Sankhyā: The Indian Journal of Statistics
Special Issue on Sample Surveys
1999, Volume 61, Series B, Pt. 1, pp. 106-132

CONFIDENCE BOUNDS FOR SURVEY-WEIGHTED QUANTILE PLOTS AND OFFSET-FUNCTION PLOTS*

By S.R. LEE
and
J.L. ELTINGE
Texas A&M University, College Station

SUMMARY. This paper discusses construction and interpretation of quantile plots using data obtained through a complex sample design. Previous quantile-plotting methods are extended through the use of survey-weighted quantile point estimators. The resulting graphical methods include normal quantile plots and related normal offset-function plots; and quantile and offset-function plots for comparison of two subpopulations.

Confidence bounds associated with each quantile plot can be based on any of three related methods of variance estimation and pivot construction. The first method is based on the Francisco and Fuller (1991) test inversion approach to confidence interval construction. The second method is based on the Woodruff (1952) direct inversion of the quantile function. This second method can be viewed as a variant on the first method, based on a local approximation to the variance of the distribution-function estimator. The third approach is based on a direct normal approximation for the distribution of quantile point estimators. In general, each of the three approaches can be used either for construction of pointwise confidence bounds; or for construction of simultaneous confidence bounds at k quantile points, e.g., the deciles of a distribution. For simultaneous inference at k predetermined points, one generally will prefer to use Bonferroni-based critical points in construction of confidence bounds, but one can also consider Scheffé-based critical points. Preference for a given method depends on trade-offs between computational burden and specific inferential goals. The proposed plotting and inference methods are applied to medical-examination data from the Third National Health and Nutrition Examination Survey (NHANES III).

1. Introduction

1.1 *Quantile plots: Two motivating examples.* Quantile plots provide relatively simple diagnostic tools to examine the distributions associated with

AMS (1991) *subject classification.* 62D05.
Key words and phrases. Normal plot; *p-p* plot; *q-q* plot; shift function; third national health and nutrition examination survey (NHANES III); Test inversion.
* This work was supported in part by the National Center for Health Statistics. The views expressed in this paper are those of the authors and do not necessarily represent the policies of the National Center for Health Statistics.

one or more populations. To review some central ideas, let $F(\cdot)$ and $G(\cdot)$ be two continuous distribution functions. In addition, for any $q \in (0,1)$, define the population quantiles for $F(\cdot)$,

$$X_{Fq} = F^{-1}(q) = inf\{x : F(x) \geq q\} \qquad \ldots(1)$$

and define the quantiles X_{Gq} similarly.

If one knew the true quantile functions X_{Fq} and X_{Gq}, then one could compare them through graphs of the pairs,

$$\{(X_{Fq}, X_{Gq}), q \in [q_L, q_U]\} \qquad \ldots(2)$$

where q_L and q_U are specified endpoints such that $0 < q_L < q_U < 1$. Examination of the plot of (2) allows one to explore similarities and differences between the two distributions. For example, the plot of (2) follows a straight line if and only if $G(ax + b) = F(x)$ for some constants a and b. In addition, this straight line has slope 1 and intercept 0 if and only if $G(x) = F(x)$ for all $x \in [F^{-1}(q_L), F^{-1}(q_U)]$.

In some practical applications of (2), distributional questions center on the extent to which the distribution function $F(\cdot)$ for an observed variable is consistent with a specific parametric model $G(\cdot)$ or family of models. For example, Section 6.1 will consider measurements of bone mineral density (BMD). Historically, BMD measurements have been assumed to follow a normal distribution; see, e.g., Daniels *et al.* (1995, p. 360). This normality assumption, and associated z scores, have in turn been used to establish lower-tail cutoff points for formal definition of two medical conditions known as osteoporosis and osteopenia; and to assess of the prevalence of osteoporosis and osteopenia in specific populations. Thus, it is substantively important to assess the adequacy of the normal-distribution assumption for BMD measurements. For the BMD example, one may do this through graphical methods based on (2), with $F(\cdot)$ equal to the distribution function for BMD in the population of interest and with $G(\cdot)$ equal to $\Phi(\cdot)$, the standard normal distribution function.

In other cases, $F(\cdot)$ and $G(\cdot)$ are distribution functions for two specified subpopulations. For example, Section 6.2 will consider a case involving a blood-chemistry variable known as lipoprotein(a). In that application, $F(\cdot)$ and $G(\cdot)$ are the distribution functions of lipoprotein(a) for white non-Hispanic males and black non-Hispanic males, respectively; and plots based on (2) are intended to explore the extent to which the distributions of lipoprotein(a) differ across these two demographic groups.

1.2 *Offset function plots.* In some cases, comparison of the distributions F and G can be enhanced by examination of offset-function plots, sometimes known as shift-function plots. Following, e.g., Doksum and Sievers (1976), define the function

$$\Delta(x) = G^{-1}\{F(x)\} - x \qquad \ldots(3)$$

If two random variables X and Y have continuous distributions F and G, respectively, then the offset function $\Delta(x)$ is the only function of x such that $\Delta(X)+X$ and Y have the same distribution for all F and G.

As with quantile-quantile plots, offset-function plots can be used to compare one population distribution to a prespecified reference distribution. For example, if F is the normal distribution with mean μ and variance σ^2, and G is the standard normal distribution, then $\Delta(x) = (\sigma^{-1} - 1)x - \sigma^{-1}\mu$. Thus, for this case a straight-line plot of $\Delta(x)$ against x is consistent with a normal distribution $F(x)$.

Offset-function plots are also useful in comparison of two subpopulation distribution functions. For example, if $\Delta(x)$ equals some constant Δ_0 for all x, then $G(x + \Delta_0) = F(x)$ for all x; or equivalently, $X_{Gq} = X_{Fq} + \Delta_0$ for all $q \in (0, 1)$. In other words, the distribution function $G(\cdot)$ is the result of shifting $F(\cdot)$ to the right by Δ_0 units on the x scale (for $\Delta_0 \geq 0$); or shifting $F(\cdot)$ to the left by $-\Delta_0$ units (for $\Delta_0 < 0$). See, e.g., Oja (1981) for a detailed study of the interpretation of the offset function $\Delta(x)$ and related quantities.

Historically, offset functions have received relatively little attention in the sample survey literature. However, Sections 4 through 6 below will show that offset-function plots can provide a useful complement to quantile-quantile plots in the analysis of complex survey data. For example, practical comparisons of quantiles often focus on whether: (a) X_{Fq} and X_{Gq} are approximately equal over a specified interval $q \in [q_L, q_U]$; (b) the differences $X_{Fq} - X_{Gq}$ are nonzero but approximately constant over $q \in [q_L, q_U]$; or (c) $X_{Fq} - X_{Gq}$ displays a more complicated pattern. As indicated in the preceding two paragraphs, cases (a) through (c) correspond, respectively, to $\Delta(x) = 0$, $\Delta(x) = \Delta_0$ and nonconstant $\Delta(x)$. Consequently, offset-function plots allow direct exploration of cases (a) through (c) in a form that may be more appealing visually than quantile-quantile plots.

In addition, Section 5 will show that some technical results for offset function confidence bounds provide a useful intermediate step in the development of confidence bounds for quantile-quantile plots. Consequently, this paper will focus attention on both offset-function and quantile-quantile plots.

1.3 *Survey-weighted quantile and offset-function plots.* To implement the ideas of Sections 1.1 and 1.2 using sample data, one generally seeks: (a) to obtain estimators for some points along the curves (2) or (3); and (b) to construct associated pointwise or simultaneous confidence bounds at the same points. For observations that are independent and identically distributed, the statistical literature has developed a relatively large body of methods for construction of quantile plots, offset function plots and related probability plots. See, e.g., Doksum (1977), Harter (1984) and Gan *et al.* (1991) and references cited therein. Also, some authors have considered related plots for some special cases in which observations are not independent and identically distributed. See, e.g., Dempster and Ryan (1985) on weighted normal plots under a heteroscedastic one-way comparison model.

However, many biological and socioeconomic datasets are obtained through complex sample designs that are not consistent with the abovementioned *i.i.d.* or one-way modeling assumptions. This leads to two problems. First, direct application of *i.i.d*-based methods to complex survey data can lead to serious biases in quantile point estimators; and can also produce confidence bounds that do not fully reflect variance inflation induced by the complex design. Second, due to the first problem, *i.i.d*-based quantile or offset-function plots can give distorted comparisons of the distributions $F(\cdot)$ and $G(\cdot)$.

To address the first problem, the complex sample survey literature has developed a number of design-based methods for construction of point estimators and confidence sets for distribution functions $F(x)$ evaluated at specific points x, and for quantile functions X_{Fq} evaluated at specific points q. See, e.g., Woodruff (1952), Rao *et al.* (1990), Francisco and Fuller (1991) and McCarthy (1993).

The present paper addresses the second problem by applying some design-based quantile ideas to the the construction of quantile plots and offset function plots with appropriate simultaneous confidence bounds at a prespecified set of points. Section 2.1 reviews some of the abovementioned design-based methods. Section 2.2 reviews asymptotic properties of the resulting distribution function and quantile estimators. These properties are used through the remainder of the paper, and form the basis of each of the proposed inference methods. Section 2.3 outlines Scheffé and Bonferroni-type methods for the construction of simultaneous confidence bounds for $F(x)$ at a fixed number k of prespecified points x.

Section 3 reviews three methods for construction of pointwise confidence intervals for quantiles, and extends these ideas to construct Scheffé and Bonferroni-type simultaneous confidence intervals for quantiles x_{qi} at k prespecified values q_i. Section 4 applies the methods of Section 3 to plotting problems associated with a normal reference distribution. Section 4.1 develops point estimators and confidence bounds for the relevant offset-function plots. Section 4.2 considers construction of quantile-quantile plots with associated confidence bounds. In both cases, the proposed bounds follow directly from methods developed previously for distribution functions and quantiles estimated with complex survey data.

Section 5 considers related ideas for the case in which $F(\cdot)$ and $G(\cdot)$ are distribution functions associated with two subpopulations. Section 5.1 develops confidence bounds for the relevant offset-function plot. Section 5.2 extends the ideas of Section 5.1 to construct vertical confidence bounds associated with a fixed number of points on a q-q plot.

