)}\} &= P\{\hat{F}(y_{(k)}) \leq \hat{F}(Y_{(t)}) < \hat{F}(y_{(r)})\} \\ &\geq P\{\alpha \leq \hat{F}(Y_{(t)}) < \beta\} \\ &= P\{\hat{F}(Y_{(t)}) < \beta\} - P\{\hat{F}(Y_{(t)}) < \alpha\} \qquad (3.76) \end{aligned}$$

where the inequality in (3.76) follows from the definitions given in (3.75).

Then,

$$P\{\hat{F}(Y_{(t)}) < \beta\} = \sum_{j=\max[1,t-N_2]}^{\min[t,N_1]} P\{\hat{F}(Y_{(t)})<\beta \mid Y_{(t)}=Y_{1(j)}\}P\{Y_{(t)}=Y_{1(j)}\}$$
$$+ \sum_{j^*=\max[1,t-N_1]}^{\min[t,N_2]} P\{\hat{F}(Y_{(t)})<\beta \mid Y_{(t)}=Y_{2(j^*)}\}P\{Y_{(t)}=Y_{2(j^*)}\}. \tag{3.77}$$

Let $\hat{F}(Y_{(t)}) = [m_1(t)N_1/n_1N] + [m_2(t)N_2/n_2N]$ where $m_i(t) = m_i$ denotes the number of observations in the sample from Stratum i with $Y \leq Y_{(t)}$. then

$$P\{\hat{F}(Y_{(t)}) < \beta \mid Y_{(t)} = Y_{1(j)}\}$$
$$= \sum_{m_1,m_2=0}^{\beta *} \binom{j}{m_1}\binom{N_1-j}{n_1-m_1}\binom{t-j}{m_2}\binom{N_2-(t-j)}{n_2-m_2} \Bigg/ \binom{N_1}{n_1}\binom{N_2}{n_2} \tag{3.78}$$

where $\sum^*$ denotes summation over all non-negative integers m_1 and m_2 such that $(m_1N_1/n_1N + m_2N_2/n_2N) < \beta$. Similarly,

$$P\{\hat{F}(Y_{(t)}) < \beta \mid Y_{(t)} = Y_{2(j^*)}\}$$
$$= \sum_{m_1,m_2=0}^{\beta *} \binom{j^*}{m_2^*}\binom{N_2-j^*}{n_2-m_2}\binom{t-j^*}{m_1}\binom{N_1-(t-j^*)}{n_1-m_1} \Bigg/ \binom{N_1}{n_1}\binom{N_2}{n_2} . \tag{3.79}$$

Then, (3.77) and, finally, (3.76) may be obtained by determining $P\{Y_{(t)} = Y_{1(j)}\}$ and $P\{Y_{(t)} = Y_{2(j^*)}\}$; several possibilities have been considered in Section B.

In Table 11 we give examples of these lower bounds under the assumption of "random" stratification for $P\{\hat{F}(Y_{(t)}) < \beta\}$ for a) $N_1 = N_2 = 50$, $n_1 = n_2 = 5$, $t = 50$; b) $N_1 = N_2 = 50$, $n_1 = 2$, $n_2 = 8$, $t = 50$; c) $N_1 = 20$, $N_2 = 80$, $n_1 = 5$, $n_2 = 5$, $t = 50$, for $\beta = .2\ (.1)\ 1.0$. The results for $N_1 = 20$, $N_2 = 80$, $n_1 = 2$, $n_2 = 8$, $t = 50$ are identical to case a).

Table 11. $P\{\hat{F}(Y_{(50)} < \beta\}$

β	N_1=50, N_2=50 n_1= 5, n_2= 5	N_1=50, N_2=50 n_1= 2, n_2= 8	N_1=20, N_2=80 n_1= 5, n_2= 5
.2	.0078	.0828	.0307
.3	.0458	.1547	.1503
.4	.1589	.3068	.2367
.5	.3703	.4271	.5000
.6	.6297	.6932	.6627
.7	.8411	.8453	.8497
.8	.9542	.9172	.9471
.9	.9922	.9936	.9873
1.0	.9994	.9994	.9994

3. Derivation of the exact confidence coefficient

For greater clarity, denote the upper confidence limit, $y_{(r)}$, by y_U and the lower confidence limit, $y_{(k)}$, by y_L. Then,

$$P\{y_L \leq Y_{(t)} < y_U\} = P\{y_L \leq Y_{(t)}\} - P\{y_U \leq Y_{(t)}\} \quad . \tag{3.80}$$

To determine $P\{y_U \leq Y_{(t)}\}$ one must first enumerate the possible "values" of U. Consider the set A of non-negative integers u_1' and u_2' such that $u_1'j_1 + u_2'j_2 \geq \beta$, $(u_1'-1)\, j_1 + u_2'j_2 < \beta$, and (of course) $y_{(u_1'+u_2')} = y_{1(u_1')}$; and the set B of non-negative integers u_1'' and u_2'' such that $u_1''j_1 + u_2''j_2 \geq \beta, u_1''j_1 + (u_2''-1)j_2 < \beta$ and $y_{(u_1''+u_2'')} = y_{2(u_2'')}$ where $j_i = N_i/n_iN$. The sets A and B include all of the values of $U = u_1 + u_2$ as well as an identification of the stratum from which the (u_1+u_2)-th order statistic (in the combined sample) comes. Thus,

$$P\{y_U \leq Y_{(t)}\} = \sum_{(u_1',u_2')\varepsilon A} P\{y_{(u_1'+u_2')} \leq Y_{(t)}\} + \sum_{(u_1'',u_2'')\varepsilon B} P\{y_{(u_1''+u_2'')} \leq Y_{(t)}\}$$

$$= \sum_{(u_1',u_2')\varepsilon A} \sum_{i=0}^{t-(u_1'+u_2')} P\{y_{(u_1'+u_2')} = Y_{(t-i)}\}$$

$$+ \sum_{(u_1'',u_2'')\varepsilon B} \sum_{i=0}^{t-(u_1''+u_2'')} P\{y_{(u_1''+u_2'')} = Y_{(t-i)}\}$$

$$= \sum_{(u_1',u_2')\varepsilon A} \sum_{i=0}^{t-(u_1'+u_2')} \sum_{j=u_1'}^{t-i} P\{y_{(u_1'+u_2')} = Y_{(t-i)} \mid$$

$$Y_{(t-i)} = Y_{1(j)}\}\, P\{Y_{(t-i)} = Y_{1(j)}\}$$

$$+ \sum_{(u_1'',u_2'')\varepsilon B} \sum_{i=0}^{t-(u_1''+u_2'')} \sum_{j^*=u_2''}^{t-i} P\{y_{(u_1''+u_2'')} = Y_{(t-i)} \mid$$

$$Y_{(t-i)} = Y_{2(j^*)}\}\, P\{Y_{(t-i)} = Y_{2(j^*)}\} \quad . \tag{3.81}$$

Then, it is easily seen that

$$P\{y_{(u_1'+u_2')} = Y_{(t-i)} | Y_{(t-i)} = Y_{1(j)}\}$$

$$= \binom{j-1}{u_1'-1}\binom{N_1-j}{n_1-u_1'}\binom{t-i-j}{u_2'}\binom{N_2-(t-i-j)}{n_2-u_2'} \Bigg/ \binom{N_1}{n_1}\binom{N_2}{n_2} \qquad (3.82)$$

and,

$$P\{y_{(u_1''+u_2'')} = Y_{(t-i)} | Y_{(t-i)} = Y_{2(j*)}\}$$

$$= \binom{j*-1}{u_2''-1}\binom{N_2-j*}{n_2-u_2''}\binom{t-i-j*}{u_1''}\binom{N_1-(t-i-j*)}{n_1-u_1''} \Bigg/ \binom{N_1}{n_1}\binom{N_2}{n_2} . \qquad (3.83)$$

Thus, $P\{y_U \leq Y_{(t)}\}$ may be obtained from (3.82) and (3.83) and a determination of $P\{Y_{(t-i)} = Y_{1(j)}\}$ and $P\{Y_{(t-i)} = Y_{2(j*)}\}$.

To find the components of the set A, note that for each value of u_2' ($u_2' = 0,1,\ldots, n_2$), if $0 \leq \{[(\beta-u_2'j_2)^-/j_1]+1\} \leq n_1$, then $([(\beta-u_2'j_2)^-/j_1]+1,u_2')\varepsilon A$. Similarly, for each value of u_1'' ($u_1'' = 0,1,\ldots, n_1$), if $0 \leq \{[(\beta-u_1''j_1)^-/j_2]+1\} \leq n_2$, then $(u_1'', [(\beta-u_1''j_1)^-/j_2]+1)\varepsilon B$.

To determine $P\{y_L \leq Y_{(t)}\}$ one must enumerate the possible "values" of L. Consider the set C of non-negative integers ℓ_1' and ℓ_2' such that $\ell_1'j_1 + \ell_2'j_2 \leq \alpha$, $(\ell_1' + 1)j_1 + \ell_2'j_2 > \alpha$, and (of course), $y_{(\ell_1'+\ell_2'+1)} = y_{1(\ell_1'+1)}$. Also, set D consists of those non-negative integers ℓ_1'' and ℓ_2'' such that $\ell_1''j_1 + \ell_2''j_2 \leq \alpha$, $\ell_1''j_1 + (\ell_2''+1)j_2 > \alpha$ and (of course), $y_{(\ell_1''+\ell_2''+1)} = y_{2(\ell_2''+1)}$. The sets C and D include all of the values of $L = \ell_1 + \ell_2$ as well as an identification of the stratum from which the $(\ell_1+\ell_2+1)$-th order statistic (in the combined sample) comes. Thus,

$$P\{y_L \leq Y_{(t)}\} = \sum_{(\ell_1', \ell_2')\varepsilon C} \sum_{i=0}^{t-(\ell_1'+\ell_2')} \sum_{j=1}^{t-i} \sum_{m=t-i+1}^{N} \sum_{p=j+1}^{m}$$

$$P\{y(\ell_1'+\ell_2') = Y_{(t-i)} \cap Y_{(t-i)} = Y_{1(j)}$$

$$\cap\ y_1(\ell_1'+1) = Y_{1(p)} \cap Y_{1(p)} = Y_{(m)}\}$$

$$+ \sum_{C} \sum_{i} \sum_{j^*=1}^{t-i} \sum_{m} \sum_{p=t-i-j^*+1}^{m}$$

$$P\{y_{(\ell_1'+\ell_2')} = Y_{(t-i)} \cap Y_{(t-i)} = Y_{2(j^*)}$$

$$\cap\ y_{1(\ell_1'+1)} = Y_{1(p)} \cap Y_{1(p)} = Y_{(m)}\}$$

$$+ \sum_{(\ell_1'',\ell_2'')\varepsilon D} \sum_{i} \sum_{j} \sum_{m} \sum_{p^*=t-i-j+1}^{m}$$

$$P\{y_{(\ell_1''+\ell_2'')} = Y_{(t-i)} \cap Y_{(t-i)} = Y_{1(j)}$$

$$\cap\ y_{2(\ell_2''+1)} = Y_{2(p^*)} \cap Y_{2(p^*)} = Y_{(m)}\}$$

$$+ \sum_{D} \sum_{i} \sum_{j^*} \sum_{m} \sum_{p^*=j^*+1}^{m}$$

$$P\{Y_{(\ell_1''+\ell_2'')} = Y_{(t-i)} \cap Y_{(t-i)} = Y_{2(j^*)}$$

$$\cap Y_{2(\ell_2''+1)} = Y_{2(p^*)} \cap Y_{2(p^*)} = Y_{(m)}\}$$

$$= \sum_{C} \sum_{i} \sum_{j} \sum_{m} \sum_{p} \binom{j-1}{\ell_1'-1}\binom{N_1-p}{n_1-\ell_1'-1}\binom{t-i-j}{\ell_2'}\binom{N_2-(m-p)}{n_2-\ell_2'} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2}$$

$$\times P\{Y_{1(j)} = Y_{(t-i)} \cap Y_{1(p)} = Y_{(m)}\}$$

$$+ \sum_{C} \sum_{i} \sum_{j^*} \sum_{m} \sum_{p} \binom{j^*-1}{\ell_2'-1}\binom{N_2-(m-p)}{n_2-\ell_2'}\binom{t-i-j^*}{\ell_1'}\binom{N_1-p}{n_1-\ell_1'-1} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2}$$

$$\times P\{Y_{2(j^*)} = Y_{(t-i)} \cap Y_{1(p)} = Y_{(m)}\}$$

$$+ \sum_{D} \sum_{i} \sum_{j} \sum_{m} \sum_{p^*} \binom{j-1}{\ell_1''-1}\binom{N_1-(m-p^*)}{n_1-\ell_1''}\binom{t-i-j}{\ell_2''}\binom{N_2-p^*}{n_2-\ell_2''-1} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2}$$

$$\times P\{Y_{1(j)} = Y_{(t-i)} \cap Y_{2(p^*)} = Y_{(m)}\}$$

$$+ \sum_{D} \sum_{i} \sum_{j^*} \sum_{m} \sum_{p^*} \binom{j^*-1}{\ell_2''-1}\binom{N_2-p^*}{n_2-\ell_2''-1}\binom{t-i-j^*}{\ell_1''}\binom{N_1-(m-p^*)}{n_1-\ell_1''} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2}$$

$$\times P\{Y_{2(j^*)} = Y_{(t-i)} \cap Y_{2(p^*)} = Y_{(m)}\} \quad . \tag{3.84}$$

To find the components of the set C, note that for each value of ℓ_2' ($\ell_2' = 0, 1, \ldots, n_2$), if $0 \leq [(\alpha - \ell_2')/j_1] \leq n_1 - 1$, then $([(\alpha - \ell_2')/j_1], \ell_2') \varepsilon C$. Similarly, for each value of ℓ_1'' ($\ell_1'' = 0, 1, \ldots, n_1$), if $0 \leq [(\alpha - \ell_1'' j_1)/j_2] \leq n_2 - 1$, then $(\ell_1'', [(\alpha - \ell_1'' j_1)/j_2)]) \varepsilon D$.

The computational expressions for $P\{Y_{1(j)} = Y_{(t-i)} \cap Y_{1(p)} = Y_{(m)}\}$, etc., are now derived. They are rather involved, computationally-wise, and no attempt has been made to program them.

Simplifying our notation, we derive an expression for

$$P\{Y_{(t)} = Y_{1(j)} \cap Y_{(t')} = Y_{1(j)}\} \quad , \quad \text{etc.}$$

<u>under the assumption $Y_{1(s)} < Y_{2(u)}$</u>. We will not explicitly write this assumption down; it is assumed to be implicit in each expression.

Now,

$$P\{Y_{(t)} = Y_{\ell(j)} \cap Y_{(t')} = Y_{\ell'(j')}\}$$

$$= P\{Y_{(t)} = Y_{\ell(j)}\} \, P\{Y_{(t')} = Y_{\ell'(j')} \mid Y_{(t)} = Y_{\ell(j)}\} , \tag{3.85}$$

$$\ell, \ell' = 1, 2 \quad .$$

The first term has been derived earlier.

For the second term, we have

$$P\{Y_{(t')} = Y_{\ell'(j')} \mid Y_{(t)} = Y_{\ell(j)} \cap Y_{1(s)} < Y_{2(u)}\}$$

$$= \frac{P\{Y_{(t')} = Y_{\ell'(j')} \cap Y_{1(s)} < Y_{2(u)} \mid Y_{(t)} = Y_{\ell(j)}\}}{P\{Y_{1(s)} < Y_{2(u)} \mid Y_{(t)} = Y_{\ell(j)}\}}$$

$$= \frac{\sum_{d} P\{Y_{(t')}=Y_{\ell'(j')} \cap Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)} | Y_{(t)}=Y_{\ell(j)}\}}{\sum_{d} P\{Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)} | Y_{(t)}=Y_{\ell(j)}\}}$$

$$= \sum_{d} P\{Y_{(t')}=Y_{\ell'(j')} | Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)} \cap Y_{(t)}=Y_{\ell(j)}\} \times P^* \tag{3.86}$$

where

$$P^* = \frac{P\{Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)} | Y_{(t)} = Y_{\ell(j)}\}}{\sum_{d} P\{Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)} | Y_{(t)} = Y_{\ell(j)}\}} \quad . \tag{3.87}$$

Using a technique previously employed, we spell out the various cases for $P\{Y_{(t')} = Y_{\ell'(j')} | Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)} \cap Y_{(t)} = Y_{\ell(j)}\}$.