Section 6 applies the ideas of Sections 2 through 5 to some bone mineral density and lipoprotein(a) data collected in the U.S. Third National Health and Nutrition Examination Survey. Section 7 reviews the main ideas considered here and discusses some possible extensions. The Appendix reviews some technical background involving simultaneous confidence sets for vectors of parameters based on Scheffé and Bonferroni approaches.

2. Confidence Sets for Distribution Functions

2.1 *Estimation of distribution functions and quantiles under a complex design.* To begin the formal development of the proposed methods, consider a finite population of size N generated through N independent and identically distributed realizations of a superpopulation model. Realizations of this superpopulation model will produce random variables that are indicators for membership in one of three subpopulations identified with the labels 1, 2 and 3; and will also provide a continuous random variable denoted X. Let $F(\cdot)$ and $G(\cdot)$ be the superpopulation distribution functions for X, conditional on membership in subpopulations 1 and 2 respectively. The third subpopulation is a remainder group that will not be considered further here. Our inferential interest will focus primarily on the superpopulation distribution functions $F(\cdot)$ and $G(\cdot)$, and functions thereof. Ideas similar to those presented here apply to the associated finite-population distribution functions, but will not be considered in detail in this paper.

For our discussion here, our finite population is partitioned into L strata with N_h primary sample units contained in stratum h; and M_{hi} elements contained in primary unit (h, i). We select a sample of elements through a stratified multistage design, with n_h primary units selected from stratum h and n_{hi} elements selected from selected primary unit (h, i). Define $n = \sum_{h=1}^{L} n_h$. For sample element (hij) in the first subpopulation, let x_{hij} be the observed X value and let W_{hij} be the associated survey weight used in estimation of (sub)population totals. Then following, e.g., Francisco and Fuller (1991), a relatively simple design-based estimator of $F(x)$, the distribution function for the first subpopulation, is,

$$\hat{F}(x) = (\sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} W_{hij})^{-1} \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} W_{hij} \delta_{hij}(x) \qquad \ldots (4)$$

where $\delta_{hij}(x) = \{1 \text{ if } x_{hij} \leq x, \ 0 \ \text{ otherwise.}\}$

Also, for a fixed k-dimensional vector $x = (x_1, \ldots, x_k)$, define the distribution function vector

$$\hat{F}(x) = [\hat{F}(x_1), \ldots, \hat{F}(x_k)]^T. \qquad \ldots (5)$$

Note that each element of $\hat{F}(x)$ is a ratio of two estimated population totals. Thus, standard linearization methods lead to an estimator $\hat{V}\{\hat{F}(x)\}$, say, of the $k \times k$ dimensional covariance matrix of the approximate distribution of $\hat{F}(x) - F(x)$; see, e.g., Shao (1996, Section 2). The remainder of this paper will assume without further comment that all covariance matrices and their estimators are invertible. Cases in which these invertibility conditions are problematic are of practical interest, but are beyond the scope of the present work.

Now consider estimation of vectors of quantiles. For a given $q \in (0,1)$, application of definition (1) to the sample distribution function $\hat{F}(x)$ leads to the direct quantile estimator,

$$\hat{X}_{Fq} = \hat{F}^{-1}(q) = inf\{x : \hat{F}(x) \geq q\}. \qquad \ldots(6)$$

Similarly, for a fixed k-dimensional vector $q = (q_1, \ldots, q_k)^T$, $q_i \in (0,1), i = 1, \ldots, k$, define the quantile vector estimator,

$$\hat{X}_{Fq} = (\hat{X}_{Fq1}, \ldots, \hat{X}_{Fqk})^T = [\hat{F}^{-1}(q_1), \ldots, \hat{F}^{-1}(q_k)]^T. \qquad \ldots(7)$$

Section 4 will also consider estimators associated with the distribution function $G(\cdot)$. Thus, define the k-dimensional sample distribution function vector $\hat{G}(\cdot)$ in parallel with expression (5); and define the k-dimensional quantile vector estimator $\hat{Y}_{Gp} = (\hat{Y}_{Gp1}, \ldots, \hat{y}_{Gpk})^T = [\hat{G}^{-1}(p_1), \ldots, \hat{G}^{-1}(p_k)]^T$ where $(p_1, \ldots, p_k)^T$ is a second fixed k-dimensional vector, $p_i \in (0,1)$, not necessarily related to $(q_1, \ldots, q_k)^T$.

Finally, Francisco and Fuller (1991) considered an estimator of the variance of the approximate distribution of a vector of quantile estimators,

$$\hat{V}(\hat{X}_{Fq}) = H^T \hat{V}\{\hat{F}(\hat{X}_{Fq})\} H \qquad \ldots(8)$$

and H is a diagonal matrix with i-th element

$$\begin{aligned} h_i &= \left[\hat{F}^{-1}[q_i + t_{1-\alpha/2,n-L}\{\hat{V}(\hat{F}(\hat{X}_{Fq}))\}_{ii}^{1/2}] - \hat{F}^{-1}[q_i - t_{1-\alpha/2,n-L}\{\hat{V}(\hat{F}(\hat{X}_{Fq}))\}_{ii}^{1/2}]\right] \\ &\cdot \left[2t_{1-\alpha/2,n-L}\{\hat{V}(\hat{F}(\hat{X}_{Fq})\}_{ii}^{1/2}\right]^{-1} \end{aligned}$$

The remainder of this paper will use covariance matrix estimators $\hat{V}\{(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T\}$ defined analogously.

2.2 *Assumed asymptotic properties.* Sections 3 and 4 below will develop inference methods based primarily on large-sample approximations for the distributions of the vectors $[\hat{F}(x)^T, \hat{G}(x)^T]^T$ and $(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)$. In this subsection we will state the principal large-sample approximations used in the remainder of the paper. Detailed justification of these approximations is beyond the scope of the present work; see, e.g., Francisco and Fuller (1991) and Shao (1996) for some general development of the approximate sampling distributions of distribution-function and quantile vectors. Throughout this discussion, the term "approximate distribution" refers to convergence in distribution of the random vectors or random matrices in question. For these approximations, let m be the total number of elements selected for the sample.

First, under moderate regularity conditions (see, e.g., Francisco and Fuller, 1991), we have the following normal approximation for a vector pivotal quantity based on the $2k$-dimensional estimator $[\hat{F}(x)^T, \hat{G}(y)^T]$, an associated covariance matrix estimator, $\hat{V}\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]$ and the inverse of its symmetric square root $[\hat{V}\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]^{-1/2}$.

(A.1) The vector $[\hat{V}\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]^{-1/2}[\{\hat{F}(x)-F(x)\}^T, \{\hat{G}(y)-G(y)\}^T]^T$ is approximately distributed as a normal random vector with mean 0 and covariance matrix equal to the $2k \times 2k$ dimensional identity matrix, I_{2k}.

Note that under approximation (A.1), pointwise approximate $(1-\alpha)100\%$ confidence intervals for a given value $F(x_i)$ follow immediately from customary normal-approximation methods. However, if $n-L$ is not large relative to the dimension $2k$, then the variability of $\hat{V}[\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]$ may make a nontrivial contribution to the variability of $[\hat{V}\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]^{-1/2}[\{\hat{F}(x) - F(x)\}^T, \{\hat{G}(y) - G(y)\}^T]^T$, which in turn makes approximation (A.1) potentially problematic. Thus, some of our work will use an alternative approximation intended to account explicitly for the variability of $\hat{V}[\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]$.

(A.2) The random vector $m^{1/2}[\{\hat{F}(x)-F(x)\}^T, \{\hat{G}(y)-G(y)\}^T]^T$ is approximately distributed as a normal random vector with mean 0 and covariance matrix $mV[\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]$. In addition, the $2k \times 2k$ matrix $(n-L)m\hat{V}[\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]$ is approximately distributed as a Wishart$(mV[\{\hat{F}(x)^T, \hat{G}(y)^T\}^T], n-L)$ random matrix and is approximately independent of $\{\hat{F}(x)^T, \hat{G}(y)^T\}^T$.

For some general background on Wishart approximations for design-based covariance matrices, see, e.g., Korn and Graubard (1990).

In parallel with approximations (A.1) and (A.2), we also will use the following large-sample approximations for the quantile estimator vectors $(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T$ and associated covariance matrices and pivotal quantities. For example, Francisco and Fuller (1991, Theorem 4) showed that under regularity conditions, $\{\hat{V}(\hat{X}_{Fq})\}^{-1/2}(\hat{X}_{Fq} - X_{Fq})$ is distributed approximately as a k-dimensional normal random vector with mean 0 and covariance matrix I, where $\hat{V}(\hat{X}_{Fq})$ is defined in expression (8) above. Approximation (A.3) is an expanded version of this idea, applied to the combined vector, $(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T$.

(A.3) For fixed vectors $(q_1, \ldots, q_k)$ and $(p_1, \ldots, p_k)$, the random vector $\{\hat{V}(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T\}^{-1/2}$ $\{(\hat{X}_{Fq} - X_{Fq})^T, (\hat{Y}_{Gp} - Y_{Gp})^T\}^T$ is approximately distributed as a normal random vector with mean 0 and covariance matrix equal to the $2k \times 2k$ dimensional identify matrix, I_{2k}.

Also, approximation (A.4) expands approximation (A.3) to account explicitly for nontrivial random variability in the variance estimator $\hat{V}\{(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T\}$.

(A.4) The random vector $m^{1/2}\{(\hat{X}_{Fq} - X_{Fq})^T, (\hat{Y}_{Gp} - Y_{Gp})^T\}^T$ is approximately distributed as a normal random vector with mean 0 and covariance matrix $mV\{(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T\}$. In addition, the $2k \times 2k$ matrix $(n-L)m\hat{V}\{(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T\}$ is approximately distributed as a Wishart$[mV\{(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T\}, n-L]$ random matrix and is approximately independent of $\{(\hat{X}_{Fq} - X_{Fq})^T, (\hat{Y}_{Gp} - Y_{Gp})^T\}^T$.

Finally, note that the approximations listed above involve the combined vectors $[\hat{F}(x)^T, \hat{G}(y)^T]^T$ and $(\hat{X}_{Fq}^T, \hat{Y}_{Gp}^T)^T$. Section 5 will use these full vectors, while Sections 2 through 4 will use only the subvectors associated with the distribution function F.

2.3 *Scheffé and Bonferroni approximate simultaneous confidence sets for distribution functions.* Now consider the use of approximation (A.2) for construction of simultaneous confidence sets for the vector of distribution function values $F(x) = \{F(x_1), \ldots, F(x_k)\}^T$ for a predetermined vector $x = (x_1, \ldots, x_k)^T$. Routine arguments (e.g., Korn and Graubard, 1990) show that under approximation (A.1), a Scheffé-type approximate $(1-\alpha)100\%$ confidence set for the k-dimensional vector $F(x)$ is,

$$\begin{aligned} S_F &= \{F = (F_1, \ldots, F_k) : [\hat{F}(x) - F]^T [\hat{V}\{\hat{F}(x)\}]^{-1} [\hat{F}(x) - F] \\ &\leq \tfrac{k(n-L)}{n-L-k+1} F_{k,n-L-k+1}(1-\alpha)\}. \end{aligned} \quad \ldots (9)$$

Consequently, simultaneous $(1-\alpha)100\%$ confidence bounds for $F(x_i), i = 1, \ldots, k$, are

$$(\hat{F}_{SLi}, \hat{F}_{SUi}) = \hat{F}(x_i) \pm [\frac{k(n-L)}{n-L-k+1} F_{k,n-L-k+1}(1-\alpha) \hat{V}\{\hat{F}(x_i)\}]^{\frac{1}{2}}, \quad i = 1, \ldots, k \quad \ldots (10)$$

The approximation (A.2) also leads to Bonferroni-type simultaneous confidence intervals for the k values $F(x_i)$,

$$(\hat{F}_{BLi}, \hat{F}_{BUi}) = \hat{F}(x_i) \pm \ t_{1-\alpha/2k, n-L} [\hat{V}\{\hat{F}(x_i)\}]^{\frac{1}{2}}, \quad i = 1, \ldots, k \quad \ldots (11)$$

where $t_{1-\alpha/2k, n-L}$ is the the $(1-\alpha/2k)$ quantile value of t distribution with $n-L$ degree of freedom.

Following, e.g., Korn and Graubard (1990) for Bonferroni-based analyses involving mean vectors, and Krieger and Pfeffermann (1997, p. 131) for Bonferroni-based comparisons of distribution functions note that for a fixed number k of predetermined points, $F(x_i)$, $i = 1, \ldots, k$, Bonferroni-type simultaneous approximate $(1-\alpha)100\%$ confidence intervals generally will be somewhat narrower than the associated Scheffé intervals. The appendix gives an illustration of this phenomenon. However, in some cases there is substantive interest in functions of $F(x)$ or x_{Fq} that are more complicated than the individual components $F(x_i)$ or X_{Fqi}, and Scheffé approaches are of interest for some of these functions. Consequently, the present work with quantiles and offset functions will consider both Scheffé and Bonferroni based methods.

3. Approximate Confidence Sets for Quantiles

3.1 *The Francisco and Fuller (1991) test inversion method.* Pointwise confidence intervals for quantiles can be obtained through either the Woodruff (1952) method or the Francisco and Fuller (1991) method. For example, a confidence set for the qth population quantile value X_{Fq} can be obtained by applying the Francisco and Fuller (1991) test inversion arguments to the confidence set for $F(x)$. For the present multivariate case, one set of simultaneous confidence intervals may be obtained through the Scheffé method. Specifically, define the test inversion $(1-\alpha)100\%$ confidence set for X_{Fq} as,

$$\begin{aligned} S_{TI} &= \{x = (x_1, \ldots, x_k) : [\hat{F}(x) - q]^T [\hat{V}\{\hat{F}(x)\}]^{-1} [\hat{F}(x) - q] \\ &\leq \tfrac{k(n-L)}{n-L-k+1} F_{k,n-L-k+1}(1-\alpha)\}, \end{aligned} \qquad \ldots(12)$$

where $x = (x_1, \ldots, x_k)$ and $q = (q_1, \ldots, q_k)$ are k-dimensional vectors.

Direct implementation of the test inversion confidence set S_{TI} is nontrivial since $\hat{V}\{\hat{F}(x)\}$ may vary with x. Francisco and Fuller (1991, Section 4) suggested an algorithm to construct pointwise confidence intervals for $F(x)$ at a prespecified $x = x_i$. For a given point estimator $\hat{F}(\cdot)$, variance estimator v_i, multiplier γ and observed ordered values $x_{(i)}$, define

$$\begin{aligned} \hat{F}_{MU}(\hat{F}, \gamma, v_i, x_{(i)}) &= \hat{F}(x_{(i)}) + \gamma(v_i)^{1/2} && \text{for } i = 1 \\ &= \max\{\hat{F}_{MU}(\hat{F}, \gamma, v_i, x_{(i-1)}), \hat{F}(x_{(i)}) + \gamma(v_i)^{1/2}\} && \text{for } i > 1 \\ \hat{F}_{ML}(\hat{F}, \gamma, v_i, x_{(i)}) &= \hat{F}(x_{(i)}) - \gamma(v_i)^{1/2} && \text{for } i = m \\ &= \min\{\hat{F}_{ML}(\hat{F}, \gamma, v_i, x_{(i+1)}), \hat{F}(x_{(i)}) - \gamma(v_i)^{1/2}\} && \text{for } i < m \end{aligned} \qquad \ldots(13)$$

where m is the total number of ultimate units in the sample, $\gamma = t_{1-\alpha/2, n-L}$, and the subscript M refers to the fact that these upper and lower bounds are constrained to be monotone nondecreasing in x.

In parallel with these bounds, Francisco and Fuller (1991) developed approximate $(1-\alpha)100\%$ pointwise confidence intervals for X_{Fq_i}, $[\hat{F}_{MU}^{-1}(q_i),\ \hat{F}_{ML}^{-1}(q_i)]$, $i = 1, \ldots, k$. Francisco and Fuller (1991) showed that under regularity conditions, the Woodruff confidence interval and Francisco and Fuller confidence interval are asymptotically equivalent, and that

$$P[F_{MU}^{-1}(q_i) \leq X_{q_i} \leq F_{ML}^{-1}(q_i)] \xrightarrow{p} 1 - \alpha$$

The pointwise Francisco-Fuller algorithm extends to the k-dimensional case in two ways. First, for a Scheffé-type confidence interval, define the interval $[\hat{F}_{MSL}(x), \hat{F}_{MSU}(x)]$ by expression (12) with $\gamma = \{k(n-L)/(n-L-k+1) \cdot F_{k,n-L-k+1}(1-\alpha)\}^{1/2}$. The resulting Francisco-Fuller-Scheffé type approximate $(1-\alpha)100\%$ simultaneous confidence intervals for $x_{q1}, \ldots, x_{qk}$ are

$$[\hat{F}_{MSU}^{-1}(q_i),\ \hat{F}_{MSL}^{-1}(q_i)], \quad i = 1, \ldots, k \qquad \ldots(14)$$

A similar procedure leads to Bonferroni confidence intervals $[\hat{F}_{MBL}(x), \hat{F}_{MBU}(x)]$, say, constructed from expression (12) with $\gamma = t_{1-\alpha/2k, n-L}$. The resulting approximate $(1-\alpha)100\%$ simultaneous quantile confidence intervals for $X_{Fq1}, \ldots, X_{Fqk}$

are

$$[\hat{F}^{-1}_{MBU}(q_i),\ \hat{F}^{-1}_{MBL}(q_i)], \quad i = 1, \ldots, k \qquad \ldots(15)$$

3.2 *A Woodruff type method.* Note that for S_{TI} defined in expression (8), the variance estimator $\hat{V}\{\hat{F}(x)\}$ may vary with x. Consequently, it is problematic to derive closed-form expressions for simultaneous confidence intervals from the confidence set S_{TI}. Thus, we use a local-variance-approximation method by additionally assuming that $V\{\hat{F}(x)\}$ is approximately constant in a neighborhood of X_q. From Francisco and Fuller (1991, condition 7), $\hat{V}\{\hat{F}(\hat{X}_{Fq})\}$ is a consistent estimator of $V\{\hat{F}(X_{Fq})\}$ and so we use $\hat{V}\{\hat{F}(\hat{X}_{Fq})\}$ as a rough estimator of $V\{\hat{F}(x)\}$ through that neighborhood. Then an approximate $(1-\alpha)$ confidence set of the S_{TI} is,

$$\begin{aligned} S_W \;=\; & \{x = (x_1, \ldots, x_k) : [\hat{F}(x) - q]^T [\hat{V}\{\hat{F}(\hat{X}_{Fq})\}]^{-1} [\hat{F}(x) - q] \\ & \leq \tfrac{k(n-L)}{n-L-k+1}\ F_{k,n-L-k+1}(1-\alpha)\}. \end{aligned} \qquad \ldots(16)$$

Let

$$\begin{aligned} S_{WF} \;=\; & \{F = (F_1, \ldots, F_k) : [F - q]^T [\hat{V}\{\hat{F}(\hat{X}_{Fq})\}]^{-1} [F - q] \\ & \leq \tfrac{k(n-L)}{n-L-k+1}\ F_{k,n-L-k+1}(1-\alpha)\} \end{aligned} \qquad \ldots(17)$$

and note that $S_W = \hat{F}^{-1}(S_{WF})$. For the individual values $q_i, i = 1, \ldots, k$, we can construct simultaneous confidence intervals from the confidence set S_{WF}.