	Case I	Case II	Case III	Case IV
	$Y_{(t)}=Y_{1(j)}$	$Y_{(t)}=Y_{1(j)}$	$Y_{(t)}=Y_{2(j)}$	$Y_{(t)}=Y_{2(j)}$
	$Y_{(t')}=Y_{1(j')}$	$Y_{(t')}=Y_{2(j')}$	$Y_{(t')}=Y_{1(j')}$	$Y_{(t')}=Y_{2(j')}$
$t < s+u+d$, $t' < s+u+d$	E_1	E_2	E_3	E_4
$t < s+u+d$, $t' = s+u+d$	E_5	E_6	E_7	E_8
$t < s+u+d$, $t' > s+u+d$	E_9	E_{10}	E_{11}	E_{12}
$t = s+u+d$, $t' > s+u+d$	E_{13}	E_{14}	E_{15}	E_{16}
$t > s+u+d$, $t' > s+u+d$	E_{17}	E_{18}	E_{19}	E_{20}

For each of these twenty cases we compute

$$P\{Y_{(t')} = Y_{\ell'(j')} | Y_{(t)} = Y_{\ell(j)} \cap Y_{1(s+d)} < Y_{2(u)} < Y_{1(s+d+1)}\}$$

$$= P\{E_i\} .$$

Case 1. $$P\{E_1\} = \binom{t'-t-1}{j'-j-1}\binom{s+d+(u-1)-t'}{s+d-j'}\bigg/\binom{s+d+(u-1)-t}{s+d-j} \qquad (3.88)$$

Case 2. $$P\{E_2\} = \binom{t'-t-1}{j'-(t-j)-1}\binom{s+d+(u-1)-t'}{(u-1)-j'}\bigg/\binom{s+d+(u-1)-t}{(u-1)-(t-j)} \qquad (3.89)$$

Case 3. $$P\{E_3\} = \binom{t'-t-1}{j'-(t-j)-1}\binom{s+d+(u-1)-t'}{s+d-j'}\bigg/\binom{s+d+(u-1)-t}{s+d-(t-j)} \qquad (3.90)$$

Case 4. $$P\{E_4\} = \binom{t'-t-1}{j'-j-1}\binom{s+d+(u-1)-t'}{(u-1)-j'}\bigg/\binom{s+d+(u-1)-t}{(u-1)-j} \qquad (3.91)$$

Case 5. $$P\{E_5\} = 0 \qquad (3.92)$$

Case 6. $$P\{E_6\} = \begin{cases} 1 & \text{if } j' = u \\ 0 & \text{otherwise} \end{cases} \qquad (3.93)$$

Case 7. $$P\{E_7\} = 0 \qquad (3.94)$$

Case 8. $$P\{E_8\} = \begin{cases} 1 & \text{if } j' = u \\ 0 & \text{otherwise} \end{cases} \qquad (3.95)$$

Case 9. $$P\{E_9\} = \binom{t'-(s+d+u)-1}{j'-(s+d)-1}\binom{N-t'}{N_1-j'}\bigg/\binom{N-(s+d+u)}{N_1-(s+d)} \qquad (3.96)$$

Case 10. $$P\{E_{10}\} = \binom{t'-(s+d+u)-1}{j'-u-1}\binom{N-t'}{N_2-j'}\Big/\binom{N-(s+d+u)}{N_2-u} \qquad (3.97)$$

Case 11. $$P\{E_{11}\} = P\{E_9\} \qquad (3.98)$$

Case 12. $$P\{E_{12}\} = P\{E_{10}\} \qquad (3.99)$$

Case 13. $$P\{E_{13}\} = 0 \qquad (3.100)$$

Case 14. $$P\{E_{14}\} = 0 \qquad (3.101)$$

Case 15. $$P\{E_{15}\} = \begin{cases} \binom{t'-t-1}{j'-(s+d)-1}\binom{N-t'}{N_1-j'}\Big/\binom{N-t}{N_1-(s+d)} & \text{if } j=u \\ 0 & \text{otherwise} \end{cases} \qquad (3.102)$$

Case 16. $$P\{E_{16}\} = \begin{cases} \binom{t'-t-1}{j'-j-1}\binom{N-t'}{N_2-j'}\Big/\binom{N-t}{N_2-j} & \text{if } j=u \\ 0 & \text{otherwise} \end{cases} \qquad (3.103)$$

Case 17. $$P\{E_{17}\} = \binom{t'-t-1}{j'-j-1}\binom{N-t'}{N_1-j'}\Big/\binom{N-t}{N_1-j} \qquad (3.104)$$

Case 18. $$P\{E_{18}\} = \binom{t'-t-1}{j'-(t-j)-1}\binom{N-t'}{N_2-j'}\Big/\binom{N-t}{N_2-(t-j)} \qquad (3.105)$$

Case 19. $$P\{E_{19}\} = \binom{t'-t-1}{j'-(t-j)-1}\binom{N-t'}{N_1-j'}\Big/\binom{N-t}{N_1-(t-j)} \qquad (3.106)$$

Case 20. $P\{E_{20}\} = \binom{t'-t-1}{j'-j-1}\binom{N-t'}{N_2-j'}\Big/\binom{N-t}{N_2-j}$. (3.107)

In order to find $P\{Y_{(t')} = Y_{\ell'(j')} | Y_{(t)} = Y_{\ell(j)}\}$ we must multiply each of the above by $P\{D_d\}/\sum_d P\{D_d\}$. In Section B we gave the expression for $P\{D_d\}$.

In the case of "Random" Stratification,

$$P\{Y_{1(j)} = Y_{(b)} \cap Y_{1(j')} = Y_{(v)}\}$$

$$= \binom{b-1}{j-1}\binom{v-b-1}{j'-j-1}\binom{N-v}{N_1-j'}\Big/\binom{N}{N_1} , \qquad (3.107a)$$

$$P\{Y_{1(j)} = Y_{(b)} \cap Y_{2(j')} = Y_{(v)}\}$$

$$= \binom{b-1}{b-j}\binom{v-b-1}{j'-(b-j)-1}\binom{N-v}{N_2-j'}\Big/\binom{N}{N_2} , \qquad (3.107b)$$

$$P\{Y_{2(j)} = Y_{(b)} \cap Y_{1(j')} = Y_{(v)}\}$$

$$= \binom{b-1}{b-j}\binom{v-b-1}{j'-(b-j)-1}\binom{N-v}{N_1-j'}\Big/\binom{N}{N_1} , \qquad (3.107c)$$

$$P\{Y_{2(j)} = Y_{(b)} \cap Y_{2(j')} = Y_{(v)}\}$$

$$= \binom{b-1}{j-1}\binom{v-b-1}{j'-j-1}\binom{N-v}{N_2-j'}\Big/\binom{N}{N_2} . \qquad (3.107d)$$

4. Woodruff's Technique

As a special case of these results, we refer to Woodruff [1952, pp. 635-646]. On page 642 of his article we read:

> "In conclusion, it appears that confidence intervals for the median and other quantiles can be approximated for any sampling design where the variance of items less than a stated value can be acceptably estimated (in general, where large samples are involved)."

The technique employed by Woodruff is similar to that which we have been discussing, with the following exceptions.

The discreteness of the empirical C.D.F. is ignored, and hence the inequality which we display in (3.76) is eliminated.

Woodruff makes no statements concerning confidence coefficients for his intervals. Instead, his approach--using the median for an example--is to pick for α and β the values $(.5 - k\sigma_p)$ and $(.5 + k\sigma_p)$ respectively, where σ_p^2 is the variance of the proportion of items in the sample less than the true median, and k is a positive constant. Hence, he talks about one, two, and three standard deviation confidence intervals for a quantile, but unless he makes an assumption such as the percentage of items in the sample which are less than the population quantile follows a normal distribution, no exact confidence coefficients can be given.

Also, since in most cases, σ_p^2 can only be estimated, this also puts some variability into the choosing of α and β .

Table 12 lists the exact confidence coefficients for the median for $(\alpha,\beta) = (.5 - k\sigma_{.5}, .5+k\sigma_{.5})$ where $k = 1$ and 2, where we are taking a simple random sample of size 9 from a population of sizes $N = 15, 27, 99, 198$.

Table 12. $P\{.5-k\sigma_p \leq \hat{F}(Y_{(t)}) < .5+k\sigma_p\}$

N	n	t	σ_p	$P\{.5-\sigma_p\leq\hat{F}(Y_{(t)})<.5+\sigma_p\}$	$P\{.5-2\sigma_p\leq\hat{F}(Y_{(t)})<.5+2\sigma_p\}$
15	9	8	.1091	.6853	.9594
27	9	14	.1387	.5803	.8968
99	9	50	.1597	.5130	.9692
198	9	99	.1633	.5024	.9651

Applying Woodruff's method to our C.D.F. technique, he would say

$$P\{-1.96 < \frac{\hat{F} - E(\hat{F})}{\sigma_{\hat{F}}} < 1.96\} \doteq .95 \tag{3.109}$$

or

$$P\{E(\hat{F}) - 1.96\ \sigma_{\hat{F}} < \hat{F} < E(\hat{F}) + 1.96\ \sigma_{\hat{F}}\} \doteq .95 \ . \tag{3.110}$$

That is, choosing

$$\begin{aligned} \alpha &= E(\hat{F}) - 1.96\ \sigma_{\hat{F}} \\ \beta &= E(\hat{F}) + 1.96\ \sigma_{\hat{F}} \end{aligned} \tag{3.111}$$

would yield approximately a 95% confidence interval. Using for $E(\hat{F})$ the quantile we are considering (e.g., $E(\hat{F}) = .5$ for the median), we now must come up with $\sigma_{\hat{F}}$.

Now

$$\hat{F}(y) = \frac{1}{N}\left\{\frac{N_1}{n}\sum_{i=1}^{n_1} x_{1i} + \frac{N_2}{n_2}\sum_{i=1}^{n_2} x_{2i}\right\} \tag{3.112}$$

where

$$x_{\ell i} = \begin{cases} 1 & \text{if } y_{\ell(i)} \le y \\ 0 & \text{otherwise} \end{cases},$$

$y_{\ell(i)}$ being the i-th sample element from the ℓ-th stratum. Then, $\mathrm{Var}(x_i) = \Pi_{\ell i}(1 - \Pi_{\ell i})$, where $\Pi_{\ell i} = P\{y_{\ell(i)} \le y\}$. Ignoring the F.P.C.,

$$\mathrm{Var}[\hat{F}(y)] \doteq \frac{N_1^2}{N^2}\,\frac{P_1(1-P_1)}{n_1} + \frac{N_2^2}{N^2}\,\frac{P_2(1-P_2)}{n_2} \tag{3.113}$$

where P_ℓ = proportion of elements in Stratum ℓ (ℓ=1,2) with $Y_{\ell(i)} \le Y_{(t)}$.

In Monte Carlo Study III (see Section E.7) the actual values for P_1 and P_2 were .6 and .2, respectively. Hence, $\sigma_{\hat{F}} = .187$. In note 7 in Section E.7, we discuss the Monte Carlo findings in terms of the Woodruff technique.

Of course, if P_1 and P_2 were unknown, $\sigma_{\hat{F}}$ would have to be estimated from the sample data.

D. The Separate Strata Method

When it is known that, for example, $Y_{1(s)} < Y_{2(u)}$, an alternative form of confidence interval may be advantageous in some situations. For instance, let N_1 and N_2 be odd and consider a confidence interval for $Y_{([N_1+N_2]/2)}$--(approximately) the finite population median. If $Y_{1([N_1+1]/2)} < Y_{2([N_2+1]/2)}$ (i.e., the median within Stratum I is less than the median within Stratum II), it is easily shown that

$Y_{([N_1+N_2]/2)} \geq Y_{1([N_1+1]/2)}$ and $Y_{([N_1+N_2]/2)} \leq Y_{2([N_2-1]/2)}$. Then, a confidence interval for $Y_{([N_1+N_2]/2)}$ of the form $[y_{1(k)}, y_{2(r)}]$ may be desirable. In Section D.1, the confidence coefficient associated with a somewhat more general form of confidence interval for $Y_{(t)}$ is considered: If $y_{1(k)} < y_{2(r)}$, the confidence interval is $[y_{1(k)}, y_{2(r)}]$, whereas if $y_{2(r)} < y_{1(k)}$ the interval is $[y_{2(r)}, y_{1(k)}]$.

1. Derivation of the confidence coefficient

The confidence coefficient associated with the confidence interval procedure defined in the previous paragraph is given by

$$P\{y_{1(k)} \leq Y_{(t)} \leq y_{2(r)}\} + P\{y_{2(r)} \leq Y_{(t)} \leq y_{1(k)}\} \ . \tag{3.114}$$

Now,

$$P\{y_{1(k)} \leq Y_{(t)} \leq y_{2(r)}\} + P\{y_{2(r)} \leq Y_{(t)} \leq y_{1(k)}\}$$

$$= \sum_{j=1}^{N_1} \{P\{y_{1(k)} \leq Y_{(t)} \leq y_{2(r)} | Y_{(t)} = Y_{1(j)}\}$$

$$+ P\{y_{2(r)} \leq Y_{(t)} \leq y_{1(k)} | Y_{(t)} = Y_{1(j)}\}\}$$

$$\times P\{Y_{(t)} = Y_{1(j)}\}$$

$$+ \sum_{j^*=1}^{N_2} \{P\{y_{1(k)} \leq Y_{(t)} \leq y_{2(r)} | Y_{(t)} = Y_{2(j^*)}\}$$

$$+ P\{y_{2(r)} \leq Y_{(t)} \leq y_{1(k)} | Y_{(t)} = Y_{2(j^*)}\}\}$$

$$\times P\{Y_{(t)} = Y_{2(j^*)}\}$$

$$= \sum_{j=1}^{N_1} \{P\{y_{1(k)} \leq Y_{(t)} | Y_{(t)} = Y_{1(j)}\} P\{y_{2(r)} \geq Y_{(t)} | Y_{(t)} = Y_{1(j)}\}$$

$$+ P\{y_{2(r)} \leq Y_{(t)} | Y_{(t)} = Y_{1(j)}\} P\{y_{1(k)} \geq Y_{(t)} | Y_{(t)} = Y_{1(j)}\}\}$$

$$\times P\{Y_{(t)} = Y_{1(j)}\}$$

$$+ \sum_{j^*=1}^{N_2} \{P\{y_{1(k)} \leq Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\} P\{y_{2(r)} \geq Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\}$$

$$+ P\{y_{2(r)} \leq Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\} P\{y_{1(k)} \geq Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\}\}$$

$$\times P\{Y_{(t)} = Y_{2(j^*)}\} \quad . \qquad (3.115)$$

Looking at these terms individually, $P\{Y_{(t)} = Y_{1(j)}\}$ and $P\{Y_{(t)} = Y_{2(j^*)}\}$ have been given earlier in (3.14), (3.15), (3.41), and (3.44) under the assumptions of "random" stratification and $\{Y_{1(s)} < Y_{2(u)}\}$.

Also,

$$P\{y_{1(k)} \leq Y_{(t)} | Y_{(t)} = Y_{1(j)}\}$$

$$= P\{y_{1(k)} \leq Y_{1(j)}\}$$

$$= \sum_{i=\max[k,j+n_1-N_1]}^{\min[j,n_1]} \binom{j}{i}\binom{N_1-j}{n_1-i} \Big/ \binom{N_1}{n_1} \quad (3.116)$$

$$P\{y_{2(r)} \leq Y_{(t)} | Y_{(t)} = Y_{1(j)}\}$$

$$= \sum_{i=r}^{\min[t-j,n]} \binom{t-j}{i}\binom{N_2-(t-j)}{n_2-i} \Big/ \binom{N_2}{n_2} \quad (3.117)$$

$$P\{y_{1(k)} \geq Y_{(t)} | Y_{(t)} = Y_{1(j)}\}$$

$$= 1 - P\{y_{1(k)} \leq Y_{1(j-1)}\}$$

$$= 1 - \sum_{i=k}^{\min[j-1,n_1]} \binom{j-1}{i}\binom{N_1-(j-1)}{n_1-i} \Big/ \binom{N_1}{n_1} \quad (3.118)$$

$$P\{y_{2(r)} \geq Y_{(t)} | Y_{(t)} = Y_{1(j)}\}$$

$$= 1 - P\{y_{2(r)} \leq Y_{(t)} | Y_{(t)} = Y_{1(j)}\} \ . \quad (3.119)$$

Similarly,

$$P\{y_{1(k)} \leq Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\}$$

$$= \sum_{i=k}^{\min[n_1,t-j^*]} \binom{t-j^*}{i}\binom{N_1-(t-j^*)}{n_1-i} \Big/ \binom{N_1}{n_1} \quad (3.120)$$

$$P\{y_{2(r)} \le Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\}$$

$$= \sum_{i=k}^{\min[j^*,n_2]} \binom{j^*}{i}\binom{N_2-j^*}{n_2-i} \Big/ \binom{N_2}{n_2} \tag{3.121}$$

$$P\{y_{1(k)} \ge Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\}$$

$$= 1 - P\{y_{1(k)} \le Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\} \tag{3.122}$$

$$P\{y_{2(r)} \ge Y_{(t)} | Y_{(t)} = Y_{2(j^*)}\}$$

$$= 1 - \sum_{i=r}^{\min[j^*-1,n_2]} \binom{j^*-1}{i}\binom{N_2-(j^*-1)}{n_2-i} \Big/ \binom{N_2}{n_2} \ . \tag{3.123}$$

If the probability that $y_{2(r)} < y_{1(k)}$ is small, an approximation for the confidence coefficient, (3.114), is given by

$$P\{y_{1(k)} \le Y_{(t)} \le y_{2(r)}\} \doteq P\{y_{1(k)} \le Y_{(t)}\} - P\{y_{2(r)} < Y_{(t)}\}. \tag{3.124}$$

Note that each of the terms of the right-hand side of (3.124) can be expressed in a relatively simple form. For the first term,

$$P\{y_{1(k)} \le Y_{(t)}\} = \sum_{j=1}^{t} \sum_{i=k}^{n_1} \binom{j}{i}\binom{N_1-j}{n_1-i} \Big/ \binom{N_1}{n_1} P\{Y_{(t)} = Y_{1(j)}\}$$
$$+ \sum_{j^*=1}^{t} \sum_{i=k}^{n_1} \binom{t-j^*}{i}\binom{N_2-(t-j^*)}{n_1-i} \Big/ \binom{N_1}{n_1} P\{Y_{(t)} = Y_{2(j^*)}\} \ . \tag{3.125}$$

Similarly,

$$P\{y_{2(r)} < Y_{(t)}\} = \sum_{j=1}^{t-1} \sum_{i=r}^{n_2} \binom{t-1-j}{i}\binom{N_2-(t-1-j)}{n_2-i}\Bigg/\binom{N_2}{n_2} P\{Y_{(t-1)} = Y_{1(j)}\}$$
$$+ \sum_{j^*=1}^{t-1} \sum_{i=r}^{n_2} \binom{j^*}{i}\binom{N_2-j^*}{n_2-i}\Bigg/\binom{N_2}{n_2} P\{Y_{(t-1)} = Y_{2(j^*)}\} \quad . \tag{3.126}$$

2. <u>A lower bound for the confidence coefficient</u>

One may obtain a lower bound for the confidence coefficient associated with the confidence interval $[y_{1(k)}, y_{2(r)}]$ for $Y_{(t)}$ when it is assumed that $Y_{1(s)} < Y_{2(u)}$. This lower bound may be evaluated very easily, and, therefore, should be useful in many applications. Three cases must be considered:

<u>Case I</u>. $u \leq t \leq s + u - 1$.