A direct application of the Scheffé method leads to the Woodruff-Scheffé approximate $(1-\alpha)100\%$ simultaneous confidence intervals, $\Big(\hat{F}^{-1}[q_i - \{k(n-L)/(n-L-k+1) \cdot F_{k,n-L-k+1}(1$

$\hat{F}^{-1}[q_i + \{k(n-L)/(n-L-k+1) \cdot F_{k,n-L-k+1}(1-\alpha)\}^{1/2} \{\hat{V}(\hat{F}(\hat{X}_{Fq}))\}^{1/2}_{ii}]\Big)$,

$i = 1, \ldots, k$ $\quad \ldots(18)$ Related Bonferroni approximate $(1-\alpha)$ simultaneous quantile confidence intervals are constructed from the confidence set S_{WF} by,

$$\Big(\hat{F}^{-1}[q_i - t_{1-\alpha/2k,n-L}\{\hat{V}(\hat{F}(\hat{X}_{Fq}))\}^{1/2}_{ii}],\ \ \hat{F}^{-1}[q_i + t_{1-\alpha/2k,n-L}\{\hat{V}(\hat{F}(\hat{X}_{Fq}))\}^{1/2}_{ii}]\Big),$$

$$i = 1, \ldots, k \qquad \ldots(19)$$

3.3 *A direct normal approximation method.* Under the normal approximation (A.3), direct application of the test-inversion reasoning from Section 3.1 leads to an approximate $(1-\alpha)100\%$ confidence set,

$$\begin{aligned} S_x \;=\; & \{x = (x_1, \ldots, x_k) : (\hat{X}_{Fq} - x)^T \{\hat{V}(\hat{X}_{Fq})\}^{-1} (\hat{X}_{Fq} - x) \\ & \leq \chi^2_k(1-\alpha)\} \end{aligned} \qquad \ldots(20)$$

where $\chi^2_k(1-\alpha)$ is the $1-\alpha$ quantile of a chi-square distribution on k degrees of freedom.

Note that expression (20) does not account explicitly for the random variability of the variance estimator $\hat{V}(\hat{x}_q)$. To reflect this variability, note that under approximation (A.4) modification of expression (20) leads to the Scheffé-type approximate $(1-\alpha)100\%$ confidence set,

$$\begin{aligned} S_x \;=\; & \{x=(x_1,\ldots,x_k):(\hat{X}_{Fq}-x)^T[\hat{V}(\hat{X}_{Fq})]^{-1}(\hat{X}_{Fq}-x) \\ & \le \tfrac{k(n-L)}{n-L-k+1}\,F_{k,n-L-k+1}(1-\alpha)\}. \end{aligned} \qquad \ldots(21)$$

Associated Bonferroni approximate $(1-\alpha)$ simultaneous confidence intervals are

$$\hat{X}_{Fq_i} \pm\; t_{1-\alpha/2k,n-L}[\{\hat{V}(\hat{X}_{Fq})\}_{ii}]^{1/2}\;,\;\; i=1,\ldots,k \qquad \ldots(22)$$

4. Comparison to a Known Reference Distribution

4.1 *Offset functions based on a normal reference distribution.* Define $\Delta_\Phi(x)$ to be the offset function (3) for the special case in which $G(\cdot)$ is the standard normal distribution function. A direct estimator of $\Delta_\Phi(x)$ is,

$$\hat{\Delta}_\Phi(x)=\Phi^{-1}\{\hat{F}(x)\}-x \qquad \ldots(23)$$

where Φ is the distribution function of the standard normal distribution.

Since the function $\Phi(\cdot)$ is not random, pointwise and simultaneous confidence bounds for $\Delta_\Phi(x)$ follow directly from the confidence bounds for $F(x)$ outlined in Sections 2 and 3. For example, let $x_0=(x_1,\ldots,x_k)$ be a prespecified k-dimensional random vector, and define the associated k-dimensional $(1-\alpha)100\%$ confidence set S_F for $F(x_0)$ by expression (5). Then $\{\Phi^{-1}(F)-x_0 : F=(F_1,\ldots,F_k)^T\in S_F\}$ gives an approximate $(1-\alpha)100\%$ confidence set of the vector $\Delta_\Phi(x_0)=[\Delta_\Phi(x_1),\ldots,\Delta_\Phi(x_k)]^T$. Similar arguments lead to simultaneous confidence bounds for the k values $\Delta_\Phi(x_i)$, $i=1,\ldots,k$. For example, approximate Bonferroni $(1-\alpha)$ simultaneous confidence intervals are

$$\begin{aligned} &\Big[\Phi^{-1}[\hat{F}(x_i)-t_{1-\alpha/2k,n-L}[\hat{V}\{\hat{F}(x_0)\}]^{1/2}_{ii}] \\ &-x_i,\Phi^{-1}[\hat{F}(x_i)+t_{1-\alpha/2k,n-L}[\hat{V}\{\hat{F}(x_0)\}]^{1/2}_{ii}]-x_i\Big], \\ &i=1,\ldots,k \end{aligned} \qquad \ldots(24)$$

Similary, approximate Scheffé $(1-\alpha)100\%$ simultaneous confidence intervals are,

$$\begin{aligned} &[\Phi^{-1}[\hat{F}(x_i)-\{k(n-L)/(n-L-k+1) \\ &\cdot F_{k,n-L-k+1}(1-\alpha)\}^{1/2}[\hat{V}\{\hat{F}(x_0)\}]^{1/2}_{ii}]-x_i, \\ &\Phi^{-1}[\hat{F}(x_i)+\{k(n-L)/(n-L-k+1) \\ &\cdot F_{k,n-L-k+1}(1-\alpha)\}^{1/2}[\hat{V}\{\hat{F}(x_0)\}]^{1/2}_{ii}]-x_i], \\ &i=1,\ldots,k \end{aligned} \qquad \ldots(25)$$

Note here that simultaneous confidence intervals for $\Delta_\Phi(x_i)$ depend only on the fixed values of x_i and the random bounds $(\hat{F}_{Li}, \hat{F}_{Ui})$.

4.2 *Quantile-quantile plots.* Now consider a quantile-quantile plot in which the standard normal quantiles $\Phi^{-1}(q_i)$ are plotted along the horizontal axis for selected $q_i, i = 1, \ldots, k$; and the associated sample quantiles $\hat{F}^{-1}(q_i)$ are plotted along the vertical axis. In parallel with the reasoning used in Section 4.1, note that the horizontal-axis standard normal quantiles are predetermined. Thus, pointwise or simultaneous confidence bounds follow directly from plotting along the vertical axis the quantile confidence bounds outlined in Section 3. Similar comments apply to comparison of the quantiles $\hat{F}^{-1}(q_i)$ to the quantiles of other predetermined reference distributions.

5. Comparison of Two Subpopulation Distribution Functions

5.1 *Use of offset functions to compare subpopulations.* We now consider the use of offset functions to compare subpopulation distribution functions. Let $x_0 = (x_1, \ldots, x_k)$ be a vector of prespecified points and consider development of a test inversion confidence set for $\Delta(x) = G^{-1}\{F(x)\} - x$. The hypothesis is of interest that for some fixed vector $d = (d_1, \ldots, d_k)$,

$$H_0 : G(x_0 + d) = F(x_0)$$

Let $\hat{H}(x_0, d)^T = [\hat{G}(x_0 + d)^T, \hat{F}(x_0)^T]$ and $H(x_0, d)^T = [G(x_0 + d)^T, F(x_0)^T]$.

Then $\hat{H}(x_0, d)$ is a vector of weighted sample ratios. Thus, under regularity conditions, asymptotic normality properties of mean vectors (see, e.g., Francisco and Fuller 1991, Theorem 1; and Krewski and Rao 1981, Theorem 3.1), indicate that $\hat{H}(x_0, d)$ is distributed approximately as a $2k$-dimensional normal random vector with mean $H(x_0, d)$ and covariance matrix $V\{\hat{H}(x_0, d)\}$, where

$$V\{\hat{H}(x_0, d)\} = \begin{pmatrix} V\{\hat{G}(x_0 + d)\} & COV\{\hat{G}(x_0 + d), \hat{F}(x_0)\} \\ COV\{\hat{F}(x_0), \hat{G}(x_0 + d)\} & V\{\hat{F}(x_0)\} \end{pmatrix}$$

Let $A = [I_k, -I_k]$, where I_k is the k-dimensional identity matrix. Then $A\hat{H}$ follows approximately a k-dimensional normal distribution with mean AH and covariance matrix $AV\{\hat{H}(x_0, d)\}A^T$. With the assumption that $(n-L)\hat{V}\{\hat{H}(x_0, d)\}$ follows an approximate $2k$-dimensional $Wishart$ $[V\{\hat{H}(x_0, d)\}, n - L]$ distribution and is approximately independent of $\hat{H}(x_0, d)$, $(n - L)A\hat{V}\{\hat{H}(x_0, d)\}$ has an approximate k-dimensional $Wishart[AV\{\hat{H}(x_0, d)\}A^T, n - L]$ distribution and is approximately independent of $A\hat{H}(x_0, d)$. Then, under H_0, $(n - L - k + 1)/(k(n - L)) \cdot [\hat{G}(x_0 + d) - \hat{F}(x_0)]^T [A\hat{V}\{\hat{H}(x_0, d)\}A^T]^{-1} [\hat{G}(x_0 + d) - \hat{F}(x_0)]$ is distributed approximately as an F random variable with k and $n - L - k + 1$ degrees of freedom. These results, and additional routine arguments, indicate that a test-inversion-based approximate $(1-\alpha)100\%$ confidence set for Δ is defined by

$$\begin{aligned} S_{TI\Delta} \;=\; & \{d=(d_1,\ldots,d_k):[\hat{G}(x_0+d)-\hat{F}(x_0)]^T \\ & \cdot[\hat{V}\{\hat{G}(x_0+d)-\hat{F}(x_0)\}]^{-1}[\hat{G}(x_0+d)-\hat{F}(x_0)] \\ & \leq \tfrac{k(n-L)}{n-L-k+1}\, F_{k,n-L-k+1}(1-\alpha)\} \end{aligned} \qquad \ldots(26)$$

5.1.1 *Local approximation (Woodruff type) confidence sets.* It is nontrivial to directly derive simultaneous confidence intervals from confidence set $S_{TI\Delta}$ because $\hat{V}\{\hat{G}(x_0+d)-\hat{F}(x_0)\}$ may vary with d. One way to addrsss this issue is to construct Woodruff type confidence intervals based on a local variance approximation. We first consider a null hypothesis $H_0:\Delta(x_i)=\Delta_0(x_i)$, $i=1,\ldots,k$, and assume that $V\{\hat{G}(x_0+d)-\hat{F}(x_0)\}$ is approximately constant for all d in some neighborhood of $\Delta_0=[\Delta_0(x_1),\ldots,\Delta_0(x_k)]$. This local variance approximation leads to an approximate $(1-\alpha)100\%$ confidence set,

$$\begin{aligned} S_{W\Delta} \;=\; & \{d=(d_1,\ldots,d_k):[\hat{G}(x_0+d)-\hat{F}(x_0)]^T \\ & \cdot[\hat{V}\{\hat{G}(x_0+\Delta_0)-\hat{F}(x_0)\}]^{-1}[\hat{G}(x_0+d)-\hat{F}(x_0)] \\ & \leq \tfrac{k(n-L)}{n-L-k+1}\, F_{k,n-L-k+1}(1-\alpha)\} \end{aligned} \qquad \ldots(27)$$

In addition, note that $S_{W\Delta}=\hat{G}^{-1}(S_{WG})-x_0$, where

$$\begin{aligned} S_{WG} \;=\; & \{G=(G_1,\ldots,G_k):[G-\hat{F}(x_0)]^T \\ & \cdot[\hat{V}\{\hat{G}(x_0+\Delta_0)-\hat{F}(x_0)\}]^{-1}[G-\hat{F}(x_0)] \\ & \leq \tfrac{k(n-L)}{n-L-k+1}\, F_{k,n-L-k+1}(1-\alpha)\} \end{aligned} \qquad \ldots(28)$$

Consequently, approximate $(1-\alpha)100\%$ Woodruff-Bonferroni type simultaneous confidence intervals for $\Delta(x_0)$ are,

$$\begin{aligned} \Big[& \hat{G}^{-1}[\hat{F}(x_i)-t_{1-\alpha/2k,n-L}[\hat{V}\{\hat{G}(x_0+\Delta_0)-\hat{F}(x_0)\}]_{ii}^{1/2}]-x_i, \\ & \hat{G}^{-1}[\hat{F}(x_i)+t_{1-\alpha/2k,n-L}[\hat{V}\{\hat{G}(x_0+\Delta_0)-\hat{F}(x_0)\}]_{ii}^{1/2}]-x_i \Big],\quad i=1,\ldots,k \end{aligned} \qquad \ldots(29)$$

Woodruff-Scheffé type simultaneous confidence intervals can be obtained by replacing $t_{1-\alpha/2k,n-L}$ with $\{k(n-L)/(n-L-k+1)\cdot F_{k,n-L-k+1}(1-\alpha)\}^{1/2}$. In the preceding confidence interval expression, if $\Delta_0(x_i)$ is not prespecified or predetermined, then one may use $\hat{V}\{\hat{G}(x_0+\hat{\Delta})-\hat{F}(x_0)\}$ as a rough estimate of $V\{\hat{G}(x_0+\Delta)-\hat{F}(x_0)\}$ throughout that neighborhood. Under regularity conditions, $\hat{V}\{\hat{G}(x_0+\hat{\Delta})-\hat{F}(x_0)\}$ is a consistent estimator of $V\{\hat{G}(x_0+\Delta)-\hat{F}(x_0)\}$.

5.1.2 *Francisco-Fuller type method.* We first consider a pointwise confidence interval and then consider extensions to the k-dimensional case. In particular, a variant on the Francisco-Fuller algorithm for quantiles can be applied to develop confidence intervals for an offset function. Define

$$\begin{aligned}
\hat{G}_U(y_{(i)}) &= \hat{G}(y_{(i)}) + t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(y_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2}, \text{ for } i = 1 \\
&= \max\Big\{\hat{G}_U(y_{(i-1)}), \hat{G}(y_{(i)}) \\
&\quad + t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(y_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2}\Big\}, \text{ for } i > 1 \\
\hat{G}_L(y_{(i)}) &= \hat{G}(y_{(i)}) - t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(y_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2}, \text{ for } i = m \\
&= \min\Big\{\hat{G}_L(y_{(i+1)}), \hat{G}(y_{(i)}) \\
&\quad - t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(y_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2}\Big\}, \text{ for } i < m
\end{aligned} \qquad \ldots(30)$$

where $x_{(1)}, \ldots, x_{(m)}$ are ordered observed values, $y_{(i)} = \hat{G}^{-1}\{\hat{F}(x_{(i)}\}$, $i = 1, \ldots, m$, and m is the total number of ultimate units in the sample. Then Francisco-Fuller type approximate $(1-\alpha)100\%$ pointwise confidence intervals for $\Delta(x_0)$ are

$$[\hat{G}_U^{-1}\{\hat{F}(x_i)\} - x_i,\ \hat{G}_L^{-1}\{\hat{F}(x_i)\} - x_i], \quad i = 1, \ldots, k \qquad \ldots(31)$$

Note that Scheffé-type confidence sets can be obtained from the preceding work by replacing $t_{1-\alpha/2,n-L}$ with an associated critical value based on the F distribution.

5.2 *Extension of offset function confidence intervals to construct confidence bounds for Q-Q plots.* Pointwise or simultaneous confidence intervals for a quantile-quantile plot can be developed in a form similar to confidence intervals for offset function. Once again, we will consider inversion of a test of the hypothesis $H_0 : G(x_0 + d) = F(x_0)$. To do this, define $\lambda = x_0 + d$ and define the function, $\Lambda(x_0) = G^{-1}\{F(x_0)\}$. Then a test inversion $(1-\alpha)100\%$ approximate confidence set for Λ is,

$$\begin{aligned}
S_{TI\Lambda} = \{&\lambda = (\lambda_1, \ldots, \lambda_k) : [\hat{G}(\lambda) - \hat{F}(x_0)]^T \\
&\cdot[\hat{V}\{\hat{G}(\lambda) - \hat{F}(x_0)\}]^{-1}[\hat{G}(\lambda) - \hat{F}(x_0)] \\
&\le \tfrac{k(n-L)}{n-L-k+1}\ F_{k,n-L-k+1}(1-\alpha)\}
\end{aligned} \qquad \ldots(32)$$

Woodruff type confidence intervals can be developed using the same local-variance assumption as used for an offset function. Specifically, under a null hypothesis $H_0 : \Lambda(x_i) = \Lambda_0(x_i), i = 1, \ldots, k$ and additional regularity conditions $V\{\hat{G}(\lambda) - \hat{F}(x_0)\}$ is approximately constant for all λ in some neighborhood of Λ_0. In addition, we will use $\hat{V}\{\hat{G}(\hat{\lambda}) - \hat{F}(x_0)\}$ as an estimate of $\hat{V}\{\hat{G}(\lambda) - \hat{F}(x_0)\}$. Here, $x_0 = [\hat{F}^{-1}(q_1), \ldots, \hat{F}^{-1}(q_k)]$ and the associated vector of estimated quantile from $\hat{G}(\cdot)$, $\hat{\lambda} = \hat{G}^{-1}(x_0) = [\hat{G}^{-1}(x_1), \ldots, \hat{G}^{-1}(x_k)]$. Then repetition of arguments similar to those used in Section 5.1 lead to Woodruff-Bonferroni type approximate $(1-\alpha)100\%$ simultaneous confidence intervals,

$$\begin{aligned}
[\quad &\hat{G}^{-1}[\hat{F}(x_i) - t_{1-\alpha/2k,n-L}[\hat{V}\{\hat{G}(\hat{\lambda}) - \hat{F}(x_0)\}]_{ii}^{1/2}], \\
&\hat{G}^{-1}[\hat{F}(x_i) + t_{1-\alpha/2k,n-L}[\hat{V}\{\hat{G}(\hat{\lambda}) - \hat{F}(x_0\}]_{ii}^{1/2}] \quad], \ i = 1, \ldots, k.
\end{aligned} \qquad \ldots(33)$$

The Francisco-Fuller type confidence intervals for a quantile-quantile plot are obtained from the same basic Francisco-Fuller algorithm used for the offset function in Section 4.1. Specifically, define

$$\begin{aligned}
\hat{G}_{\Lambda U}(\hat{\lambda}_{(i)}) &= \hat{G}(\hat{\lambda}_{(i)}) + t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(\hat{\lambda}_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2} \text{ for } i = 1 \\
&= \max\Big\{\hat{G}_{\Lambda U}(\hat{\lambda}_{(i-1)}), \hat{G}(\hat{\lambda}_{(i)}) \\
&\quad + t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(\hat{\lambda}_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2}\Big\} \text{ for } i > 1 \\
\hat{G}_{\Lambda L}(\hat{\lambda}_{(i)}) &= \hat{G}(\hat{\lambda}_{(i)}) - t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(\hat{\lambda}_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2} \text{ for } i = m \\
&= \min\Big\{\hat{G}_{\Lambda L}(\hat{\lambda}_{(i+1)}), \hat{G}(\hat{\lambda}_{(i)}) \\
&\quad - t_{1-\alpha/2,n-L}[\hat{V}\{\hat{G}(\hat{\lambda}_{(i)}) - \hat{F}(x_{(i)})\}]^{1/2}\Big\} \text{ for } i < m
\end{aligned} \quad \ldots(34)$$

where $x_{(1)}, \ldots, x_{(m)}$ are ordered observed values, $\hat{\lambda}_{(i)} = \hat{G}^{-1}\{\hat{F}(x_{(i)})\}$, $i = 1, \ldots, m$, and m is the total number of ultimate units in the sample. The resulting pointwise confidence intervals for Λ are,

$$[\hat{G}_{\Lambda U}^{-1}\{\hat{F}(x_i)\},\ \hat{G}_{\Lambda L}^{-1}\{\hat{F}(x_i)\}], \quad i = 1, \ldots, k \qquad \ldots(35)$$

For the k-dimensional case, one may obtain related Scheffé-type confidence intervals by replacing $t_{1-\alpha/2,n-L}$ with an associated critical value based on the F distribution.