With $u \leq t \leq s + u - 1$, it is easily shown that $Y_{(t)} \geq Y_{1(t-u+1)}$ and $Y_{(t)} \leq Y_{2(u-1)}$. Let A_1 denote the event "$y_{1(k)} \leq Y_{1(t-u+1)}$", and A_2 the event "$y_{2(r)} \geq Y_{2(u)}$". Then it can be shown that A_1, A_2, and $Y_{1(s)} < Y_{2(u)}$ imply that $y_{1(k)} \leq Y_{(t)} \leq y_{2(r)}$ for any value of $Y_{(t)}$ in the intervals specified above. Thus,

$$P\{y_{1(k)} \leq Y_{(t)} \leq y_{2(r)}\} \geq P_1P_2 \tag{3.127}$$

where $P_1 = P\{A_1\} = P\{y_{1(k)} \leq Y_{1(t-u+1)}\}$ and $P_2 = P\{A_2\} = P\{y_{2(r)} \geq Y_{2(u)}\}$. It should be noted that both P_1 and P_2 may be evaluated very easily (see (2.7)) since each of A_1 and A_2 refers to simple random sampling from a single stratum.

Case II. $s + u + 1 \leq t \leq s + N_2$

With $s + u + 1 \leq t \leq s + N_2$, it is easily shown that $Y_{(t)} \geq Y_{1(s+1)}$ and $Y_{(t)} \leq Y_{2(t-s)}$. Let B_1 denote the event "$y_{1(k)} \leq Y_{1(s)}$", and B_2 the event "$y_{2(r)} \geq Y_{2(t-s)}$". Then, it can be shown that B_1, B_2, and $Y_{1(s)} < Y_{2(u)}$ imply that $y_{1(k)} \leq Y_{(t)} \leq y_{2(r)}$ for any value of $Y_{(t)}$ in the intervals specified above. Thus,

$$P\{y_{1(k)} \leq Y_{(t)} \leq y_{2(r)}\} \geq P_1^* P_2^* \tag{3.128}$$

where $P_1^* = P\{B_1\} = P\{y_{1(k)} \leq Y_{1(s)}\}$ and $P_2^* = P\{B_2\} = P\{y_{2(r)} \geq Y_{2(t-s)}\}$. Again, P_1^* and P_2^* are easy to evaluate.

Case III. $t = s + u$

With $Y_{1(s)} < Y_{2(u)}$, $Y_{(t)} = Y_{2(u)}$. Thus, a confidence interval of the form $[y_{2(k)}, y_{2(r)}]$ is appropriate for this case.

3. Tables for the Separate Strata Method

In Tables 13-17 we tabulate the probability of coverage of $Y_{(t)}$ by the interval determined by $y_{1(k)}$ and $y_{2(r)}$, under the assumption $Y_{1(s)} < Y_{2(u)}$. The cases are completely analogous to Tables 5-9 in Section C on the Combined method. Notationally, "$P\{\{y_{1(k)}, y_{2(r)}\}$ covers $Y_{(t)} | Y_{1(s)} < Y_{2(u)}\}$" means "the probability that the interval determined by $y_{1(k)}$ and $y_{2(r)}$ covers $Y_{(t)}$, given $Y_{1(s)} < Y_{2(u)}$".

Table 13. $P\{\{y_{1(k)}, y_{2(r)}\}$ covers $Y_{(t)} | Y_{1(s)} < Y_{2(u)}\}$

$N_1 = N_2 = 10,\ n_1 = n_2 = 3,\ t = 4$

			r = 1			r = 2			r = 3	
k	s	u= 1	5	10	1	5	10	1	5	10
1	1	.667	.664	.667	.619	.564	.560	.596	.520	.512
	5	.833	.667	.667	.833	.619	.561	.833	.596	.513
	10	.833	.695	.667	.833	.681	.575	.833	.671	.533
2	1	.501	.556	.560	.174	.171	.174	.126	.092	.091
	5	.333	.502	.559	.333	.174	.173	.333	.126	.092
	10	.333	.447	.546	.333	.200	.174	.333	.176	.100
3	1	.427	.506	.512	.057	.087	.091	.007	.007	.007
	5	.033	.427	.511	.033	.056	.091	.033	.007	.007
	10	.033	.327	.491	.033	.037	.083	.033	.011	.007

Table 14. $P\{\{y_{1(k)}, y_{2(r)}\}$ covers $Y_{(t)} | Y_{1(s)} < Y_{2(u)}\}$

$N_1 = N_2 = 10,\ n_1 = n_2 = 3,\ t = 8$

			r = 1			r = 2			r = 3	
k	s	u= 1	5	10	1	5	10	1	5	10
1	1	.434	.430	.434	.745	.716	.715	.838	.809	.805
	5	.637	.434	.433	.939	.792	.715	.960	.891	.806
	10	1.000	.531	.434	1.000	.882	.733	1.000	.943	.826
2	1	.684	.711	.715	.549	.546	.549	.416	.382	.380
	5	.645	.637	.714	.682	.549	.548	.663	.480	.381
	10	.933	.627	.697	.933	.639	.549	.933	.617	.404
3	1	.772	.803	.805	.344	.375	.380	.097	.096	.097
	5	.574	.719	.804	.248	.279	.379	.181	.097	.097
	10	.467	.611	.784	.467	.252	.356	.467	.159	.097

Table 15. $P\{\{y_{1(k)}, y_{2(r)}\}$ covers $Y_{(t)} | Y_{1(s)} < Y_{2(u)}\}$

$N_1 = N_2 = 10,\ n_1 = 2,\ n_2 = 4,\ t = 4$

		r = 1			r = 2			r = 3			r = 4		
k	s u=	1	5	10	1	5	10	1	5	10	1	5	10
1	1	.649	.677	.681	.499	.481	.482	.442	.384	.380	.439	.375	.369
	5	.667	.650	.681	.667	.498	.482	.667	.442	.381	.667	.438	.369
	10	.667	.638	.673	.667	.535	.486	.667	.507	.396	.667	.505	.386
2	1	.555	.635	.640	.141	.186	.192	.050	.044	.045	.046	.032	.032
	5	.133	.556	.639	.133	.141	.192	.133	.050	.045	.133	.046	.032
	10	.133	.451	.619	.133	.115	.180	.133	.067	.046	.133	.066	.035

Table 16. $P\{\{y_{1(k)}, y_{2(r)}\}$ covers $Y_{(t)} | Y_{1(s)} < Y_{2(u)}\}$

$N_1 = N_2 = 10,\ n_1 = 2,\ n_2 = 4,\ t = 8$

		r = 1			r = 2			r = 3			r = 4		
k	s u=	1	5	10	1	5	10	1	5	10	1	5	10
1	1	.472	.486	.490	.653	.638	.639	.706	.681	.679	.694	.662	.658
	5	.572	.445	.489	.844	.670	.639	.862	.751	.680	.860	.755	.659
	10	.978	.494	.480	.978	.770	.646	.978	.833	.696	.978	.835	.682
2	1	.807	.835	.837	.550	.584	.589	.274	.280	.283	.180	.162	.161
	5	.650	.762	.836	.439	.480	.588	.347	.258	.283	.340	.214	.161
	10	.622	.675	.818	.622	.440	.563	.622	.322	.277	.622	.308	.174

Table 17. $P\{[y_{1(k)}, y_{2(r)}]$ covers $Y_{(t)} | Y_{1(s)} < Y_{2(u)}\}$

$N_1 = N_2 = 10,\ n_1 = 4,\ n_2 = 2,\ t = 8$

k	s u=	r = 1: 1	5	10	r = 2: 1	5	10
1	1	.508	.489	.490	.867	.840	.837
	5	.743	.534	.490	.983	.912	.838
	10	1.000	.642	.500	1.000	.958	.855
2	1	.625	.636	.639	.628	.591	.589
	5	.722	.608	.639	.864	.697	.590
	10	1.000	.668	.632	1.000	.819	.615
3	1	.652	.676	.679	.293	.282	.283
	5	.594	.608	.679	.475	.309	.283
	10	.867	.594	.663	.867	.431	.290
4	1	.622	.655	.658	.142	.158	.161
	5	.422	.561	.657	.113	.108	.160
	10	.333	.457	.634	.333	.109	.148

In studying these tables, we note some interesting points.

1. In Tables 13 and 14, for $(s,u) = (1,10)$ (essentially random stratification), the probabilities associated with $(k,r) = (a,b)$ are equal to those for $(k,r) = (b,a)$, $a = 1, 2, 3$, $b = 1, 2, 3$. This is due to the fact that, for random stratification, $P\{y_{1(k)} \leq Y_{(t)}\} = P\{y_{2(k)} \leq Y_{(t)}\}$ under proportional allocation.

2. In Table 13, the entries for any fixed k and (s,u) pair are non-increasing as r increases. This is due to the fact that, while increasing r does increase the length of the confidence interval $[y_{1(k)}, y_{2(r)}]$, at the same time it decreases the length of $[y_{2(r)}, y_{1(k)}]$. Hence, since we are working with a low quantile $(t = 4)$, the interval

$[y_{2(r)}, y_{1(k)}]$ will cover $Y_{(t)}$ less often, and so decrease the probability of coverage. Similar comments can be made for the remaining tables.

3. The entries for a fixed (k,r) pair differ considerably for various combinations of (s,u). Considerable gains (or losses) in confidence coefficients are possible. Hence, if good estimates of (s,u) are known, considerable gain can be made in the coefficients using the "separate" technique.

4. Comparing Tables 16 and 17, in the entries for $(s,u) = (1,10)$, for $(k,r) = (a,b)$ in Table 16 and $(k,r) = (b,a)$ in Table 17, $a = 1, 2, 3, 4$; $b = 1, 2$; we have equality in confidence coefficients. Keeping in mind that $(s,u) = (1,10)$ corresponds to random stratification, this result is not surprising.

5. In all tables, the entries for $(s,u) = (10,1)$ (complete ordering of strata), the confidence coefficient is constant for k fixed, $r = 1, 2, 3, 4$. This is because $Y_{(t)}$ is in Stratum I and hence, $y_{2(r)}$ is greater than $Y_{(t)}$ and the confidence interval depends only on $y_{1(k)}$.

E. Methods of comparing the alternative procedures

Using any of the confidence interval procedures suggested in Chapters II and III one may obtain a confidence interval with known confidence coefficient. However, it is not apparent which of the methods to use. First, given that a stratified simple random sample is to be selected, the "combined strata" approach, the sample C.D.F. method, and the "separate strata" approach should be compared. For instance, one may wish to determine his "loss" by using the simpler "combined strata" approach rather than the C.D.F. method when proportional sample size allocation has not been used. Second, it is desirable to compare simple random with stratified simple random sampling.

Making assumptions comparable to those used in deriving the confidence coefficients, one may compare (1) simple random with stratified simple random sampling, and (2) the three confidence interval methods for stratified sampling.

1. Simple random sampling vs. the "combined" method

We first consider the comparison of confidence intervals obtained by using (a) simple random sampling, and (b) stratified simple random sampling with the combined method. For simplicity, we consider one-sided confidence intervals for $Y_{(t)}$: let r be such that $P\{Y_{(t)} \leq y_{(r)}\} \doteq \gamma$, in the simple random sampling method, and r' be such that $P\{Y_{(t)} \leq y_{(r')}\} \doteq \gamma$ in the combined method. Then,

$$P\{y_{(r')} \leq y_{(r)}\} = \sum_{v=r}^{N-(n-r)} P\{y_{(r')} \leq y_{(r)} | y_{(r)} = Y_{(v)}\} \, P\{y_{(r)} = Y_{(v)}\} \, . \tag{3.129}$$

The first term has been derived earlier in this Chapter in (3.8) ff. Also,

$$P\{y_{(r)} = Y_{(v)}\} = \binom{v-1}{r-1}\binom{N-v}{n-r}\Big/\binom{N}{n} \quad . \tag{3.130}$$

2. Simple random sampling vs. the C.D.F. method

Turning to the comparison of confidence intervals obtained by using (a) simple random sampling, and (b) stratified simple random sampling with the C.D.F. method, we have: let r and β be chosen so that $P\{Y_{(t)} \leq y_{(r)}\} \doteq \gamma$ and $P\{Y_{(t)} \leq y_U\} \doteq \gamma$. Then, to compare the two intervals (one-sided) one may use $P\{y_U \leq y_{(r)}\}$:

$$P\{y_U \leq y_{(r)}\} = \sum_{v=r}^{N-(n-r)} P\{y_U \leq y_{(r)} | y_{(r)} = Y_{(v)}\} P\{y_{(r)} = Y_{(v)}\} \quad (3.131)$$

where $P\{y_{(r)} = Y_{(v)}\}$ is given in (3.130) and $P\{y_U \leq Y_{(v)}\}$ may be obtained from (3.81).

3. <u>Simple random sampling vs. the Separate method</u>

Let r and r' be chosen so that $P\{Y_{(t)} \leq y_{(r)}\} \doteq \gamma$ and $P\{Y_{(t)} \leq y_{2(r')}\} \doteq \gamma$. Then

$$P\{y_{2(r')} \leq y_{(r)}\} = \sum_{v=r}^{N-(n-r)} P\{y_{2(r')} \leq y_{(r)} | y_{(r)} = Y_{(v)}\} P\{y_{(r)} = Y_{(v)}\} . \quad (3.132)$$

The latter term is given in (3.130), and

$$P\{y_{2(r')} \leq Y_{(v)}\} = \sum_j P\{y_{2(r')} \leq Y_{(v)} | Y_{(v)} = Y_{1(j)}\} P\{Y_{(v)} = Y_{1(j)}\}$$

$$+ \sum_{j^*} P\{y_{2(r')} \leq Y_{(v)} | Y_{(v)} = Y_{2(j^*)}\} P\{Y_{(v)} = Y_{2(j^*)}\} . \quad (3.133)$$

The latter terms correspond to $P\{B_j^1\}$ and $P\{B_{j^*}^2\}$ as given in (3.14) and (3.15) for "random" stratification and (3.41) and (3.42) for "$Y_{1(s)} < Y_{2(u)}$". $P\{y_{2(r')} \leq Y_{(v)} | Y_{(v)} = Y_{1(j)}\}$ is given in (3.117), and $P\{y_{2(r')} \leq Y_{(v)} | Y_{(v)} = Y_{2(j^*)}\}$ is given in (3.121).

4. Combined method vs. the C.D.F. method

Let $y_{(r)}$ be such that $P\{Y_{(t)} \leq y_{(r)}\} \doteq \gamma$ in the combined method, and β such that $P\{Y_{(t)} \leq y_U\} \doteq \gamma$ in the C.D.F. method. We then have

$$P\{y_U \leq y_{(r)}\} = \sum_v P\{y_U \leq Y_{(v)} | y_{(r)} = Y_{(v)}\} \, P\{y_{(r)} = Y_{(v)}\} \quad . \qquad (3.134)$$

Now,

$$P\{y_{(r)} = Y_{(v)}\} = P\{y_{(r)} \leq Y_{(v)}\} - P\{y_{(r)} \leq Y_{(v-1)}\} \,, \qquad (3.135)$$

which can be found in Section B of this chapter.

$P\{y_U \leq Y_{(v)}\}$ is given in (3.81).

5. C.D.F. technique vs. separate strata method

Let $y_{2(r)}$ be such that $P\{Y_{(t)} \leq y_{2(r)}\} \doteq \gamma$, and β such that $P\{Y_{(t)} \leq y_U\} \doteq \gamma$. Then

$$P\{y_U \leq y_{2(r)}\} = \sum_v P\{y_U \leq Y_{(v)} | y_{2(r)} = Y_{(v)}\} \, P\{y_{2(r)} = Y_{(v)}\} \,. \quad (3.136)$$

The first term in the summation is contained in (3.81), and

$$P\{y_{2(r)} = Y_{(v)}\} = \sum_{p=r}^{N_2} P\{y_{2(r)} = Y_{2(p)} | Y_{2(p)} = Y_{(v)}\} \; P\{Y_{2(p)} = Y_{(v)}\}$$

$$= \sum_p \binom{p-1}{r-1}\binom{N_2-p}{n_2-r} \Big/ \binom{N_2}{n_2} \times P\{B^2_{p*}\} \,, \qquad (3.137)$$

where $P\{B^2_{p*}\}$ is found in Section B of this chapter.

6. Comparison of SRS confidence intervals with the separate strata technique - 2 sided

We wish to compare the SRS confidence interval $[y_{(k)}, y_{(r)}]$ to the separate strata interval $\{[y_{1(k')}, y_{2(r')}]$ or $[y_{2(r')}, y_{1(k')}]\}$

where k, r, k', and r' are picked so that the confidence coefficients of the two intervals are approximately the same.

In particular we want to find

$$P\{y_{(k)} \leq y_{1(k')} \leq y_{(r)} \cap y_{(k)} \leq y_{2(r')} \leq y_{(r)}\}$$

$$= P\{y_{(k)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq y_{(r)}\} \quad . \tag{3.138}$$

This is interpreted as meaning the probability if the separate strata confidence interval is included inside the SRS confidence interval.

We have

$$P\{y_{(k)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq y_{(r)}\}$$

$$= \sum_b \sum_v P\{y_{(k)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq y_{(r)} \mid y_{(k)} = Y_{(b)} \cap y_{(r)} = Y_{(v)}\}$$

$$\times P\{y_{(k)} = Y_{(b)} \cap y_{(r)} = Y_{(v)}\} \qquad [b < v]$$

$$= \sum_b \sum_v P\{Y_{(b)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{(v)}\} P\{y_{(k)} = Y_{(b)} \cap y_{(r)} = Y_{(v)}\} \quad . \tag{3.139}$$

Looking at the terms individually, we have

$$P\{y_{(k)} = Y_{(b)} \cap y_{(r)} = Y_{(v)}\} = \binom{b-1}{k-1}\binom{v-b-1}{r-k-1}\binom{N-v}{n-r} \bigg/ \binom{N}{n} \quad . \tag{3.140}$$

$$
\begin{aligned}
P\{Y_{(b)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{(v)}\}
&= \sum_{j} \sum_{j'} P\{Y_{1(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{1(j')}\} \\
&\quad \times P\{Y_{1(j)} = Y_{(b)} \cap Y_{1(j')} = Y_{(v)}\} \\
&+ \sum_{j} \sum_{j'} P\{Y_{1(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{2(j')}\} \\
&\quad \times P\{Y_{1(j)} = Y_{(b)} \cap Y_{2(j')} = Y_{(v)}\} \\
&+ \sum_{j} \sum_{j'} P\{Y_{2(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{1(j')}\} \\
&\quad \times P\{Y_{2(j)} = Y_{(b)} \cap Y_{1(j')} = Y_{(v)}\} \\
&+ \sum_{j} \sum_{j'} P\{Y_{2(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{2(j')}\} \\
&\quad \times P\{Y_{2(j)} = Y_{(b)} \cap Y_{2(j')} = Y_{(v)}\} \ .
\end{aligned}
\tag{3.141}
$$

In this expression, the formulas for

$$P\{Y_{i(j)} = Y_{(b)} \cap Y_{i'(j')} = Y_{(v)}\} \ , \quad i, i' = 1, 2,$$

are derived in the Section C.3. under the assumption $Y_{1(s)} < Y_{2(u)}$.

$$P\{Y_{1(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{1(j')}\}$$

$$= P\{y_{1(k')} \in [Y_{1(j)}, Y_{1(j')}]\}$$

$$\times P\{y_{2(r')} \in [Y_{2(b-j+1)}, Y_{2(v-j')}]\} \quad . \tag{3.142}$$

Now,

$$P\{y_{1(k')} \in [Y_{1(j)}, Y_{1(j')}]\}$$

$$= P\{y_{1(k')} \leq Y_{1(j')}\} - P\{y_{1(k')} \leq Y_{1(j-1)}\} \tag{3.143}$$

and

$$P\{y_{2(r')} \in [Y_{2(b-j+1)}, Y_{2(v-j')}]\}$$

$$= P\{y_{2(r')} \leq Y_{2(v-j')}\} - P\{y_{2(r')} \leq Y_{2(b-j)}\} \quad . \tag{3.144}$$

Similarly,

$$P\{Y_{1(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{2(j')}\}$$

$$= [P\{y_{1(k')} \leq Y_{1(v-j')}\} - P\{y_{1(k')} \leq Y_{1(j-i)}\}]$$

$$\times [P\{y_{2(r')} \leq Y_{2(j')}\} - P\{y_{2(r')} \leq Y_{2(b-j)}\}] \quad , \tag{3.145}$$

$$P\{Y_{2(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{1(j')}\}$$

$$= [P\{y_{1(k')} \leq Y_{1(j')}\} - P\{y_{1(k')} \leq Y_{1(b-j)}\}]$$

$$\times [P\{y_{2(r')} \leq Y_{2(v-j')}\} - P\{y_{2(r')} \leq Y_{2(j-1)}\}] , \qquad (3.146)$$

and

$$P\{Y_{2(j)} \leq \begin{bmatrix} y_{1(k')} \\ y_{2(r')} \end{bmatrix} \leq Y_{2(j')}\}$$

$$= [P\{y_{1(k')} \leq Y_{1(v-j')}\} - P\{y_{1(k')} \leq Y_{1(b-j)}\}]$$

$$\times [P\{y_{2(r')} \leq Y_{2(j')}\} - P\{y_{2(r')} \leq Y_{2(j-1)}\}] . \qquad (3.147)$$

The expressions for each of the component parts of the form $P\{y_{i(a)} \leq Y_{i(c)}\}$ are given in Chapter II.

Comparing alternative methods by evaluating expressions such as $P\{y_U \leq y_{(r)}\}$ in (3.131) seems to be a formidable numerical task: for given values of t and γ, it is necessary to determine those values of r and β satisfying $P\{Y_{(t)} \leq y_{(r)}\} \doteq \gamma$ and $P\{Y_{(t)} \leq y_U\} \doteq \gamma$. Then for each such (r,β), (3.131) must be evaluated.

A common way of choosing among alternative (two-sided) confidence interval procedures is to compare their expected lengths. However, this requires that for each stratum the distribution of Y must be specified. Further, even for random sampling, simple (exact) expressions for

$E(y_{(r)})$ etc., are available only for a few distributions. Thus, it is even difficult to give general rules for choosing (in simple random sampling) among those values of k and r for which $P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} \doteq \gamma$ where γ and t are specified. Hence, we turn to Monte Carlo studies to compare the methods.

7. Monte Carlo results--two strata

Several Monte Carlo studies were performed to compare the techniques we have suggested. In Monte Carlo Study I, the fixed parameters were $N = 20$, $n = 6$, and $t = 8$. The population Π_{20} was $Y_{(i)} = i$, $i = 1,\ldots, 20$. For the Simple Random Sample technique, discussed in Chapter II, a S.R.S. of size 6 was selected for Π_{20}, and then for pairs $(k,r) = (1,3), (1,4), (1,5), (2,5), (2,6)$, and $(3,6)$, the proportion of times the random interval covered $Y_{(t)} = Y_{(8)} = 8$ was calculated, along with the mean interval length and the standard error of the mean length. Table 18 gives the results of this study for 1000 replications.

For the remaining techniques, stratification of the population was required. In Monte Carlo Study II, this was achieved by first randomly stratifying Π_{20} in Study I into Π^1_{10} and Π^2_{10} for Strata I and II, respectively, and then checking if the stratification met the specification $Y_{1(s)} < Y_{2(u)}$, s and u predetermined. If not, another random stratification was made, and the specification was checked again. Upon achieving the appropriate specification, simple random samples of sizes $n_1 = 4$ and $n_2 = 2$ were drawn from Π^1_{10} and Π^2_{10}, respectively. The three techniques of the "Combined", "Separate", and "C.D.F." methods were then applied, in each case recording the coverage or non-coverage

Table 18. Monte Carlo Study I - S.R.S. results

$Y_{(i)} = i,\ t = 8$

k	r	p^a	p_e^b	l^c	s.d.^{d}
1	3	.642	.640	6.18	.090
1	4	.904	.898	9.09	.092
1	5	.971	.969	12.10	.094
2	5	.803	.805	8.97	.095
2	6	.808	.808	11.97	.095
3	6	.460	.455	8.99	.095

[a]proportion of coverages

[b]theoretical proportion of coverages

[c]mean length

[d]standard error of mean length $= \sqrt{\dfrac{\sum_{i=1}^{1000} (\ell_i - \bar{\ell})^2}{999(1000)}}$

of $Y_{(t)} = Y_{(8)} = 8$, and the length of the interval. For each stratification scheme the latter process was repeated five times, and then a new stratification was made. The entire process was repeated 400 times for each parameter pair for the three techniques. The parameters used for (s,u) were $(1,10)$ (essentially random stratification), $(5,5)$, and $(7,4)$. Table 19 gives the results of this study.

In Monte Carlo Study III, we employed as data the population of the twenty largest cities in the United States in the 1970 census. These were stratified into two strata of sizes ten each by the 1940 census figures. In Table 20 we list these cities and their populations, as

Table 19. Monte Carlo Study II $Y_{(i)} \propto i$, t = 8

				Combined Method						Separate Method				C.D.F. Method				
s	u	k	r	p^a	p_e^b	ℓ^c	s.d.[d]	k_1	r_2	p	p_e	ℓ	s.d.	α	β	p	ℓ	s.d.
1	10	1	3	.648	.640	6.04	.061	1	1	.492	.490	5.16	.084	.26	.51	.721	7.44	.073
		1	4	.904	.898	9.04	.067	1	2	.836	.837	10.20	.111	.26	.63	.860	9.75	.077
		1	5	.971	.969	12.00	.068	2	1	.637	.639	5.34	.076	.26	.76	.896	11.68	.077
		2	4	.730	.735	5.96	.060	2	2	.586	.589	7.04	.094	.26	.88	.918	13.84	.072
		2	5	.806	.805	8.93	.068	3	1	.681	.679	7.15	.093	.38	.76	.745	9.64	.077
		2	6	.809	.813	11.92	.068	3	2	.286	.283	5.40	.077	.38	.88	.752	11.49	.075
		3	6	.445	.455	8.95	.066	4	1	.656	.658	10.19	.109					
5	5	1	3	.598	.594	5.70	.060	1	1	.526	.534	5.75	.090	.26	.51	.727	7.30	.071
		1	4	.882	.885	8.64	.067	1	2	.916	.912	11.56	.100	.26	.63	.856	9.81	.077
		1	5	.973	.975	11.86	.068	2	1	.606	.608	4.97	.072	.26	.76	.893	11.55	.079
		2	4	.753	.757	5.94	.061	2	2	.704	.697	8.34	.096	.26	.88	.910	13.39	.072
		2	5	.854	.847	9.07	.068	3	1	.606	.608	6.06	.088	.38	.76	.745	9.64	.078
		2	6	.864	.858	12.18	.066	3	2	.307	.309	5.78	.078	.38	.99	.755	11.33	.072
		3	6	.530	.521	9.31	.068	4	1	.560	.561	8.72	.109	.38	.88			
7	4	1	3	.541	.543	5.33	.055	1	1	.630	.636	7.12	.097	.26	.51	.745	7.19	.069
		1	4	.880	.872	8.47	.064	1	2	.960	.961	12.96	.087	.26	.63	.857	9.32	.072
		1	5	.978	.981	11.40	.069	2	1	.650	.656	5.64	.076	.26	.76	.921	11.72	.073
		2	4	.782	.784	5.71	.059	2	2	.827	.823	10.06	.089	.26	.88	.925	13.04	.068
		2	5	.895	.893	9.07	.066	3	1	.582	.580	5.09	.072	.38	.76	.722	9.23	.076
		2	6	.905	.906	12.36	.066	3	2	.430	.420	7.02	.083	.38	.88	.734	10.96	.073
		3	6	.592	.595	9.50	.066	4	1	.462	.454	7.16	.104					

[a]proportion of coverages

[b]theoretical proportion of coverages

[c]mean length of interval

[d]standard error of mean length

stratified. One thousand replications of the following procedures were performed; the results are found in Table 21. A simple random sample of size 6 was drawn from the entire (unstratified) population, and for given (k,r), the coverage or non-coverage of $Y_{(8)}$ noted, as well as the length of the confidence interval. Then simple random samples of sizes 4 and 2 were drawn from Strata I and II, respectively. For fixed (k,r) (for the combined method), (k_1,r_2) (for the separate method), and (α,β) (for the C.D.F. method), the coverages and lengths were recorded.

Table 20. Monte Carlo Study III - data

City	Stratum I	Stratum II
Phoenix	580[a]	
New Orleans	586	
St. Louis		608
Memphis	621	
Boston		628
San Antonio	650	
San Diego	676	
San Francisco	704	
Milwaukee	710	
Cleveland		739
Indianapolis	743	
Washington, D.C.		764
Dallas	836	
Baltimore		895
Houston	1213	
Detroit		1493
Philadelphia		1927
Los Angeles		2782
Chicago		3323
New York		[illegible]

[a] populations given in thousands

Table 21. Monte Carlo Study III

City Data. $t = 8$

S.R.S. Method					Combined Method				
k	r	p[a]	l[b]	s.d.[c]	k	r	p	l	s.d.
1	4	.892	342.10	12.73	1	4	.871	184.96	4.18
1	5	.971	958.32	26.24	1	5	.982	569.37	19.13
2	5	.807	907.70	25.96	2	5	.923	534.37	18.98
2	6	.815	3336.63	82.09	2	6	.936	2572.94	76.43

[a] proportion of coverages

[b] mean length of interval

[c] standard error of mean length

Table 21. (continued)

Separate Method					C.D.F. Method				
k_1	r_2	p	l	s.d.	α	β	p	l	s.d.
1	1	.620	415.95	20.75	.26	.51	.751	419.21	19.88
1	2	.974	2573.22	77.79	.26	.63	.888	532.55	18.82
2	1	.657	393.12	20.36	.26	.76	.931	2592.32	78.11
2	2	.879	2516.98	77.78	.26	.88	.941	2569.12	75.97
					.38	.76	.699	2500.00	77.21
					.38	.88	.718	2588.81	76.86

Comparing the Monte Carlo studies, several points are obvious.

1. The $(s,u) = (1,10)$ stratification is essentially random stratification, and the S.R.S. and combined methods are, of course, equivalent.

2. In the case of the C.D.F. method, the lowest α that can be used is $\alpha > .25$, since $\alpha > \alpha_0 = \max\{j_1, j_2\}$. Furthermore, because of the sizes of the possible jumps in our study (.25 or .125), the only "critical" points we need to look at are .25, .375, .50, .625, .75, and .875; hence the limits for α and β as they appear in Tables 19 and 21.

3. In Monte Carlo Study III, with the highly skewed distribution, the average length "exploded" when using confidence interval limits that included "high" data values. Hence, in using any of the techniques with highly skewed data, care should be used to pick, if possible, intervals that will not include unnecessarily large (or small) data values.

4. In comparing the S.R.S. and Combined methods in Monte Carlo Study III, it is seen that large gains in probability of coverage and smaller interval length can be made using the Combined method.

5. In the case of a "uniform" distribution, the interval lengths for the Separate method and the "C.D.F." method are longer than for the combined method, holding confidence coefficients approximately equal.

6. In the Skewed distribution study, the combined method was highly superior to the separate and the C.D.F. method in terms of interval length.

7. In considering the Woodruff method (see Section C.4), for the city data we have $\sigma_p = .187$. Hence, $2\sigma_p$ confidence limits, in terms of α and β, would be $.4 \pm 2\,(.187) = (.026, .774)$. Because of the .25 lower bound for α, the best approximation we have is (.26, .76), which yielded a coverage proportion of .931, indicating, in this particular example, that the Woodruff technique may be quite accurate for even small samples and populations.

8. Good agreement between the proportion of coverages by the Monte Carlo method and the theoretical proportion is found throughout the tables.

9. Another comparison, obtainable from Tables 19 and 20, can be made by choosing a desired confidence coefficient and tabulating the intervals which most nearly achieve this number, along with their expected lengths. In Table 22 we do this for 65%, 90%, and 95% confidence levels.

10. For the city data (Table 20), we have $Y_{1(8)} < Y_{2(4)}$. Comparing the probabilities in Table 21 to those in Tables 18 and 19 for (s,u) = (7,4), it is noted that they are in fairly good agreement. This would indicate that making "slight" errors in estimating the (s,u) appropriate for a given situation would not seriously effect the resulting outcomes, and also that the actual confidence coefficients for $Y_{1(s)} < Y_{2(u)}$ apply to specific populations--the city size data is only one possibility out of those permutations represented by $Y_{1(8)} < Y_{2(4)}$.

Table 22. Comparison of intervals for fixed confidence coefficients

Method	Stratification Specification		Level: 65%			Level: 90%			Level: 95%		
	s	u	Limits	p	Length	Limits	p	Length	Limits	p	Length
SRS	-	-									
Combined	1	10	1-3	.640	6.18	1-4	.898	9.09	1-5	.969	12.10
Combined	5	5	n.a.[a]			1-4	.885	8.64	1-5	.975	11.86
Combined	7	4	n.a.			2-5	.893	9.07	1-5	.981	11.40
Separate	1	10	2-1	.637	5.34	n.a.			n.a.		
Separate	5	5	n.a.			1-2	.912	11.56	n.a.		
Separate	7	4	2-1	.656	5.64	n.a.			1-2	.961	12.96
C.D.F.	1	10	.26-.51	.721	7.44	.26-.76	.896	11.68	n.a.		
C.D.F.	5	5	.26-.51	.727	7.30	.26-.76	.893	11.55	n.a.		
C.D.F.	7	4	.38-.76	.722	9.23	.26-.76	.921	11.72	n.a.		

[a]none applicable

F. Joint Confidence Intervals--2 Strata

Our notation for the population and sample is as defined earlier. We wish to find a joint confidence interval for $Y_{(t)}$ and $Y_{(t')}$ $(t < t')$ of the form

$$y_{(k)} \leq Y_{(t)} \leq y_{(r)} \cap y_{(k')} \leq Y_{(t')} \leq y_{(r')}$$

where $k \leq k'$, $r \leq r'$, $k < r$, $k' < r'$.

Now

$$\begin{aligned} P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)} \cap y_{(k')} \leq Y_{(t')} \leq y_{(r')}\} \\ &= P\{y_{(k)} \leq Y_{(t)} \cap y_{(k')} \leq Y_{(t')}\} \\ &\quad - P\{y_{(r)} < Y_{(t)}\} \\ &\quad - P\{y_{(r')} < Y_{(t')}\} \\ &\quad + P\{y_{(r)} < Y_{(t)} \cap y_{(r')} < Y_{(t')}\} \\ &= P\{A\} - P\{B\} - P\{C\} + P\{D\} . \end{aligned} \tag{3.148}$$

We see first that

$$P\{B\} = P\{y_{(r)} \leq Y_{(t-1)}\}$$

and

$$P\{C\} = P\{y_{(r')} \leq Y_{(t'-1)}\} \quad .$$

The formulas for computing these are given in Section B.

$P\{D\}$ can be found from $P\{A\}$ by replacing

$$\begin{array}{lll} k & \text{by} & r \\ k' & \text{by} & r' \\ t & \text{by} & t-1 \\ t' & \text{by} & t'-1 \quad . \end{array}$$

We now turn to the evaluation of $P\{A\}$. We break up our work into four cases.

Case 1: $Y_{(t)} = Y_{1(j)} \cap Y_{(t')} = Y_{1(j')}$

Case 2: $Y_{(t)} = Y_{1(j)} \cap Y_{(t')} = Y_{2(j')}$

Case 3: $Y_{(t)} = Y_{2(j)} \cap Y_{(t')} = Y_{1(j')}$

Case 4: $Y_{(t)} = Y_{2(j)} \cap Y_{(t')} = Y_{2(j')}$

Hence,

$$P\{A\} = \sum_{i=1}^{t} P\{y_{(k)} \leq Y_{(t)} \cap y_{(k')} \leq Y_{(t')} | \text{Case } i\} \times P\{\text{Case } i\} \ . \tag{3.149}$$

Turning to the first term, we have

$$P\{y_{(k)} \leq Y_{(t)} \cap y_{(k')} \leq Y_{(t')} | \text{Case } 1\}$$

$$= \sum_{j=1}^{t} \sum_{j'=j+1}^{t'} \sum_{i=0}^{n-k} \sum_{i'=i+k-k'}^{n-k-k'+i} P\{\text{exactly } (k+i) \text{ observations} \leq Y_{(t)} \text{ and exactly } (k'+i')-(k+i) \text{ observations in } [Y_{(t+1)}, Y_{(t')}] \, | Y_{(t)}=Y_{1(j)} \cap Y_{(t')}=Y_{1(j')}\}$$

$$= \sum \sum \sum \sum P\{G_1\} \ . \tag{3.150}$$

Similar expressions can be given for the other three cases, call them $\sum\sum\sum\sum P\{G_i\}$, $i = 2, 3, 4$.