6. Application to NHANES III Data

6.1 *The third national health and nutrition examination survey.* The Third National Health and Nutrition Examination Survey (NHANES III) was a large-scale stratified multistage sample survey carried out over 1988-1994. This survey was intended to assess the health and nutritional status of the U.S. civilian non-institutionalized population. There was special interest in estimation of parameters of certain demographic subpopulations, e.g., non-Hispanic blacks, Mexican-Americans, children and senior citizens. Consequently, sampling probabilities were determined such that members of these special demographic groups were oversampled relative to the balance of the population. Consequently, there was substantial variability in selection probabilities across different sample persons. Survey weights were computed to account for these unequal probabilities, and also included poststratification and nonresponse adjustments. For additional general background on NHANES III, see National Center for Health Statistics (1996).

For purposes of design based analysis, the NHANES III design generally is treated as arising from 49 strata with 2 primary sample units selected per stratum. Each primary unit was roughly equivalent to a county or a portion of a large metropolitan county. Within a selected primary unit, additional stages of sampling were carried out at the level of segments (small groups of households), households and individual persons. Standard ratio-type design-based methods were used to compute the relevant point estimators and variance estimators for the subpopulation distribution functions $F(\cdot)$ and $G(\cdot)$ of interest.

In keeping with the discussion in Section 1.1, the present paper will focus on subpopulation distribution functions and related quantities for two variables: bone mineral density and lipoprotein(a).

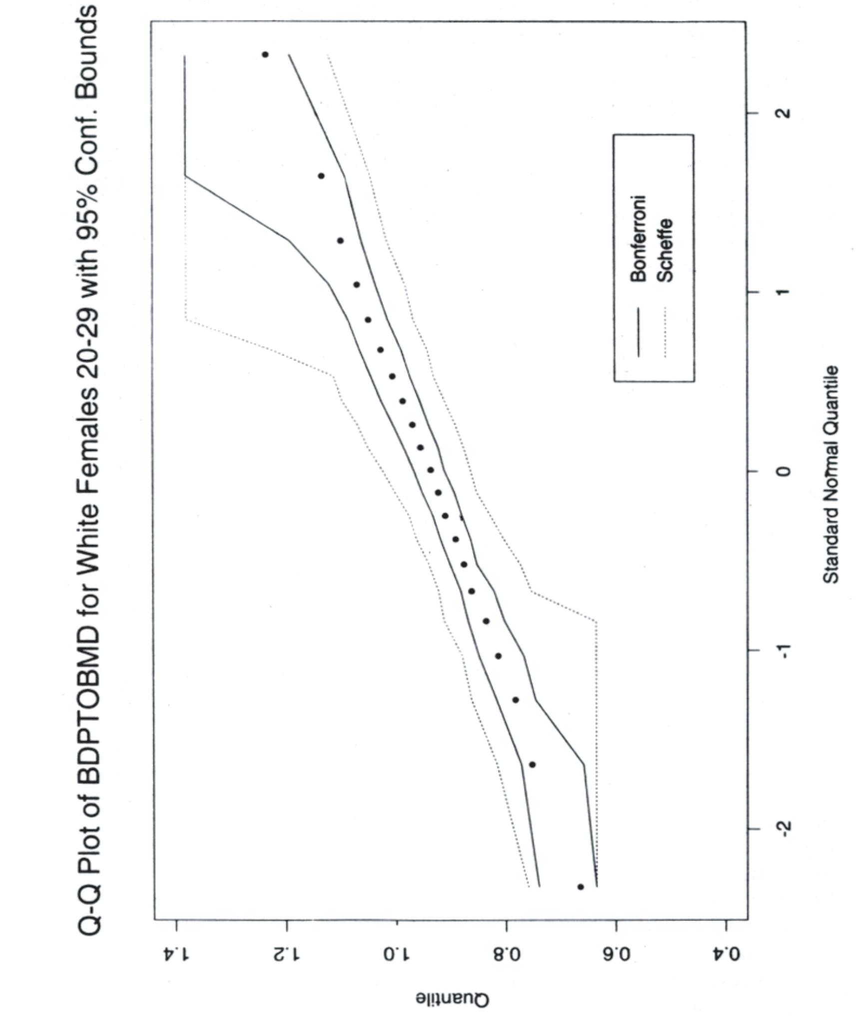
6.2 *Bone mineral density: Comparison to a normal reference distribution.* NHANES III collects several measurements of bone mineral density (BMD), based on detailed examination of "bone scan" photographs of the femur and hip bones. The present analysis will restrict attention to the variable BDPTOBMD, "total bone mineral density." This generally is considered the best single measure of bone mineral density.

As noted in Section 1.1, low measurements of bone mineral density generally are associated with increased risk of broken bones, especially in elderly women. The bone literature generally seeks to define low BMD values in terms of the quantiles of BMD measurements taken on white females aged 20-29. Historically, the relevant quantile estimates were computed from observed BMD means and variances for samples from this subpopulation, in conjunction with the assumption that BMD measurements were normally distributed within the subpopulation of white females aged 20-29.

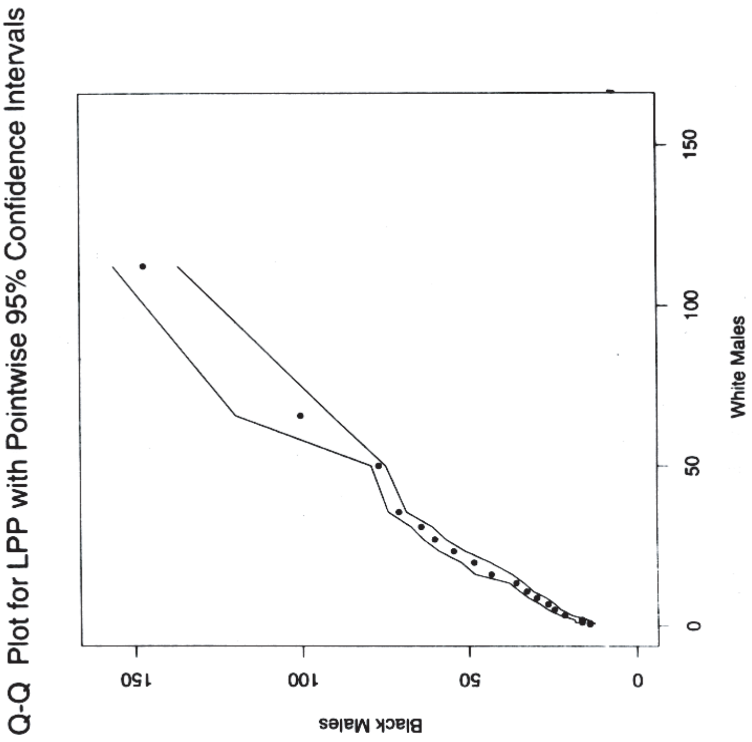
Previous BMD studies generally had been based on clinical studies, rather than population-based studies. Thus, when the NHANES III BMD data became available, it was of interest to examine the extent to which these new data were consistent with the abovementioned normality assumption.

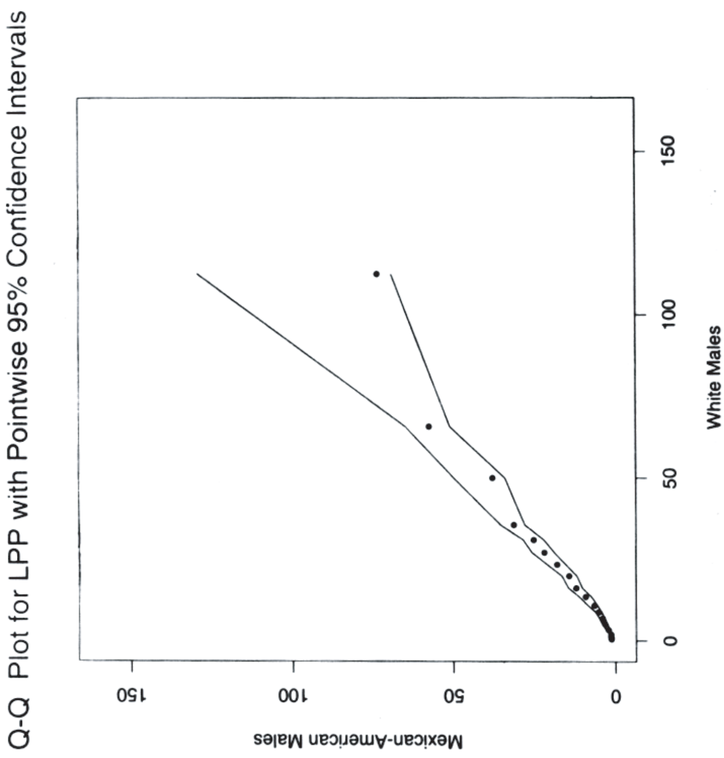
To address this issue, Figure 1 presents a plot of the estimated BMD quantiles for white females aged 20-29 against the associated standard normal quantiles (solid circles). The 21 plotted point estimates correspond to the values $q = 0.01, 0.05(0.05)0.95, 0.99$. For these 21 points, Figure 1 also includes simultaneous 95% confidence bounds based on the Scheffé (dotted lines) and Bonferroni (solid line) methods, respectively. In keeping with the formal development in Sections 2 through 4, we emphasize that the simultaneous bounds in the vertical dimension are evaluated with respect to the fixed values identified on the horizontal axis; the same comment applies to the plots presented in Section 5.3 below. Within the variability indicated by these confidence bounds, the plotted point estimates fall approximately along a straight line. Thus, the BDPTOBMD data for white females aged 20-29 appear to be consistent with the normal-distribution assumption described above.