<u>Case 1</u>. $$P\{G_1\} = \sum_{\ell}\sum_{\ell'} \binom{j}{\ell}\binom{t-j}{k+i-\ell}\binom{j'-j}{\ell'}\binom{(t'-t)-(j'-j)}{(k'+i')-(k+i)-\ell'} \times \binom{N_1-j'}{n_1-(\ell+\ell')}\binom{N_2-(t'-j')}{n_2-(k'+i')+(\ell+\ell')}\Bigg/\binom{N_1}{n_1}\binom{N_2}{n_2} \quad . \tag{3.151}$$

In this case, ℓ represents the number of sample observations in Stratum I less than or equal to $Y_{(t)}$; ℓ' represents the number of sample observations in Stratum I in the range $[Y_{(t+1)}, Y_{(t')}]$.

<u>Case 2</u>. $$P\{G_2\} = \sum_{\ell}\sum_{\ell'} \binom{j}{\ell}\binom{t-j}{k+i-\ell}\binom{j'-(t-j)}{\ell'}\binom{t'-j'-j}{(k'+i')-(k+i)-\ell'} \times \binom{N_2-j'}{n_2-(k+i-\ell)-\ell'}\binom{N_1-(t'-j')}{n_1-\ell-[(k'+i')-(k+i)-\ell']}\Bigg/\binom{N_1}{n_1}\binom{N_2}{n_2} \quad . \tag{3.152}$$

Here, ℓ is as in Case 1, ℓ' represents the number of sample observations in Stratum II in the range $[Y_{(t+1)}, Y_{(t')}]$.

<u>Case 3</u>. $$P\{G_3\} = \sum_{\ell}\sum_{\ell'} \binom{t-j}{k+i-\ell}\binom{j}{\ell}\binom{j'-(t-j)}{\ell'}\binom{t'-j-j'}{(k'+i')-(k+i)-\ell'} \times \binom{N_1-j'}{n_1-(k+i-\ell)-\ell'}\binom{N_2-(t'-j')}{n_2-[(k'+i')-(k+i)]+\ell'-\ell}\Bigg/\binom{N_1}{n_1}\binom{N_2}{n_2} \quad . \tag{3.153}$$

In Case 3, ℓ is the number of sample observation in Stratum II in the range $[Y_{(t+1)}, Y_{(t')}]$, ℓ' is as in Case I.

<u>Case 4</u>. $$P\{G_4\} = \sum_{\ell}\sum_{\ell'} \binom{t-j}{k+i-\ell}\binom{j}{\ell}\binom{(t'-j')-(t-j)}{(k'+i')-(k+i)-\ell'}\binom{j'-j}{\ell'}$$
$$\times \binom{N_1-(t'-j')}{n_1-(k'+i')+\ell+\ell'}\binom{N_2-j'}{n_2-(\ell+\ell')}\Bigg/\binom{N_1}{n_1}\binom{N_2}{n_2} \quad . \tag{3.154}$$

Here, ℓ is as in Case 3, ℓ' as in Case 2.

In Section C we derived the expressions for $P\{Y_{(t)} = Y_{1(j)} \cap Y_{(t')} = Y_{1(j')}\}$, etc., under the assumption of $Y_{1(s)} < Y_{2(u)}$. Substituting these expressions into (3.149) yields $P\{A\}$.

IV. CONFIDENCE INTERVALS WITH THREE STRATA

In this chapter we extend the results of the previous chapter to three strata.

A. Definitions and Notations

Let π_N be a population of size N whose elements have distinct Y-values $Y_{(1)} < Y_{(2)} < \ldots < Y_{(N)}$. Assume that π_N has been divided into three strata of sizes N_1, N_2 and N_3, where $N_1 + N_2 + N_3 = N$. Denote the (ordered) Y-values associated with the elements of Stratum h by

$$Y_{h(1)} < Y_{h(2)} < \ldots < Y_{h(N_h)} \ , \quad h = 1, 2, 3.$$

Stratified random sampling is performed, the sample sizes within the strata being n_1, n_2, and n_3, respectively. We denote the Y-values (ordered) as

$$y_{h(1)} < y_{h(2)} < \ldots < y_{h(n_h)} \ , \quad h = 1, 2, 3.$$

Combining and ordering the sample values yields the combined sample

$$y_{(1)} < y_{(2)} < \ldots < y_{(n)} \ , \quad n = n_1 + n_2 + n_3 \ .$$

B. The Combined Method

We first investigate confidence intervals for $Y_{(t)}$ of the form $[y_{(k)}, y_{(r)}]$, where $1 \leq k < r \leq n$. As earlier,

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} = P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} \leq Y_{(t-1)}\} . \qquad (4.1)$$

We confine our attention to the term $P\{y_{(k)} \leq Y_{(t)}\}$.

1. <u>General derivation of $P\{y_{(k)} \leq Y_{(t)}\}$</u>

Let A_i be the event "exactly i observations in the combined sample have values less than or equal to $Y_{(t)}$"; $D_{\ell(j)}$ the event "$Y_{\ell(j)} = Y_{(t)}$", $\ell = 1, 2, 3$; and $E_{\ell(p)}$ the event "$Y_{\ell(p)} < Y_{(t)} < Y_{\ell(p+1)}$", $\ell = 1, 2, 3$.

Then

$$\begin{aligned}
P\{y_{(k)} \leq Y_{(t)}\} &= P\{\text{at least } k \text{ sample observations have values less than or equal to } Y_{(t)}\} \\
&= \sum_{i=k}^{\{{t \atop n}\}} \{A_i\} \\
&= \sum_i \sum_{j=\max[1,t-(N-N_1)]}^{\min[t,N_1]} \sum_{p=\max[0,t-j-N_3]}^{\min[t-j,N_2]} P\{A_i \cap D_{1(j)} \cap E_{2(p)}\} \\
&+ \sum_i \sum_{j=\max[1,t-(N-N_2)]}^{\min[t,N_2]} \sum_{p=\max[0,t-j-N_3]}^{\min[t-j,N_1]} P\{A_i \cap D_{2(j)} \cap E_{1(p)}\} \\
&+ \sum_i \sum_{j=\max[1,t-(N-N_3)]}^{\min[t,N_3]} \sum_{p=\max[0,t-j-N_2]}^{\min[t-j,N_1]} P\{A_i \cap D_{3(j)} \cap E_{1(p)}\} . \qquad (4.2)
\end{aligned}$$

Furthermore,

$$P\{A_i \cap D_{1(j)} \cap E_{2(p)}\} = P\{A_i | D_{1(j)} \cap E_{2(p)}\} P\{D_{1(j)} \cap E_{2(p)}\} \qquad (4.3)$$

and

$$P\{A_i | D_{1(j)} \cap E_{2(p)}\} = \sum_{m_1+m_2=0}^{i}{}^{*} \binom{j}{m_1}\binom{N_1-j}{n_1-m_1}\binom{p}{m_2}$$

$$\times \binom{N_2-p}{n_2-m_2}\binom{t-j-p}{i-(m_1+m_2)}\binom{N_3-(t-j-p)}{n_3-(i-(m_1+m_2))}$$

$$\times \left[\binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N_3}{n_3}\right]^{-1} \tag{4.4}$$

where $\sum_{m_1+m_2=0}^{i}{}^{*}$ indicates the summation is over all non-negative integers m_1 and m_2 such that $0 \le m_1+m_2 \le i$. Also, we have

$$\max \begin{Bmatrix} 0 \\ n_1+j-N_1 \end{Bmatrix} \le m_1 \le \min \begin{Bmatrix} j \\ n_1 \end{Bmatrix} \tag{4.5}$$

$$\max \begin{Bmatrix} 0 \\ n_2+p-N_2 \end{Bmatrix} \le m_2 \le \min \begin{Bmatrix} n_2 \\ p \end{Bmatrix} \tag{4.6}$$

$$\max \begin{Bmatrix} 0 \\ n_3+(t-j-p)-N_3 \end{Bmatrix} \le (i-m_1-m_2) \le \min \begin{Bmatrix} t-j-p \\ n_3 \end{Bmatrix} . \tag{4.7}$$

$P\{D_{1(j)} \cap E_{2(p)}\}$ depends on the stratification assumptions and is derived in Subsections 2 and 3.

Similarly,

$$P\{A_i \cap D_{2(j)} \cap E_{1(p)}\} = P\{A_i | D_{2(j)} \cap E_{1(p)}\} \, P\{E_{1(p)} \cap D_{2(j)}\} \quad , \tag{4.8}$$

and

$$P\{A_i | D_{2(j)} \cap E_{1(p)}\} = \sum_{m_1+m_2=0}^{i}{}^{*} \binom{j}{m_2}\binom{N_2-j}{n_2-m_2}\binom{p}{m_1}\binom{N_1-p}{n_1-m_1}$$

$$\times \binom{t-j-p}{i-(m_1+m_2)}\binom{N_3-(t-j-p)}{n_3-(i-(m_1+m_2))}$$

$$\times \left[\binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N_3}{n_3}\right]^{-1} \quad . \tag{4.9}$$

Also,

$$P\{A_i \cap D_{3(j)} \cap E_{1(p)}\} = P\{A_i | D_{3(j)} \cap E_{1(p)}\} \; P\{E_{1(p)} \cap D_{3(j)}\} \quad , \tag{4.10}$$

and

$$P\{A_i | D_{3(j)} \cap E_{1(p)}\} = \sum_{m_3+m_1=0}^{i}{}^{*} \binom{j}{m_3}\binom{N_3-j}{n_3-m_3}\binom{p}{m_1}\binom{N_1-p}{n_1-m_1}$$

$$\times \binom{t-j-p}{i-(m_1+m_3)}\binom{N_2-(t-j-p)}{n_2-(i-(m_1+m_3))}$$

$$\times \left[\binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N_3}{n_3}\right]^{-1} \quad . \tag{4.11}$$

2. Random stratification

We say stratification is random if each possible stratification is equally likely to occur. Under this assumption, we have the following, using the notation of the previous subsection:

$$P\{D_{1(j)} \cap E_{2(p)}\} = P\{E_{2(p)} | D_{1(j)}\} \, P\{D_{1(j)}\}$$

$$= \frac{\binom{t-j}{p}\binom{N-N_1-(t-j)}{N_2-p}\binom{t-1}{j-1}\binom{N-t}{N_1-j}}{\binom{N-N_1}{N_2}\binom{N}{N_1}}, \tag{4.12}$$

$$P\{D_{2(j)} \cap E_{1(p)}\} = P\{E_{1(p)} | D_{2(j)}\} \, P\{D_{2(j)}\}$$

$$= \frac{\binom{t-j}{p}\binom{N-N_2-(t-j)}{N_1-p}\binom{t-1}{j-1}\binom{N-t}{N_2-j}}{\binom{N-N_2}{N_1}\binom{N}{N_2}}, \tag{4.13}$$

and

$$P\{D_{3(j)} \cap E_{1(p)}\} = P\{E_{1(p)} | D_{3(j)}\} \, P\{D_{3(j)}\}$$

$$= \frac{\binom{t-j}{p}\binom{N-N_3-(t-j)}{N_1-p}\binom{t-1}{j-1}\binom{N-t}{N_3-j}}{\binom{N-N_3}{N_1}\binom{N}{N_3}}. \tag{4.14}$$

Theorem 3.1 tells us that the "combined" method with random stratification is equivalent to simple random sampling from the non-stratified population.

3. $\underline{Y_{1(s)} < Y_{2(u)} < Y_{3(v)}}$: "Ordered" Stratification

We assume that the strata are such that $Y_{1(s)} < Y_{2(u)} < Y_{3(v)}$, analogous to subsection B.5 of Chapter III. We wish to find expressions for $P\{D_{1(j)} \cap E_{2(p)}\}$, $P\{D_{2(j)} \cap E_{1(p)}\}$, and $P\{D_{3(j)} \cap E_{1(p)}\}$.

Our approach is to consider the event

$$\left\{\begin{array}{lll} Y_{1(s+r_1)} & < Y_{2(u)} < & Y_{1(s+r_1+1)} \\ Y_{1(s+r_1+r_2)} & < Y_{3(v)} < & Y_{1(s+r_1+r_2+1)} \\ Y_{2(u+r_3)} & < Y_{3(v)} < & Y_{2(u+r_3+1)} \\ Y_{2(u)} & = Y_{(r_4)} & \\ Y_{3(v)} & = Y_{(q)} & \end{array}\right\} . \tag{4.15}$$

The last expression is equivalent to

$$q = s + r_1 + r_2 + u + r_3 + v \quad . \tag{4.16}$$

These events are illustrated in Figure 7.

By $P\{\cdot \mid \underline{R} = \underline{r} \cap (s,u,v)\}$ we mean the probability of event $\{\cdot\}$, conditional on a specific configuration as given by (4.15) and $Y_1(s) < Y_{2(u)} < Y_{3(v)}$. By $\underline{r}$ we mean the vector $\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix}$. When we write $\sum_{\underline{r}}^{*}$ we mean the summation over all components of $\underline{r}$ which are consistent with the logic of the situation.

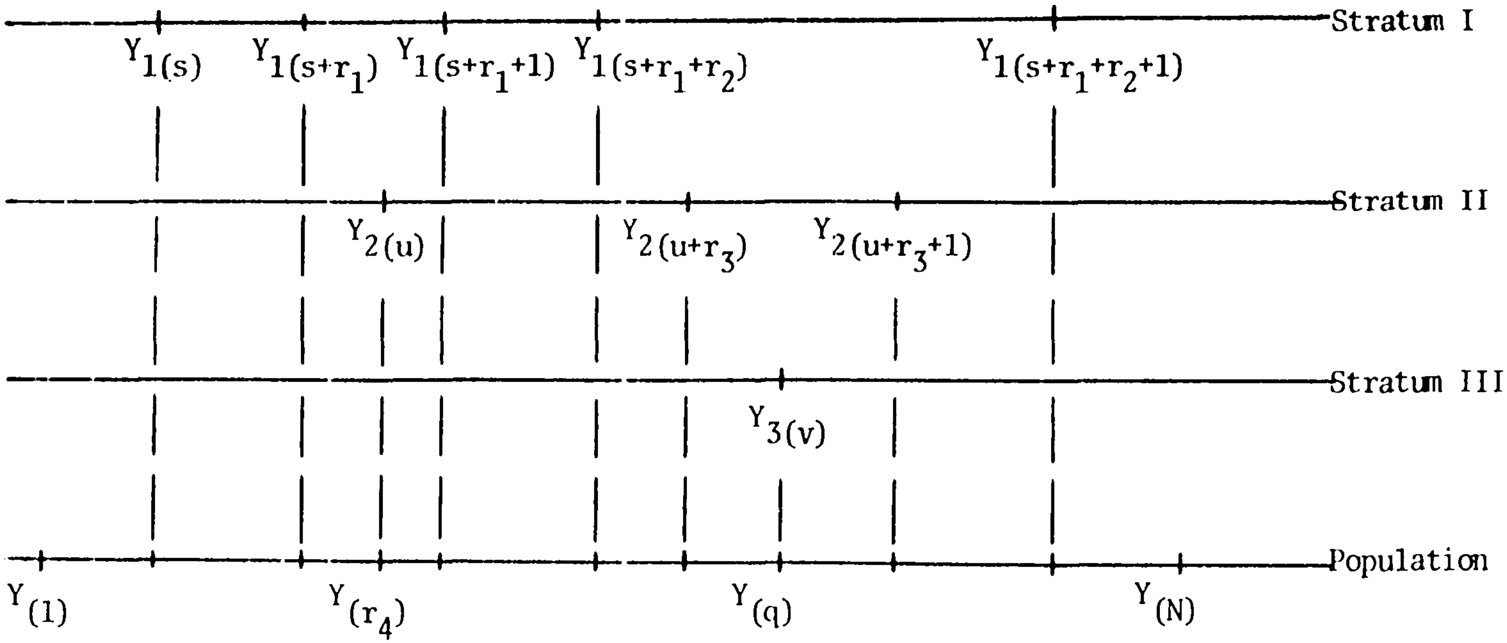

Figure 7. Stratification notation.

Then,

$$P\{D_{1(j)} \cap E_{2(p)} | (s,u,v)\}$$

$$= \sum_{\underline{r}}^{*} P\{D_{1(j)} \cap E_{2(p)} | \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{\underline{R} = \underline{r} | (s,u,v)\}$$

$$= \sum_{\underline{r}}^{*} P\{E_{2(p)} | D_{1(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{D_{1(j)} | \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{\underline{R} = \underline{r} | (s,u,v)\} \quad . \tag{4.18}$$

Similarly,

$$P\{D_{2(j)} \cap E_{1(p)} | (s,u,v)\}$$

$$= \sum_{\underline{r}}^{*} P\{E_{1(p)} | D_{2(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{D_{2(j)} | \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{\underline{R} = \underline{r} | (s,u,v)\} \tag{4.19}$$

and

$$P\{D_{3(j)} \cap E_{1(p)} | (s,u,v)\}$$

$$= \sum_{\underline{r}}^{*} P\{E_{1(p)} | D_{3(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{D_{3(j)} | \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$\times P\{\underline{R} = \underline{r} | (s,u,v)\} \quad . \tag{4.20}$$

We first consider $P\{\underline{R} = \underline{r} \mid (s,u,v)\}$. For fixed s, u, v, r_1, r_2, r_3, r_4, N_1, N_2 and N_3, we count the number of possible arrangements of the population into the three strata.

For Stratum I, we have

$$\binom{r_4-1}{s+r_1}\binom{q-r_4-1}{r_2}\binom{N-q}{N_1-(s+r_1+r_2)} \tag{4.21}$$

possible arrangements.

Then, for Stratum II, we have

$$\binom{r_4-(s+r_1)-1}{u-1}\binom{q-r_4-1-r_2}{r_3}\binom{N-q-(N_1-(s+r_1+r_2))}{N_2-(u+r_3)} \tag{4.22}$$

possible arrangements.

Stratum III is then determined.

Hence,

$$P\{\underline{R} = \underline{r} \mid (s,u,v)\}$$

$$\begin{aligned}
&= \binom{r_4-1}{s+r_1}\binom{q-r_4-1}{r_2}\binom{N-q}{N_1-(s+r_1+r_2)} \\
&\quad \times \binom{r_4-(s+r_1)-1}{u-1}\binom{q-r_4-1-r_2}{r_3}\binom{N-q-(N_1-(s+r_1+r_2))}{N_2-(u+r_3)} \\
&\times \Bigg[\sum_{\underline{r}}{}^{*}\binom{r_4-1}{s+r_1}\binom{q-r_4-1}{r_2}\binom{N-q}{N_1-(s+r_1+r_2)} \\
&\quad \times \binom{r_4-(s+r_1)-1}{u-1}\binom{q-r_4-1-r_2}{r_3}\binom{N-q-(N_1-(s+r_1+r_2))}{N_2-(u+r_3)}\Bigg]^{-1}
\end{aligned} \tag{4.23}$$

To evaluate $P\{E_{2(p)} | D_{1(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\} = P\{B_2^1\}$, we consider five cases:

Case I: $t < r_4$

Case II: $t = r_4$

Case III: $r_4 < t < q$

Case IV: $t = q$

Case V: $t > q$.

Case I:

$$P_I\{B_2^1\} = \frac{\binom{t-j}{p}\binom{r_4-(t-j)-(s+r_1)-1}{u-1-p}}{\binom{r_4-(s+r_1)-1}{u-1}} \qquad \text{for} \quad 0 \le p \le \min\begin{Bmatrix} t-j \\ u-1 \end{Bmatrix}$$

$$= \quad 0 \qquad \text{otherwise} \quad . \tag{4.24}$$

Case II:

$$P_{II}\{B_2^1\} = \quad 0 \quad \text{for all} \quad p \quad . \tag{4.25}$$

Case III:

$$P_{III}\{B_2^1\} = \frac{\binom{t-r_4-(j-(s+r_1))}{p-u}\binom{q-1-t-(s+r_1+r_2-j)}{u+r_3-p}}{\binom{q-r_4-r_2-1}{r_3}} \qquad \text{for} \quad u \le p \le u+r_3$$

$$= \quad 0 \qquad \text{otherwise} \quad . \tag{4.26}$$

Case IV:

$$P_{IV}\{B_2^1\} = 0 \quad \text{for all} \quad p \quad . \tag{4.27}$$

Case V:

$$P_V\{B_2^1\} = \frac{\binom{t-q-(j-(s+r_1+r_2))}{p-(u+r_3)}\binom{N-t-(N_1-j)}{N_2-p}}{\binom{N-q-(N-(s_1+r_1+r_2))}{N_2-(u+r_3)}} \qquad \text{for} \quad u+r_3 \leq p \leq N_2$$

$$= 0 \qquad \text{otherwise} \quad . \tag{4.28}$$

Introducing an indicator function

$${}_{\beta}t_{b}^{a} = \begin{cases} 1 & \text{if} \quad a \leq t < \beta \\ 0 & \text{otherwise} \end{cases} \tag{4.29}$$

we have

$$P\{E_{2(p)} \mid D_{1(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$= P_I\{B_2^1\} \; {}_{\beta}t_{0}^{r_4} + P_{III}\{B_2^1\} \; {}_{\beta}t_{r_4+1}^{q} + P_V\{B_2^1\} \; {}_{\beta}t_{q+1}^{N} \quad . \tag{4.30}$$

To evaluate $P\{D_{1(j)} \mid \underline{R} = \underline{r} \cap (s,u,v)\} = P\{C_1\}$ we consider the same five cases:

Case I:

$$P_I\{C_1\} = \frac{\binom{t-1}{j-1}\binom{r_4-t-1}{s+r_1-j}}{\binom{r_4-1}{s+r_1}} \tag{4.31}$$

Case II:

$$P_{II}\{C_1\} = 0 \tag{4.32}$$

Case III:

$$P_{III}\{C_1\} = \frac{\binom{t-r_4-1}{j-(s+r_1)-1}\binom{q-t-1}{s+r_1+r_2-j}}{\binom{q-r_4-1}{r_2}} \tag{4.33}$$

Case IV:

$$P_{IV}\{C_1\} = 0 \tag{4.34}$$

Case V:

$$P_V\{C_1\} = \frac{\binom{t-q-1}{j-(s+r_1+r_2)-1}\binom{N-t}{N_1-j}}{\binom{N-q}{N_1-(s+r_1+r_2)}} \quad . \tag{4.35}$$

Hence,

$$P\{D_{1(j)} \mid \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$= P_I\{C_1\} \sum_{t=0}^{r_4} + P_{III}\{C_1\} \sum_{t=r_4+1}^{q} + P_V\{C_1\} \sum_{t=q+1}^{N} \quad . \tag{4.36}$$

To evaluate $P\{E_{1(p)} \mid D_{2(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\} = P\{B_1^2\}$, we use a technique parallel to that just given, using the same five cases.

Case I:

$$P_I\{B_1^2\} = \frac{\binom{t-j}{p}\binom{r_4-(t-j)-u}{s+r_1-p}}{\binom{r_4-u}{s+r_1}} \qquad \text{if} \quad 0 \le p \le \begin{Bmatrix} t-j \\ s+r_1 \end{Bmatrix}$$

$$= \quad 0 \qquad \text{otherwise} \tag{4.37}$$

Case II:

$$P_{II}\{B_1^2\} = \begin{cases} 1 & \text{if} \quad p = s+r_1 \\ 0 & \text{otherwise} \end{cases} \tag{4.38}$$

Case III:

$$P_{III}\{B_1^2\} = \frac{\binom{t-r_4-(j-u)}{p-(s+r_1)}\binom{q-1-t-(u+r_3-j)}{s+r_1+r_2-p}}{\binom{q-r_4-r_3-1}{r_2}} \qquad \text{if} \quad s+r_1 \le p \le s+r_1+r_2$$

$$= \quad 0 \qquad \text{otherwise} \tag{4.39}$$

Case IV:

$$P_{IV}\{B_1^2\} = \quad 0 \quad \text{for all} \quad p \tag{4.40}$$

Case V:

$$P_V\{B_1^2\} = \frac{\binom{t-q-(j-(u+r_3))}{p-(s+r_1+r_2)}\binom{N-t-(N_2-j)}{N_1-p}}{\binom{N-q-(N_2-(u+r_3)}{N_1-(s+r_1+r_2)}} \qquad \text{if} \quad s+r_1+r_2 \leq p \leq N_1$$

$$= \quad 0 \qquad\qquad \text{otherwise} \quad . \tag{4.41}$$

Thus, we have

$$P\{E_{1(p)} \mid D_{2(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$= P_I\{B_1^2\}\ \beta\ {}_{0}^{r_4}t + P_{II}\{B_1^2\}\ \delta_{t,r_4}$$

$$+ P_{III}\{B_1^2\}\ \beta\ {}_{r_4+1}^{q}t + P_V\{B_1^2\}\ \beta\ {}_{q+1}^{N}t \quad . \tag{4.42}$$

Also, to evaluate $P\{D_{2(j)} \mid \underline{R} = \underline{r} \cap (s,u,v)\} = P\{C_2\}$, we have

Case I:

$$P_I\{C_2\} = \frac{\binom{t-1}{j-1}\binom{r_4-1-t}{u-1-j}}{\binom{r_4-1}{u-1}} \tag{4.43}$$

Case II:

$$P_{II}\{C_2\} = \begin{cases} 1 & \text{if} \quad j=u \\ 0 & \text{otherwise} \end{cases} \tag{4.44}$$

Case III:

$$P_{III}\{C_2\} = \frac{\binom{t-r_4-1}{j-u-1}\binom{q-1-t}{u+r_3-j}}{\binom{q-r_4-1}{r_3}} \tag{4.45}$$

Case IV:

$$P_{IV}\{C_2\} = 0 \tag{4.46}$$

Case V:

$$P_V\{C_2\} = \frac{\binom{t-q-1}{j-(u+r_3)-1}\binom{N-t}{N_2-j}}{\binom{N-q}{N_2-(u+r_3)}} \quad . \tag{4.47}$$

Then

$$P\{D_{2(j)} \mid \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$= P_I\{C_2\} \; \beta \overset{r_4}{\underset{0}{t}} + P_{II}\{C_2\} \; \delta_{t,r_4}$$

$$+ P_{III}\{C_2\} \; \beta \overset{q}{\underset{r_4+1}{t}} + P_V\{C_2\} \; \beta \overset{N}{\underset{q+1}{t}} \quad . \tag{4.48}$$

Finally, to evaluate $P\{E_{1(p)} \mid D_{3(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\} = P\{B_1^3\}$, using the same five cases, we have:

Case I:

$$P_I\{B_1^3\} = \frac{\binom{t-j}{p}\binom{s+r_1+u-1-(t-j)}{s+r_1-p}}{\binom{u+(s+r_1)-1}{s+r_1}} \quad \text{if } \left\{\begin{matrix} 0 \\ t-j-u-1 \end{matrix}\right\} \le p \le \left\{\begin{matrix} t-j \\ s+r_1 \end{matrix}\right\}$$

$$= \quad 0 \qquad \text{otherwise} \tag{4.49}$$

Case II:

$$P_{II}\{B_1^3\} = 0 \quad \text{for all} \quad p \tag{4.50}$$

Case III:

$$P_{III}\{B_1^3\} = \frac{\binom{t-(s+r_1+u)-j}{p-(s+r_1)}\binom{q-1-t-((v-1)-j)}{s+r_1+r_2-p}}{\binom{q-v-u-(s+r_1)}{r_2}} \quad \text{if} \quad s+r_1 \le p \le s+r_1+r_2$$

$$= \quad 0 \qquad \text{otherwise} \tag{4.51}$$

Case IV:

$$P_{IV}\{B_1^3\} = \begin{cases} 1 & \text{if} \quad p=s+r_1+r_2 \\ 0 & \text{otherwise} \end{cases} \tag{4.52}$$

Case V:

$$P_V\{B_1^3\} = \frac{\binom{t-q-(j-v)}{p-(s+r_1+r_2)}\binom{N-t-(N_3-j)}{N_1-p}}{\binom{N-q-(N_3-v)}{N_1-(s+r_1+r_2)}} \quad \text{if} \quad s+r_1+r_2 \le p \le N_1$$

$$= \quad 0 \qquad \text{otherwise} \quad . \tag{4.53}$$

Thus,

$$P\{E_{1(p)} \mid D_{3(j)} \cap \underline{R} = \underline{r} \cap (s,u,v)\}$$
$$= P_I\{B_1^3\}\beta t_0^{r_4} + P_{III}\{B_1^3\}\beta t_{r_4+1}^{q} + P_{IV}\{B_1^3\}\delta_{t,q} + P_V\{B_1^3\}\beta t_{q+1}^{N} \quad . \tag{4.54}$$

Also, to evaluate $P\{D_{3(j)} \mid \underline{R} = \underline{r} \quad (s,u,v)\} = P\{C_3\}$, we have

Case I:

$$P_I\{C_3\} = \frac{\binom{t-1}{j-1}\binom{r_4-1-t}{r_4-(s+r_1+u-1)-j}}{\binom{r_4-1}{r_4-(s+r_1+u)}} \tag{4.55}$$

Case II:

$$P_{II}\{C_3\} = 0 \tag{4.56}$$

Case III:

$$P_{III}\{C_3\} = \frac{\binom{t-r_4-1}{j-(r_4-(s+r_1+u))-1}\binom{q-1-t}{v-1-j}}{\binom{q-1-r_4}{v-(r_4-(s+r_1+u)-1)}} \tag{4.57}$$

Case IV:

$$P_{IV}\{C_3\} = \begin{cases} 1 & \text{if } j=v \\ 0 & \text{otherwise} \end{cases} \tag{4.58}$$

Case V:

$$P_V\{C_3\} = \frac{\binom{t-q-1}{j-v-1}\binom{N-t}{N_3-j}}{\binom{N-q}{N_3-v}} \quad . \tag{4.59}$$

Thus,

$$P\{D_{3(j)} \mid \underline{R} = \underline{r} \cap (s,u,v)\}$$

$$= P_I\{C_3\}\ \beta\ t_0^{r_4} + P_{III}\{C_3\}\ \beta\ t_{r_4+1}^{q}$$

$$+ P_{IV}\{C_3\}\ \delta_{t,q} + P_V\{C_3\}\ \beta\ t_{q+1}^{N} \quad . \qquad (4.60)$$

Substituting (4.23), (4.30) and (4.36) into (4.18); (4.23), (4.42), and (4.48) into (4.19); and (4.23), (4.54), and (4.60) into (4.20); and these into (4.3), (4.8) and (4.10) yields $P\{y_{(k)} \leq Y_{(t)} \mid (s,u,v)\}$.

C. Confidence Intervals Derived from the Sample C.D.F.

1. Definition of the confidence interval

We now consider the confidence interval procedure based on the sample C.D.F. for $L = 3$ strata. First, the sample C.D.F. is defined by

$$\hat{F}_{(y)} = \begin{cases} 0 & \text{if } y < y_{(1)} \\ j\,\dfrac{N_1}{n_1 N} + k\,\dfrac{N_2}{n_2 N} + (i-(j+k))\,\dfrac{N_3}{n_3 N} & \\ & \text{if } y_{(i)} \leq y < y_{(i+1)} \cap y_{(i)} = y_{1(j)} \cap y_{2(k)} < y_{(i)} < y_{2(k+1)} \\ k\,\dfrac{N_1}{n_1 N} + j\,\dfrac{N_2}{n_2 N} + (i-(j+k))\,\dfrac{N_3}{n_3 N} & \\ & \text{if } y_{(i)} \leq y < y_{(i+1)} \cap y_{(i)} = y_{2(j)} \cap y_{1(k)} < y_{(i)} < y_{1(k+1)} \\ k\,\dfrac{N_1}{n_1 N} + (i-(j+k))\,\dfrac{N_2}{n_2 N} + j\,\dfrac{N_3}{n_3 N} & \\ & \text{if } y_{(i)} \leq y < y_{(i+1)} \cap y_{(i)} = y_{3(j)} \cap y_{1(k)} < y_{(i)} < y_{1(k+1)} \\ 1 & \text{if } y \geq y_{(n)} \quad . \end{cases} \qquad (4.61)$$

To each pair (α,β) where $\alpha_0 < \alpha < \beta < 1$ and $\alpha_0 = \max_i\{(N_i/n_iN)\}$, there corresponds a unique pair of integers (k,r) where $1 \leq k \leq r \leq n$ such that

$$\begin{aligned} \hat{F}(y_{(k)}) &\leq \alpha < \hat{F}(y_{(k+1)}) \\ \hat{F}(y_{(r-1)}) &< \beta \leq \hat{F}(y_{(r)}) \end{aligned} \tag{4.62}$$

where $\hat{F}(y_{(0)}) = 0$, $\hat{F}(y_{(n+1)}) = 1$.

For given values of α and β, the confidence interval for $Y_{(t)}$ is given by $[y_{(k)}, y_{(r)})$ where $y_{(k)}, y_{(r)}$ are defined by (4.62). Of course, the integers k and r will vary in repeated sampling from Π_N.

2. A lower bound for the confidence coefficient

Proceeding as in Section C of Chapter III,

$$P\{y_{(k)} \leq Y_{(t)} < y_{(r)}\} \geq P\{\hat{F}(Y_{(t)}) < \beta\} - P\{\hat{F}(Y_{(t)}) < \alpha\}. \tag{4.63}$$

Let $D_{\ell(j)}$ be the event "$Y_{(t)} = Y_{\ell(j)}$", $\ell = 1, 2, 3,$; and $E_{\ell(p)}$ the event "$Y_{\ell(p)} < Y_{(t)} < Y_{\ell(p+1)}$" for $\ell = 1, 2, 3$.

Then,

$$\begin{aligned} P\{\hat{F}(Y_{(t)}) < \beta\} = &\sum_j \sum_p P\{\hat{F}(Y_{(t)}) < \beta | D_{1(j)} \cap E_{2(p)}\} P\{D_{1(j)} \cap E_{2(p)}\} \\ &+ \sum_j \sum_p P\{\hat{F}(Y_{(t)}) < \beta | D_{2(j)} \cap E_{1(p)}\} P\{D_{2(j)} \cap E_{1(p)}\} \\ &+ \sum_j \sum_p P\{\hat{F}(Y_{(t)}) < \beta | D_{3(j)} \cap E_{1(p)}\} P\{D_{3(j)} \cap E_{1(p)}\} . \end{aligned} \tag{4.64}$$

Letting $\hat{F}(Y_{(t)}) = \sum_{i=1}^{3} [m_i(t)]N_i/n_iN$ where $m_i(t) = m_i$ denotes the number of observations in the sample from stratum i with $Y \leq Y_{(t)}$,

$$P\{\hat{F}(Y_{(t)}) < \beta | D_{1(j)} \cap E_{2(p)}\} =$$
$$\sum_{m_1,m_2,m_3=0}^{\beta *} \binom{j}{m_1}\binom{N_1-j}{n_1-m_1}\binom{p}{m_2}\binom{N_2-p}{n_2-m_2}\binom{t-j-p}{m_3}\binom{N_3-(t-j-p)}{n_3-m_3} \Bigg/ \binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N_3}{n_3} \tag{4.65}$$

where $\sum^*$ denotes summation over all non-negative integers m_1, m_2 and m_3 such that $\sum_{i=1}^{3} (m_iN_i/n_iN) < \beta$. One may obtain $P\{\hat{F}(Y_{(t)}) < \beta | D_{2(j)} \cap E_{1(p)}\}$ from (4.65) by interchanging (m_1,n_1,N_1) and (m_2,n_2,N_2); and $P\{\hat{F}(Y_{(t)}) < \beta | D_{3(j)} \cap E_{1(p)}\}$ may be obtained from $P\{\hat{F}(Y_{(t)}) < \beta | D_{3(j)} \cap E_{1(p)}\}$ may be obtained from $P\{\hat{F}(Y_{(t)}) < \beta | D_{2(j)} \cap E_{1(p)}\}$ by interchanging (m_2,n_2,N_2) and (m_3,n_3,N_3).