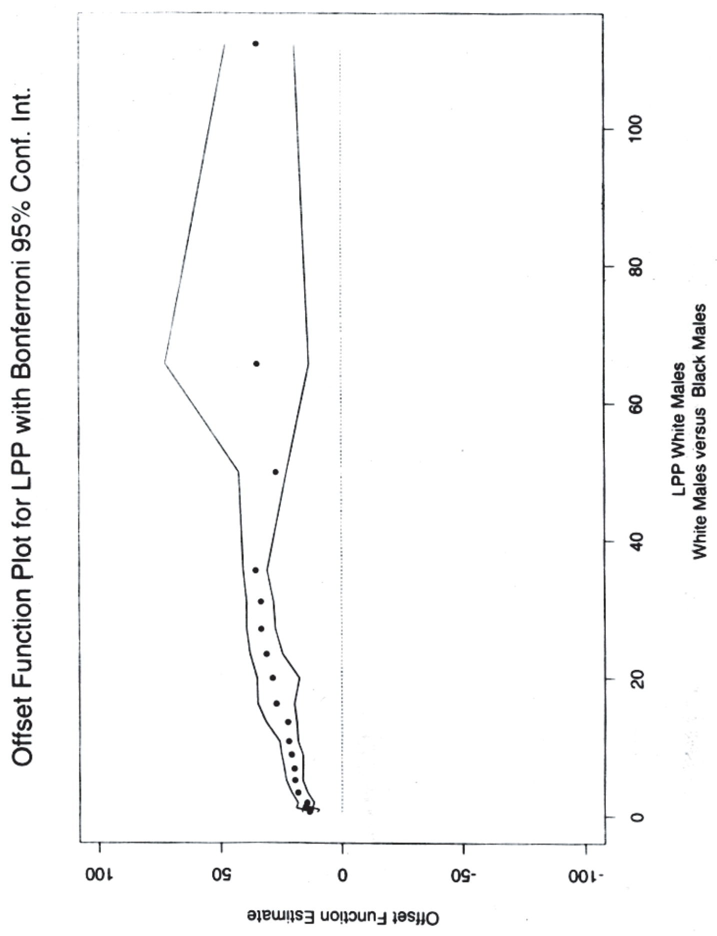
6.3 *Lipoprotein(a): Comparison of subpopulations.* The lipoprotein(a) (LPP) data considered here were obtained through analysis of blood samples that were collected as part of the medical examination component of NHANES III. One complicating characteristic of the LPP measurements is that a substantial proportion of the measurements were below the detection limit for the measuring

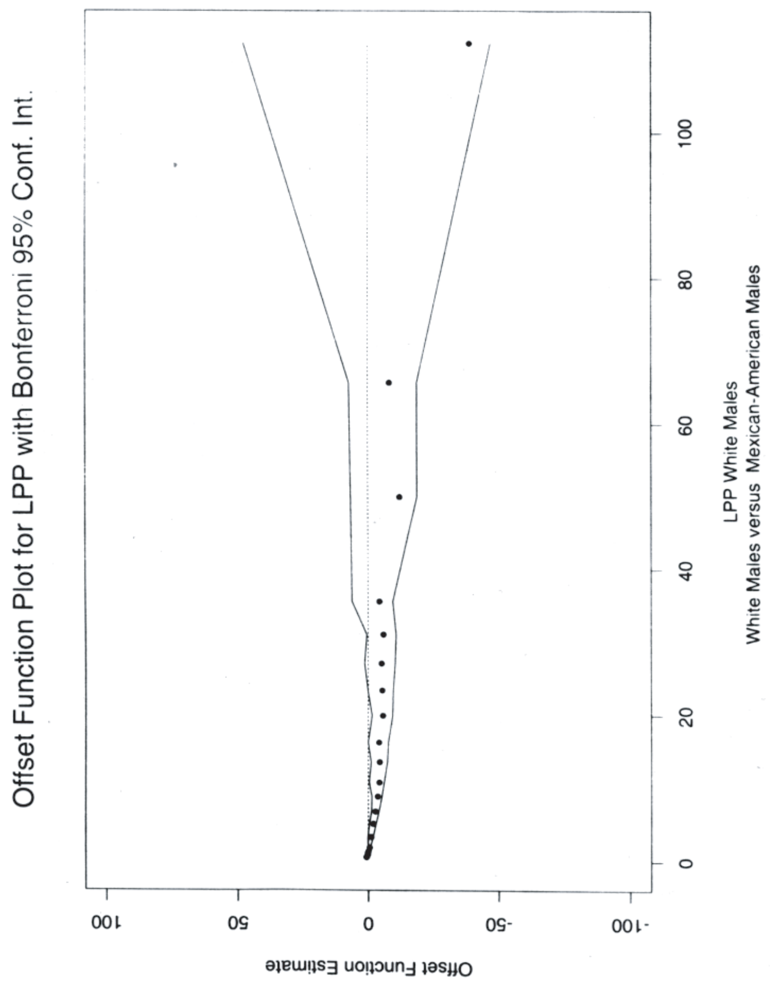


Q-Q Plot of BDPTOBMD for White Females 20-29 with 95% Conf. Bounds
Bonferroni
Scheffe
Standard Normal Quantile
Quantile
-2
-1
0
1
2
1.4
1.2
1.0
0.8
0.6
0.4


Q-Q Plot for LPP with Pointwise 95% Confidence Intervals
Black Males
White Males
0
50
100
150
0
50
100
150


Q-Q Plot for LPP with Pointwise 95% Confidence Intervals
White Males
Mexican-American Males
0
50
100
150
0
50
100
150


Offset Function Plot for LPP with Bonferroni 95% Conf. Int.
Offset Function Estimate
100
50
0
-50
-100
0
20
40
60
80
100
LPP White Males
White Males versus Black Males



Offset Function Plot for LPP with Bonferroni 95% Conf. Int.
Offset Function Estimate
100
50
0
-50
-100
0
20
40
60
80
100
LPP White Males
White Males versus Mexican-American Males

equipment, and thus were recorded as zero values. For various subpopulations, this proportion of recorded zeros varied between approximately ten and thirty percent.

Some previous studies have suggested that high levels of lipoprotein(a) might be associated with increased risk of cardiovascular disease, while other studies have reported results to the contrary. In addition, previous studies had suggested the presence of elevated levels of lipoprotein(a) in some minority groups. See, e.g., Ridker *et al.* (1993), Alfthan *et al.* (1994), Wild *et al.* (1997) and references cited therein. Consequently, there was special interest in comparing the upper quantiles of LPP of minority and non-minority subpopulations.

6.3.1 *Quantile-quantile plots.* To address this, we compared the quantiles of LPP for adult males in the white non-Hispanic, black non-Hispanic and Mexican American subpopulations. Comparisons for females in these race-ethnic groups were similar, but somewhat less distinct, and will not be discussed further here.

Specifically, Figure 2 displays an estimated quantile-quantile plot of LPP for the adult male white non-Hispanic and adult male black non-Hispanic subpopulations, respectively (solid circles). The 21 plotted point estimates correspond to the values $q = 0.01, 0.05(0.05)0.95, 0.99$. Note especially that the plotted points, and associated pointwise 95% confidence bounds (solid lines), fall above a line with slope equal to 1 and intercept equal to zero. This is consistent with the abovementioned anticipation of elevated levels of LPP among some minority groups. However, Figure 3 displays a corresponding estimated quantile-quantile plot of LPP for the adult male white non-Hispanic and adult male Mexican-American subpopulations, respectively. This latter plot does not indicate any pronounced distinction between the associated two demographic groups. Thus, it appears that statements regarding possible elevated levels of LPP among minorities would need to be couched in terms of specific minority subgroups.

6.3.2 *Offset-function plots.* The quantile-quantile plots in Figures 2 and 3 offer some indication of the relationship between the quantiles of the subpopulations of interest. However, offset-function plots can give somewhat more focused indications of the specific ways in which the subpopulation distributions may differ. For example, after examining Figure 2, one may ask whether the different LPP quantiles for the two subpopulations may be explained by a simple constant offset quantity Δ_0 (at least in the upper tails), or whether a more complicated offset function pattern exists. To examine this, Figure 4 presents an offset function plot for the adult male black non-Hispanic and adult male white non-Hispanic subpopulations (solid circles) with associated simultaneous Bonferroni 95% confidence bounds (solid line). Note especially that the offset function appears to be approximately constant for $q \geq 0.6$. Also, in keeping with Figure 2, note the pronounced indication that $\Delta(x) > 0$ for the nontrivial LPP measurements, even after accounting for the variability reflected in the confidence bounds.

Finally, Figure 5 presents an offset function plot to compare the adult male

Mexican-American and adult male white non-Hispanic subpopulations. Note especially that the offset-function point estimates generally are less than or equal to zero, corresponding to an LPP distribution that may be slightly *lower* in the adult male Mexican-American subpopulation. However, note also that the Bonferroni confidence bounds are fairly wide relative to the offset-function point estimates, especially for $q \geq 0.75$. To some degree, this reflects the limitations on power in tests associated with quantile-quantile plots or offset function plots.

7. **Discussion**

7.1 *Review of proposed methods.* This paper has considered construction of quantile-quantile and offset-function plots based on data collected through a complex sample survey. Principal attention has focused on construction of pointwise and simutaneous confidence bounds for these plots. In developing the proposed methods, three points have been of special interest. First, for plots involving comparison of one population to a specified reference distribution (e.g., the standard normal distribution), confidence bounds follow directly from confidence bounds developed previously for individual quantiles, e.g., in Woodruff (1952) and Francisco and Fuller (1991).

Second, it is somewhat more complicated to develop appropriate confidence bounds for plots that compare two subpopulation distributions. In particular, the resulting bounds involve the approximate joint distribution of the weighted sample distribution functions $\hat{F}(x)$ and $\hat{G}(x)$. Methods similar to the Bahadur approximation lead to an approximation for the distribution of the sample offset function $\hat{G}^{-1}\{\hat{F}(x)\} - x$, and thus to approximate confidence bounds for the superpopulation offset function. Additional arguments then lead to confidence bounds applicable to quantile-quantile plots. Third, as a technical note, in most practical applications for a given level of nominal simutaneous coverage $1 - \alpha$, one generally will obtain narrower confidence bounds through use of Bonferroni methods rather than Scheffé-type methods; this is consistent with related ideas developed by Korn and Graubard (1990) for regression coefficient vectors.

7.2 *Variants and extensions.* The ideas in this paper can be extended in several ways. First, the methods proposed here are based on the large-sample distributional approximations outlined in Section 2.2. A detailed development of sufficient design and superpopulation conditions for these approximations is of interest but is beyond the scope of the present paper. Similar comments apply to the distributional approximations used for the sample offset function $\hat{G}^{-1}\{\hat{F}(x)\} - x$. Assessment of the adequacy of these distributional approximations will be of special interest for cases in which the effective degrees of freedom term for $\hat{V}[\{\hat{F}(x)^T, \hat{G}(y)^T\}^T]$ is small for some subpopulations. This small-degrees-of-freedom problem is of practical interest for some NHANES III analyses.