Thus, (4.64) and, finally, (4.63) may be obtained by determining $P\{D_{1(j)} \cap E_{2(p)}\}$, $P\{D_{2(j)} \cap E_{1(p)}\}$ and $P\{D_{3(j)} \cap E_{1(p)}\}$: several suggestions have been explored in the previous section.

3. Derivation of the confidence coefficient

As in Section C of Chapter III denote the upper confidence limit, $y_{(r)}$, by y_U and the lower confidence limit, $y_{(k)}$, by y_L. Then,

$$P\{y_L \leq Y_{(t)} < y_U\} = P\{y_L \leq Y_{(t)}\} - P\{y_U \leq Y_{(t)}\} . \tag{4.66}$$

To determine $P\{y_U \leq Y_{(t)}\}$, first consider the set A_1 of non-negative integers u_1',u_2',u_3' such that $u_1'j_1 + u_2'j_2 + u_3'j_3 \geq \beta$,

$(u_1'-1)j_1 + u_2'j_2 + u_3'j_3 < \beta$ and (of course) $y_{(u_1'+u_2'+u_3')} = y_{1(u_1')}$; the set A_2 of non-negative integers u_1'',u_2'',u_3'' such that $u_1''j_1 + u_2''j_2 + u_3''j_3 \geq \beta$, $u_1''j_1 + (u_2''-1)\ j_2 + u_3''j_3 < \beta$ and $y_{(u_1''+u_2''+u_3'')} = y_{2(u_2'')}$; and the set A_3 of non-negative integers u_1''',u_2''',u_3''' such that $u_1'''j_1 + u_2'''j_2 + u_3'''j_3 \geq \beta$, $u_1'''j_1 + u_2'''j_2 + (u_3'''-1)j_3 < \beta$ and $y_{(u_1'''+u_2'''+u_3''')} = y_{3(u_3''')}$, where $j_i = N_i/n_iN$. Proceeding as in Section C of the previous chapter,

$$P\{y_U \leq Y_{(t)}\} = \sum_{(u_1',u_2',u_3')\varepsilon A_1} \sum_{i=0}^{t-\Sigma u_i'} P\{y_{(u_1'+u_2'+u_3')} = Y_{(t-i)}\}$$
$$+ \sum_{(u_1'',u_2'',u_3'')\varepsilon A_2} \sum_{i=0}^{t-\Sigma u_i''} P\{y_{(u_1''+u_2''+u_3'')} = Y_{(t-i)}\}$$
$$+ \sum_{(u_1''',u_2''',u_3''')\varepsilon A_3} \sum_{i=0}^{t-\Sigma u_i'''} P\{y_{(u_1'''+u_2'''+u_3''')} = Y_{(t-i)}\}. \tag{4.67}$$

Now,

$$P\{y_{(u_1'+u_2'+u_3')} = Y_{(t-i)}\} =$$

$$\sum_{j=u_1'} \sum_{p=u_2'} P\{y_{(u_1'+u_2'+u_3')} = Y_{(t-i)} | D^i_{1(j)} \cap E^i_{2(p)}\} P\{D^i_{1(j)} \cap E^i_{2(p)}\} \tag{4.68}$$

where $D^i_{\ell(j)}$ is the event "$Y_{(t-i)} = Y_{\ell(j)}$" for (ℓ=1,2,3); and $E^i_{\ell(p)}$ is the event "$Y_{\ell(p)} < Y_{(t-i)} < Y_{\ell(p+1)}$" for ($\ell$=1,2,3). Similarly,

$$P\{y_{(u_1''+u_2''+u_3'')} = Y_{(t-i)}\} =$$

$$\sum_{j=u_2''}\sum_{p=u_1''} P\{y_{(u_1''+u_2''+u_3'')} = Y_{(t-i)} | D^i_{2(j)} \cap E^i_{1(p)}\} P\{D^i_{2(j)} \cap E^i_{1(p)}\} \tag{4.69}$$

and,

$$P\{y_{(u_1'''+u_2'''+u_3''')} = Y_{(t-i)}\} =$$

$$\sum_{j=u_3'''}\sum_{p=u_1'''} P\{y_{(u_1'''+u_2'''+u_3''')} = Y_{(t-i)} | D^i_{3(j)} \cap E^i_{1(p)}\} P\{D^i_{3(j)} \cap E^i_{1(p)}\} . \tag{4.70}$$

Then, it is easily seen that

$$P\{y_{(u_1'+u_2'+u_3')} = Y_{(t-i)} | D^i_{1(j)} \cap E^i_{2(p)}\} =$$
$$\binom{j-1}{u_1'-1}\binom{N_1-j}{n_1-u_1'}\binom{p}{u_2'}\binom{N_2-p}{n_2-u_2'}\binom{t-i-j-p}{u_3'}\binom{N_3-(t-i-j-p)}{n_3-u_3'} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N_3}{n_3} . \tag{4.71}$$

$P\{y_{(u_1''+u_2''+u_3'')} = Y_{(t-i)} | D^i_{2(j)} \cap E^i_{1(p)}\}$ can be obtained from (4.71) by interchanging (u_1', n_1, N_1) and (u_2', n_2, N_2) and then replacing u_k' with u_k'', while $P\{y_{(u_1'''+u_2'''+u_3''')} = Y_{(t-i)} | D^i_{3(j)} \cap E^i_{1(p)}\}$ can be obtained from $P\{y_{(u_1''+u_2''+u_3'')} = Y_{(t-i)} | D^i_{2(j)} \cap E^i_{1(p)}\}$ by interchanging (u_2'', n_2, N_2) and (u_3'', n_3, N_3) and then replacing u_k'' with u_k'''.

Thus, $P\{y_U \leq Y_{(t)}\}$ may be obtained from the expressions given above plus a determination of $P\{D^i_{1(j)} \cap E^i_{2(p)}\}$, $P\{D^i_{2(j)} \cap E^i_{1(p)}\}$ and $P\{D^i_{3(j)} \cap E^i_{1(p)}\}$. $P\{y_L \leq Y_{(t)}\}$ may be derived in an analogous manner.

To find the components of the set A_1, note that for each value of u_2' ($u_2'=0,1,\ldots,n_2$) and u_3' ($u_3'=0,1,\ldots,n_3$), if $0 \leq \{[(\beta - u_2'j_2 - u_3'j_3)^{-}/j_1] + 1\} \leq n_1$, then $([(\beta - u_2'j_2 - u_3'j_3)^{-}/j_1] + 1, u_2', u_3') \in A_1$. The components of the sets A_2 and A_3 may be determined in a similar manner.

V. ADDITIONAL APPLICATIONS AND EXTENSIONS

A. Applications

1. Tolerance regions - two strata situation

We now demonstrate how our work in Chapter III can be converted into the context of tolerance regions. In particular, assume a fixed β, $0 < \beta < 1$, and consider the interval $[y_{(k)}, y_{(r)}]$ in the method of the combined sample approach. Then, $[y_{(k)}, y_{(r)}]$ is a β-content tolerance region at confidence level γ if

$$P\{[y_{(k)}, y_{(r)}] \text{ contains at least } 100\ \beta\% \text{ of population values}\} = \gamma . \tag{5.1}$$

We see immediately that if $[y_{(k)}, y_{(r)}]$ is to contain at least 100 β% of the population values, at least $[\beta N]+1$ elements from our population must have associated values in the interval. (If βN is an integer, replace $[\beta N]+1$ by βN in the following derivation.)

Turning to the computation of γ, we have

$$\gamma = P\{\text{at least } [\beta N]+1 \text{ elements from population are in } [y_{(k)}, y_{(r)}]\}$$

$$= \sum_{i=0}^{N-(n-r)-([\beta N]+1)-k} \ \sum_{t=k}^{N-([\beta N]+1)-i} P\{y_{(k)}=Y_{(t)} \cap y_{(r)}=Y_{([\beta N]+1+t-i)}\}. \tag{5.2}$$

For ease of notation, let $t' = [\beta N]+1+t-i$, and let $D_{\ell(j)}$ be the event "$Y_{(t)} = Y_{\ell(j)}$", $\ell = 1,2$; and $D'_{\ell(j')}$ the event "$Y_{(t')} = Y_{\ell(j')}$", $\ell = 1,2$.

Then,

$$
\begin{aligned}
P\{y_{(k)} &= Y_{(t)} \cap y_{(r)} = Y_{(t')}\} \\
&= \sum_{j'=2}^{N_1} \sum_{j=1}^{j'-1} P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} \mid D_{1(j)} \cap D'_{1(j')}\} \\
&\qquad \times P\{D_{1(j)} \cap D'_{1(j')}\} \\
&+ \sum_{j'=1}^{N_2} \sum_{j=1}^{N_1} P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} \mid D_{1(j)} \cap D'_{2(j')}\} \\
&\qquad \times P\{D_{1(j)} \cap D'_{2(j')}\} \\
&+ \sum_{j'=1}^{N_1} \sum_{j=1}^{N_2} P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} \mid D_{2(j)} \cap D'_{1(j')}\} \\
&\qquad \times P\{D_{2(j)} \cap D'_{1(j')}\} \\
&+ \sum_{j'=2}^{N_2} \sum_{j=1}^{j'-1} P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} \mid D_{2(j)} \cap D'_{2(j')}\} \\
&\qquad \times P\{D_{2(j)} \cap D'_{2(j')}\} \quad . \qquad (5.3)
\end{aligned}
$$

The formulas for $P\{D_{\ell(j)} \cap D'_{\ell'(j')}\}$ are given in Section C of Chapter III.

Finally:

$$P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} | D_{1(j)} \cap D'_{1(j')}\}$$

$$= \sum_{\ell'=1}^{r-k} \sum_{\ell=1}^{k} \binom{j-1}{\ell-1}\binom{j'-j-1}{\ell'-1}\binom{N_1-j'}{n_1-(\ell+\ell')}\binom{t-j}{k-\ell}$$

$$\times \binom{(t'-t)-(j'-j)}{r-\ell'-k}\binom{N_2-(t'-j')}{n_2-(r-\ell-\ell')} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2} \qquad (5.4)$$

$$P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} | D_{1(j)} \cap D'_{2(j')}\}$$

$$= \sum_{\ell'} \sum_{\ell} \binom{j-1}{\ell-1}\binom{t'-j-j'}{r-k-\ell'}\binom{N_1-(t'-j')}{n_1-(r-k)+(\ell'-\ell)}$$

$$\times \binom{t-j}{k-\ell}\binom{j'-(t-j)-1}{\ell'-1}\binom{N_2-j'}{n_2-(k-\ell)-\ell'} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2} \qquad (5.5)$$

$$P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} | D_{2(j)} \cap D'_{1(j')}\}$$

$$= \sum_{\ell'} \sum_{\ell} \binom{t-j}{k-\ell}\binom{j'-(t-j)-1}{\ell'-1}\binom{N_1-j'}{n_1-(k-\ell)-\ell'}$$

$$\times \binom{j-1}{\ell-1}\binom{t'-j'-j}{r-k-\ell'}\binom{N_2-(t'-j')}{n_2-(r-k)+(\ell'-\ell)} \Big/ \binom{N_1}{n_1}\binom{N_2}{n_2} \qquad (5.6)$$

$$P\{y_{(k)} = Y_{(t)} \cap y_{(r)} = Y_{(t')} | D_{2(j)} \cap D'_{2(j')}\}$$

$$= \sum_{\ell'} \sum_{\ell} \binom{t-j}{k-\ell} \binom{t'-j'-(t-j)}{r-\ell'-k} \binom{N_1-(t'-j')}{n_1-r+(\ell+\ell')}$$

$$\times \binom{j-1}{\ell-1} \binom{j'-j-1}{\ell'-1} \binom{N_2-j'}{n_2-\ell'-\ell} \bigg/ \binom{N_1}{n_1} \binom{N_2}{n_2} \quad . \tag{5.7}$$

Using the C.D.F. method in contrast to the combined method, we have

$$\gamma = P\{\text{at least } [\beta N] \text{ elements from population in } [y_L, y_U]$$

$$= \sum_{i=0}^{N-[\beta N]} \sum_{t=1}^{N-[\beta N]+1-i} P\{y_L = Y_{(t)} \cap y_U = Y_{([\beta N]+1+t+i)}\} \quad . \tag{5.8}$$

These terms can be calculated using methods derived in Chapter III.

2. Best finite population problem

Using the notation of Chapter III, we now consider the problem:

Given two populations, we wish to arrive at a decision of which is the "better" of the two, in the sense that Population 2 is "better" than Population 1 if $Y_{1(s)} < Y_{2(u)}$.

The correspondence between Chapter II and this section is: interpret the strata to be the Populations, so that $Y_{i(j)}$ is the j-th ordered element in the i-th population, i = 1,2; $y_{i(j)}$ is the j-th ordered observation in the sample drawn from Population i.

Our decision as to the "better" of the two populations is based on comparing $y_{1(k)}$ and $y_{2(r)}$. If $y_{1(k)} < y_{2(r)}$, we will say Population 2 is "better". We now derive the probability of a correct decision.

$$P\{y_{1(k)} < y_{2(r)} | Y_{1(s)} < Y_{2(u)}\}$$

$$= \sum_{n=r}^{N} \sum_{m=k}^{n-(r-k)} P\{y_{1(k)} < y_{2(r)} \cap y_{1(k)} = Y_{(m)}$$

$$\cap\, y_{2(r)} = Y_{(n)} | Y_{1(s)} < Y_{2(u)}\}$$

$$= \sum_{m<n} P\{y_{1(k)} < y_{2(r)} | y_{1(k)} = Y_{(m)} \cap y_{2(r)} = Y_{(n)}$$

$$\cap\, Y_{1(s)} < Y_{2(u)}\}$$

$$\times\, P\{y_{1(k)} = Y_{(m)} \cap y_{2(r)} = Y_{(n)} | Y_{1(s)} < Y_{2(u)}\} \;. \quad (5.9)$$

The first term in this summation is identically 1, so it remains to find an expression for the latter term.

$$P\{y_{1(k)} = Y_{(m)} \cap y_{2(r)} = Y_{(n)} | Y_{1(s)} < Y_{2(u)}\}$$

$$= \sum_{j} \sum_{j'} P\{y_{1(k)} = Y_{(m)} \cap y_{2(r)} = Y_{(n)} |$$

$$Y_{(m)} = Y_{1(j)} \cap Y_{(n)} = Y_{2(j')} \cap Y_{1(s)} < Y_{2(u)}\}$$

$$\times\, P\{Y_{(m)} = Y_{1(j)} \cap Y_{(n)} = Y_{2(j')} | Y_{1(s)} < Y_{2(u)}\} \;. \quad (5.10)$$

Now,

$$P\{y_{1(k)} = Y_{(m)} \cap y_{2(r)} = Y_{(n)} \mid$$

$$Y_{(m)} = Y_{1(j)} \cap Y_{(n)} = Y_{2(j')} \cap Y_{1(s)} < Y_{2(u)}\}$$

$$= \frac{\binom{j-1}{k-1}\binom{1}{1}\binom{N_1-j}{n_1-k}\binom{j'-1}{r-1}\binom{1}{1}\binom{N_2-j'}{n_2-r}}{\binom{N_1}{n_1}\binom{N_2}{n_2}} \quad . \tag{5.11}$$

Also, $P\{Y_{(m)} = Y_{1(j)} \cap Y_{(n)} = Y_{2(j')} \mid Y_{1(s)} < Y_{2(u)}\}$ is Case II in the Section F of Chapter III.

Hence, in the case of two populations, the probability of a correct decision is obtainable.

B. Extension: Cluster Sampling

1. Introduction

The problem of finding a non-parametric confidence interval for the population median with cluster sampling as the sampling design has been investigated by Chapman [1970]. In his (unpublished) Ph.D. thesis, the confidence coefficients for the confidence intervals in cluster sampling are approximated. His assumptions include that the random variable of interest has a continuous distribution over the entire population as well as within each cluster. Hence, it is assumed that each cluster is of infinite size, and that there are an infinite number of clusters.

We approach the same problem, but with the following assumptions:

1) Population size: N.
2) Each element of the population has a distinct Y-value associated with it.
3) There are K clusters, of sizes M_i, $i = 1, 2, \ldots, K$, $(\sum M_i = N)$
4) When sampling from the i-th cluster, a simple random sample of size m_i is drawn.

2. <u>One cluster chosen</u>

We first consider the situation in which one cluster (say cluster i) is chosen from the K clusters, and then a simple random sample of size m_i chosen from that cluster. Letting Q be the number of items in the sample with associated Y-values less than or equal to $Y_{(t)}$, we have

$$P\{y_{(k)} \leq Y_{(t)} \leq y_{(r)}\} = P\{y_{(k)} \leq Y_{(t)}\} - P\{y_{(r)} \leq Y_{(t-1)}\} \tag{5.12}$$

and

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{q=k}^{t} P\{Q = q\} \; . \tag{5.13}$$

If we let A_i be the event "cluster i is chosen" and B_{t_i} the event "exactly t_i elements in cluster i have Y-values less than or equal to $Y_{(t)}$", it follows that

$$P\{Q = q\} = \sum_{i=1}^{K} \sum_{t_i=0}^{M_i} P\{Q = q \cap A_i \cap B_{t_i}\}$$

$$= \sum_{i} \sum_{t_i} P\{Q = q | A_i \cap B_{t_i}\} \, P\{B_{t_i} | A_i\} \, P\{A_i\} \quad . \tag{5.14}$$

Considering the components of (5.14) separately, we have

$$P\{Q = q | A_i \cap B_{t_i}\} = \binom{t_i}{q} \binom{M_i - t_i}{m_i - q_i} \bigg/ \binom{M_i}{m_i} \quad , \tag{5.15}$$

and

$$P\{A_i\} = \pi_i \quad , \tag{5.16}$$

where $\pi_i = 1/K$ if the cluster is chosen by simple random sampling, $\pi_i = M_i/N$ if the cluster is chosen with probability proportional to size, etc.

Turning to $P\{B_{t_i} | A_i\}$, we consider several cases.

If we assume random clustering,

$$P\{B_{t_i} | A_i\} = \binom{t}{t_i} \binom{N-t}{M_i - t_i} \bigg/ \binom{N}{M_i} \quad . \tag{5.17}$$

If we assume "ordered" clustering, i.e., $Y_{1(s)} < Y_{2(u)}$ for $K = 2$ clusters, or $Y_{1(s)} < Y_{2(u)} < Y_{3(v)}$ for $K = 3$ clusters, previously derived formulas can be slightly altered to obtain $P\{B_{t_i} | A_i\}$.

Referring to Chapter III, $P\{B_{t_i} | A_i\}$ is quite analogous to $P\{B_j^i\} = P\{Y_{(t)} = Y_{i(j)}\}$, for $P\{B_{t_i} | A_i\} = P\{Y_{i(t_i)} \leq Y_{(t)} < Y_{i(t_i+1)}\}$.

Thus, referring to (3.41) and (3.44), for $K = 2$ clusters and $Y_{1(s)} < Y_{2(u)}$, replacing "j" by "t_i", "$\binom{t-1}{j-1}$" with "$\binom{t}{t_i}$", "N_i" by "M_i", and keeping in mind that we just require that "$Y_{i(t_i)} \leq Y_{(t)}$" and not "$Y_{i(t_i)} = Y_{(t)}$", we have

$$
\begin{aligned}
P\{B_{t_1} | A_1\} = \Bigg\{ & \binom{t}{t_1} \sum_d \binom{s+d+(u-1)-t}{s+d-t_1} \binom{N-(s+d+u)}{M_1-(s+d)} \alpha^t_{s+d+(u-1)} \\
& + \binom{t-1}{t_1} \binom{N-t}{M_1-t_1} \delta_{t,s+d+u} \\
& + \binom{N-t}{M_1-t_1} \sum_d \binom{s+d+(u-1)}{s+d} \binom{t-(s+d+u)}{t_1-(s+d)} \left(1-\alpha^t_{s+d+u}\right) \Bigg\} \\
& \times \left[\sum_d \binom{s+d+(u-1)}{s+d} \binom{N-(s+d+u)}{M_1-(s+d)} \right]^{-1}
\end{aligned}
\tag{5.18}
$$

and

$$
\begin{aligned}
P\{B_{t_2} | A_2\} = \Bigg\{ & \binom{t}{t_2} \sum_d \binom{s+d+(u-1)-t}{u-1-t_2} \binom{N-(s+d+u)}{M_2-u} \alpha^t_{s+d+(u-1)} \\
& + \binom{t}{t_2} \binom{N-t}{M_2-t_2} \delta_{t,s+d+u} \\
& + \binom{N-t}{M_2-t_2} \sum_d \binom{s+d+(u-1)}{u-1} \binom{t-(s+d+u)}{t_2-u} \left(1-\alpha^t_{s+d+u}\right) \Bigg\} \\
& \times \left[\sum_d \binom{s+d+(u-1)}{s+d} \binom{N-(s+d+u)}{M_1-(s+d)} \right]^{-1} .
\end{aligned}
\tag{5.19}
$$

3. Two clusters selected

We now consider the situation in which two clusters (say Cluster i and j) are selected from the K clusters. Simple random samples of sizes m_i and m_j are selected from these clusters, and the samples ordered as $y_{(1)} \cdots y_{(m_i+m_j)}$. Letting Q be the number of items in the ordered sample with associated Y-values less than or equal to $Y_{(t)}$, we have, as before

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{q=r}^{t} P\{Q = q\} \quad . \tag{5.20}$$

Now

$$P\{Q = q\} = \sum_{i<j} \sum_{t_i+t_j=0}^{t_*} P\{Q = q \cap A_i \cap A_j \cap B_{t_i} \cap B_{t_j}\}$$

$$= \sum \sum{}^{*} P\{Q = q | A_i \cap A_j \cap B_{t_i} \cap B_{t_j}\}$$

$$\times P\{B_{t_i} \cap B_{t_j} | A_i \cap A_j\} \times P\{A_i \cap A_j\} \quad . \tag{5.21}$$

Then

$$P\{Q = q | A_i \cap A_j \cap B_{t_i} \cap B_{t_j}\}$$

$$= \sum_{\ell=0}^{q} \binom{t_i}{\ell}\binom{M_i - t_i}{m_i - \ell}\binom{t_j}{q-\ell}\binom{M_j - t_j}{m_j - (q-\ell)} \Bigg/ \binom{M_i}{m_i}\binom{M_j}{m_j} \quad . \tag{5.22}$$

Also,

$$P\{A_i \cap A_j\} = \pi_{ij} \quad , \tag{5.23}$$

where $\Pi_{ij} = 1/\binom{k}{2}$ if the two clusters are chosen by simple random sampling.

Under the assumption of random clustering, we have

$$P\{B_{t_i} \cap B_{t_j} | A_i \cap A_j\} = \frac{\binom{t}{t_i}\binom{t-t_i}{t_j}\binom{N-t}{M_i-t_i}\binom{N-t-(M_i-t_i)}{M_j-t_j}}{\binom{N}{M_i}\binom{N-M_i}{M_j}} \quad . \tag{5.24}$$

A different approach to this same problem is suggested by Chapman. We alter our assumptions by having all K clusters equal-sized, M. Let P be a random variable, representing the proportion of elements in an individual cluster with associated Y-values less than or equal to $Y_{(t)}$. In this case, $P \in \{0, \frac{1}{M}, \frac{2}{M}, \ldots, 1\}$.

4. <u>One cluster chosen</u>

Suppose one cluster is selected at random from the K clusters, and a simple random sample of size m chosen from it. Letting R = number of items in sample with $Y \leq Y_{(t)}$, and $f(\cdot)$ the density function of P, we have

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{r=k}^{m} P\{R=r\} \quad . \tag{5.25}$$

Now

$$P\{R=r\} = \sum_{p} P\{R=r \cap P=p\} = \sum_{p} P\{R=r | P=p\} \times P\{P=p\}$$

$$= \sum_{p} \binom{Mp}{r}\binom{M(1-p)}{m-r} \Big/ \binom{M}{m} \times f(p) = \sum_{t'=0}^{M} \binom{t'}{r}\binom{M-t'}{m-r} \Big/ \binom{M}{m} f\left(\frac{t'}{M}\right) \quad . \tag{5.26}$$

Turning to $f(\cdot)$, if the number of items in the selected cluster less than or equal to $Y_{(t)}$ has a hypergeometric distribution, that is

$$f(t'/m) = \binom{t}{t'}\binom{N-t}{M-t'}\bigg/\binom{N}{M} , \tag{5.27}$$

then, after some simplification

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{i=k} \binom{t}{i}\binom{N-t}{m-i}\bigg/\binom{N}{m} , \tag{5.28}$$

which is equivalent to taking a simple random sample of size m from the entire population, ignoring clustering. Thus, the hypergeometric assumption is equivalent to "random" clustering.

If, instead, we have

$$f(t'/N) = \binom{b+t'-1}{b-1}\binom{N-t'+a-1}{a-1}\bigg/\binom{N+a+b-1}{a+b-1} , \tag{5.29}$$

which is the negative hypergeometric density with parameters (b,a), and is the discrete analogue of the Beta distribution, we have

$$P\{y_{(k)} \leq Y_{(t)}\} = \sum_{i=k}^{m} \frac{\binom{b+i-1}{i}\binom{a+m-i-1}{m-i}}{\binom{m+a+b-1}{m}} . \tag{5.30}$$

Similar work can be done for the situation in which two clusters are selected.

VI. BIBLIOGRAPHY

Abdel-Aty, S. H. 1954. Ordered variables in discontinuous distributions. Statistica Neerlandica 8: 61-82.

Abramowitz, Milton and Irene A. Stegun. 1965. Handbook of mathematical functions. New York, N.Y., Dover Publications, Inc.

Anderson, T. W. and S. M. Samuels. 1967. Some inequalities among binomial and Poisson probabilities. In Proceedings of the Fifth Berkeley Symposium. Vol. I. pp. 1-12. Berkeley, California, University of California Press.

Blom, Gunnar. 1958. Statistical estimates and transformed Beta-variables. New York, N.Y., John Wiley and Sons, Inc.

Bradley, James V. 1968. Distribution-free statistical tests. Englewood Cliffs, N.J., Prentice-Hall, Inc.

Chapman, David Welland. 1970. Cluster sampling and approximate distribution-free confidence intervals for a median. Unpublished Ph.D. thesis. Ithaca, N.Y., Library, Cornell University.

Chung, J. H. and D. B. DeLury. 1950. Confidence limits for the hypergeometric distribution. Toronto, Canada, University of Toronto Press.

Cochran, W. G. 1963. Sampling techniques. 2nd ed. New York, N.Y., John Wiley and Sons, Inc.

David, F. N. and N. L. Johnson. 1954. Statistical treatment of censored data. Biometrika 41: 228-240.

David, H. A. 1970. Order Statistics. New York, N.Y., John Wiley and Sons, Inc.

David, H. A. and R. S. Mishriky. 1968. Order statistics for discrete populations and grouped samples. J. Amer. Stat. Assn. 63: 1390-1398.

Feller, William. 1957. An introduction to probability theory and its applications. Vol. 1. 2nd ed. New York, N.Y., John Wiley and Sons, Inc.

Fisher, R. A. and F. Yates. 1949. Statistical tables for biological, agricultural, and medical research. London, Oliver and Boyd, Ltd.

Gupta, Shanti S. 1960. Order statistics from the Gamma distribution. Technometrics 2: 243-262.

Guttman, Irwin. 1967. Tolerance regions--a survey of its literature: I. Introduction and discussion of distribution-free tolerance regions. University of Wisconsin Department of Statistics Technical Report 123.

Hansen, Morris, William Hurwitz, and William Madow. 1953. Sample survey methods and theory. Vol. II: Theory. New York, N.Y., John Wiley and Sons, Inc.

Harter, H. Leon. 1961. Expected values of normal order statistics. Biometrika 48: 151-165.

Harter, H. Leon. 1970. Order statistics and their use in testing and estimation. Vol. 2: Estimates based on order statistics of samples from various populations. Washington, D.C., U. S. Govt. Print. Off.

Hastings, C., Frederick Mosteller, John Tukey, and Charles Winsor. 1947. Low moments for small samples: a comparative study of order statistics. Ann. Math. Stat. 18: 413-426.

Hill, Bruce M. 1968. Posterior distributions of percentiles: Bayes' theorem for sampling from a population. J. Amer. Stat. Assn. 63: 677-691.

Hoeffding, Wassily. 1956. On the distribution of the number of successes in independent trials. Ann. Math. Stat. 27: 713-721.

Hogg, Robert V. and Allen T. Craig. 1970. Introduction to mathematical statistics. 3rd ed. New York, N.Y., The Macmillan Company.

Johnson, Norman L. and Samuel Kotz. 1969. Discrete distributions. Boston, Mass., Houghton Mifflin Company.

Kabe, D. G. 1969. Some distribution problems of order statistics from discrete distributions. Institute of Statistical Mathematics Annals 21: 551-556.

Katz, Leo. 1953. Confidence intervals for the number showing a certain characteristic in a population when sampling is without replacement. J. Amer. Stat. Assn. 48: 256-261.

Kempthorne, Oscar. 1967. The classical problem of inference--goodness of fit. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. I. pp. 235-249. Berkeley, California, University of California Press.

Kendall, M. G. and A. Stuart. 1963. The advanced theory of statistics. Vol. 1: 2nd ed. London, Charles Griffin and Company, Ltd.

Khatri, C. G. 1963. Distributions of order statistics for discrete case. Institute of Statistical Mathematics Annals 14: 167-171.

Knuth, Donald E. 1968. The art of computer programming. Vol. 1: Fundamental algorithms. Reading, Mass., Addison-Wesley Publishing Company.

Knuth, Donald E. 1969. The art of computer programming. Vol. 2: Seminumerical algorithms. Reading, Mass., Addison-Wesley Publishing Company.

Krishnaiah, P. R. and M. Haseeb Rizui. 1967. A note on moments of Gamma order statistics. Technometrics 9: 315-318.

Lehmann, E. L. 1959. Testing statistical hypotheses. New York, N.Y., John Wiley and Sons, Inc.

Lieberman, G. J. and D. B. Owen. 1961. Tables of the hypergeometric probability distribution. Stanford, Calif., Stanford University Press.

Loynes, R. M. 1966. Some aspects of the estimation of quantiles. Royal Stat. Soc. J. 28: 497-512.

MacKinnon, W. J. 1964. Table for both the sign test and distribution-free confidence intervals of the median for sample sizes to 1000. J. Amer. Stat. Assn. 59: 935-956.

McCarthy, Philip J. 1965. Stratified sampling and distribution-free confidence intervals for a median. J. Amer. Stat. Assn. 60: 772-783.

Madansky, Albert. 1962. More on length of confidence intervals. J. Amer. Stat. Assn. 57: 586-589.

Mosteller, Frederick. 1946. On some useful "inefficient" statistics. Ann. Math. Stat. 17: 377-408.

Nair, K. R. 1940. Table of confidence intervals for the median in samples from any continuous population. Sankhyā 4: 551-558.

Noether, Gottfried E. 1967. Elements of non-parametric statistics. New York, N.Y., John Wiley and Sons, Inc.

Noether, Gottfried E. 1948. On confidence limits for quantiles. Ann. Math. Stat. 19: 416-419.

Parzen, Emanuel. 1960. Modern probability theory and its applications. New York, N.Y., John Wiley and Sons, Inc.

Pearson, Karl. 1901. Note on Francis Galton's problem. Biometrika 1: 390-399.

Plackett, R. L. 1958. Linear estimation from censored data. Ann. Math. Stat. 29: 131-142.

Pratt, John W. 1961. Length of confidence intervals. J. Amer. Stat. Assn. 56: 549-566.

Rao, C. R. 1965. Linear statistical inference and its applications. New York, N.Y., John Wiley and Sons, Inc.

Riordan, John. 1968. Combinatorial identities. New York, N.Y., John Wiley and Sons, Inc.

Särndal, Carl-Eric. 1964. A unified derivation of some non-parametric distributions. J. Amer. Stat. Assn. 59: 1042-1053.

Savur, S. R. 1937. The use of the median in tests of significance. Proc. Indian Acad. Sci., Sect. A, 5: 564-576.

Saw, J. G. 1960. A note on the error after a number of terms of the David-Johnson series for the expected values of normal order statistics. Biometrika 47: 79-86.

Scheffe, H. and J. W. Tukey. 1945. Non-parametric estimation. I. Validity of order statistics. Ann. Math. Stat. 16: 187-192.

Sedransk, J. 1969. Some elementary properties of systematic sampling. Skandinavisk Aktuarietidskrift (1969): 39-47.

Sen, Pranab Kumar. 1959. On the moments of the sample quantiles. Calcutta Stat. Assn. Bull. 9: 1-20.

Sillitto, G. P. 1951. Interrelations between certain linear systematic statistics of samples from any continuous population. Biometrika 38: 377-382.

Steck, G. P. and W. J. Zimmer. 1968. The relation between Neyman and Bayes confidence intervals for the hypergeometric parameter. Technometrics 10: 199-203.

Teichroew, D. 1956. Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. Ann. Math. Stat. 27: 410-426.

Thompson, W. R. 1936. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. Ann. Math. Stat. 7: 122-128.

U. S. Army Ordinance Corps. 1952. Tables of the cumulative binomial probabilities. Washington, D.C., U. S. Govt. Print. Off.

Walker, A. M. 1968. A note on the asymptotic distribution of sample quantiles. Roy. Stat. Soc. J. - B - 30: 570-575.

Walsh, John E. 1962a. Handbook of non-parametric statistics. Princeton, N.J., D. van Nostrand Company, Inc.

Walsh, John E. 1962b. Non-parametric confidence intervals and tolerance regions. In Sarhan, Ahmed E. and Bernard G. Greenberg, eds. Contributions to order statistics. pp. 136-140. New York, N.Y., John Wiley and Sons, Inc.

Wilks, Samuel S. 1962. Mathematical statistics. New York, N.Y., John Wiley and Sons, Inc.

Wilks, Samuel S. 1948. Order statistics. Amer. Math. Soc. Bull. 54: 6-50.

Woodruff, R. S. 1952. Confidence intervals for medians and other position measures. J. Amer. Stat. Assn. 47: 635-646.

VII. ACKNOWLEDGMENTS

This thesis has been prepared under the direction of Dr. Joseph Sedransk, whose persistent efforts, guidance and encouragement have been invaluable in its completion.

For financial support of my graduate studies and thesis work, I am indebted to the National Aeronautics and Space Agency for the NASA Traineeship awarded to me while a student at Iowa State University (1964-1966). I am also grateful for the National Science Faculty Fellowship, granted by NSF, which enabled me to continue my studies and research at the University of Wisconsin (1969-1970).

In addition, my employer, Cornell College has provided financial support and computer facilities during my thesis preparation.

The efforts of the staff and faculty of the Statistics Department at Iowa State University have been greatly appreciated. Dr. T. A. Bancroft advised and counselled me in my early graduate work. I also thank Dr. H. T. David who kindly consented to be the chairman of my Graduate Committee.

Through all the trials and tribulations of graduate work and thesis preparation, my wife, Sue, has been at my side (or at home with our children) offering both encouragement and sympathy. Thanks, Sam!

Finally, I wish to recognize Mrs. Dale Donald, skilled typist and translator, whose artful talents have produced this finished product.