Second, Sections 2 through 6 restricted attention to simple weighted point estimators of $F(\cdot)$ and $G(\cdot)$ as considered in, e.g., Woodruff (1952) and Francisco and Fuller (1991). However, some authors have developed design-based point estimators of $F(\cdot)$ and $G(\cdot)$ make explicit use of auxiliary information; see, e.g., Rao *et al.* (1990). The principal ideas of Sections 2 through 6 can be applied to these somewhat more complicated cases, with appropriate changes in the required offset function point estimators and confidence bounds.

Third, other authors have used model-based approaches to estimation of distribution functions and quantiles with complex survey data. See, e.g., Chambers and Dunstan (1986), Chambers *et al.* (1992), Dorfman (1993), Dorfman and Hall (1993), Krieger and Pfeffermann (1997) and references cited therein. The main ideas of Sections 3 through 5 can also be used under a model-based approach. In future comparisons of various design-based and model-based methods for quantile-quantile and offset-function plotting, it may be especially useful to examine the following issues.

(a) Pointwise and simultaneous coverage rates, evaluated with respect to the sampling design, a specified model, or both.

(b) Similar evaluation of the distribution of the widths of confidence intervals, and the distribution of the associated volumes of k-dimensional confidence sets.

(c) Performance of related omnibus tests for goodness of fit, both for comparison of one population distribution to a reference distribution; and for comparison of two subpopulation distributions.

Appendix

SIMULTANEOUS CONFIDENCE INTERVALS FOR PARAMETER VECTORS

We now consider Scheffé and Bonferroni simultaneous confidence intervals. Scheffé simultaneous confidence intervals can be extracted directly from simultaneous confidence sets formed by the Scheffé s-method (e.g., Seber 1977, pp. 128-129). Following the reasoning outlined in Korn and Graubard (1990), consider a k-dimensional point estimator $\hat{\theta}$ that follows a normal distribution with mean θ and covariance matrix $V(\hat{\theta})$; let $\hat{V}(\hat{\theta})$ have an expectation of $V(\hat{\theta})$; and assume that $\hat{V}(\hat{\theta})$ follows a multiple of a Wishart$_k$ distribution on $n-L$ degrees of freedom, independent of $\hat{\theta}$. In addition, define the intervals,

$$\hat{\theta}_i \pm \ [\frac{k(n-L)}{n-L-k+1}\ F_{k,n-L-k+1}(1-\alpha)\ \{\hat{V}(\hat{\theta})\}_{ii}]^{\frac{1}{2}},\ \ i=1,\ldots,k \qquad \ldots(A.1)$$

where $\hat{\theta}_i$ is the ith element of the vector $\hat{\theta}$ and $\{\hat{V}(\hat{\theta})\}_{ii}$ is the ith diagonal element of $\hat{V}(\hat{\theta})$.

Alternative simultaneous confidence intervals can be constructed by the Bonferroni procedure based on the t distribution,

$$\hat{\theta}_i \pm \ t_{1-\alpha/2k,n-L}[\{\hat{V}(\hat{\theta})\}_{ii}]^{\frac{1}{2}}, \quad i=1,\ldots,k \qquad \ldots(A.2)$$

where $t_{1-\alpha/2k,n-L}$ is the the $(1-\alpha/2k)$ quantile value of t distribution with $n-L$ degree of freedom.

Table 1 below compares critical values of these two simultaneous confidence intervals for $\alpha = 0.05$. For $\alpha = 0.05$, the Bonferroni confidence intervals are always shorter than Scheffé's interval except in the simple case of $k = 1$. For related comments comparing Bonferroni and Scheffé simultaneous confidence intervals, see, e.g., Korn and Graubard (1990), Christensen (1973) and Miller (1966, pp. 62-69). In general, Scheffé confidence intervals are preferred in one wishes to carry out simultaneous inference for arbitrary linear combinations of θ. However, the present paper will consider confidence intervals only for the elements of the quantile vector θ, and for prespecified linear combinations of θ, e.g., the interquartile range.

Table 1. Critical values for Scheffé and Bonferroni confidence intervals for $\alpha = 0.05$

$$F^* = k(n-L)/(n-L-k+1) \cdot F_{k,n-L-k+1}(1-\alpha)$$
$$t^* = t_{1-\alpha/2k,n-L}$$

$n-L$		k = 1	2	3	4	9
10	F^*	2.228139	3.075529	3.904871	4.852282	29.53501
	t^*	2.228139	2.633767	2.870073	3.038243	3.518150
22	F^*	2.073873	2.695139	3.197607	3.661819	6.117110
	t^*	2.073873	2.405473	2.591212	2.720139	3.073740
44	F^*	2.015368	2.564853	2.980773	3.340785	4.866075
	t^*	2.015368	2.320711	2.488968	2.604550	2.916337

Acknowledgement. The authors thank D. Brody, M.A. Carroll, A. Looker and V.L. Parsons of the NCHS for providing the NHANES III data used in this paper, and for helpful comments on related application issues; Alan Dorfman, Kenneth Koehler and two anonymous referees for very helpful comments on earlier versions of this work; and Richard Groeneveld for the Oja (1981) reference.

References

Alfthan, G., Pekkanen, J., Jauhiainen, M., Pitkäniemi, J., Karvonen, M., Tuomilehto, J., Salonen, J.T. and Ehnholm, C. (1994). Relation of serum homocysteine and lipoprotein(a) concentrations to atherosclerotic disease in a prospective finnish population based study. *Atherosclerosis* **106**, 9-19.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edn., New York: Wiley.

Chambers, R.L., Dorfman, A.H. and Hall, P. (1992). Properties of estimators of the finite population distribution function. *Biometrika* **79**, 577-582.

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika* **73**, 597-604.

Christensen, L.R. (1973). Simultaneous statistical inference in the normal multiple linear regression model, *Jour. Amer. Statist. Assoc.,* **68**, 457-461.

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.

DANIELS, E.D., PETTIFOR, J.M., SCHNITZLER, C.M., RUSSELL, S.W. AND PATEL, D.N. (1995). Ethnic differences in bone density in female South African nurses. *Jour. Bone and Mineral Research* **10**, 359-367.

DEMPSTER, A.P. AND RYAN, L.M. (1985). Weighted normal plots. *Jour. Amer. Statist. Assoc.* **80**, 845-850.

DOKSUM, K.A. (1977). Some graphical methods in statistics. A review and some extensions. *Statistica Neerlandica* **31**, 53-68.

DOKSUM, K.A. AND SIEVERS, G.L. (1976). Plotting with confidence: graphical comparisons of two populations. *Biometrika* **63**, 421-434.

DORFMAN, A.H. (1993). A comparison of design-based and model-based estimators of the finite population distribution function. *Australian Journal of Statistics* **35**, 29-41.

DORFMAN, A.H. AND HALL, P. (1993). Estimators of the finite population distribution function using nonparametric regression. *Ann. Statist.* **21**, 1452-1475.

DURBIN, J. (1968). The probability that the sample distribution function lies between two parallel straight lines. *Ann. Math. Statist.* **39**, 398-411.

FRANCISCO, C.A. AND FULLER, W.A. (1991). Quantile estimation with a complex survey design, *Ann. Statist.* **19**, 454-469.

GAN, F.F., KOEHLER, K.J. AND THOMPSON, J.C. (1991). Probability plots and distribution curves for assessing the fit of probability models. *Amer. Statist.* **45**, 14-21.

HARTER, H.L. (1984). Another look at plotting positions. *Communications in Statistics, A* **13**, 1613-1633.

HOLT, D., SCOTT, A.J., AND EWINGS, P.O. (1980). Chi-squared tests with survey data, *Jour. Royal Statist. Soc.,* Ser A, **143**, 302–320.

JOHNSON, N.L. AND KOTZ, S. (1970). *Continuous Univariate Distributions,* Volume 2, Boston, Hougton Mifflin.

KORN, E.L. AND GRAUBARD, B.I. (1990). Simultaneous testing of regression coefficients with complex survey data : use of Bonferroni t statistic, *The American Statistician*, **44**, 270-275.

KORN, E.I. AND GRAUBARD, B.G. (1995). Analysis of large health surveys: accounting for the sampling design, *Jour. Royal Statist. Soc.,* Ser. A **158**, 263-295.

KREWSKI, D. AND RAO, J.N.K. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods, *Ann. Statist.* **9**, 1010-1019.

KRIEGER, A.M. AND PFEFFERMANN, D. (1997). Testing of distribution functions from complex sample surveys. *Jour. Off. Statist.* **13**, 123-142.

MCCARTHY, P.J. (1993). Standard error and confidence interval estimation for the median. *Jour. Off. Statist.* **9**, 673-689.

MILLER, R.G. (1966). *Simultaneous Statistical Inference*, New York: Mcgraw-Hill.

NATIONAL CENTER FOR HEALTH STATISTICS (1996). *NHANES III Reference Manuals and Reports.* CD-ROM, GPO 017-022-01358-4, United States Government Printing Office, Washington, DC.

OJA, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.* **8**, 154-168.

RAO, J.N.K., KOVAR, J.G. AND MANTEL, H.J. (1990). On estimating distribution functions and quantiles from survey data. *Biometrika* **77**, 365-375.

RAO, J.N.K. AND SCOTT, A.J. (1981). The analysis of categorical data from complex surveys: chi-squared tests for goodness-of-fit and independence in two-way tables, *Jour. Amer. Statist. Assoc.,* **76**, 221–230.

RAO, J.N.K. AND THOMAS, D.R. (1989). Chi-squared tests for contigency tables. Pp. 89–114 in *Analysis of Complex Surveys,* Skinner, C.J., Holt, D.and Smith, T.M.F. eds, New York: Wiley.

RIDKER, P.M., HENNEKENS, C.H. AND STAMPFER, M.J. (1993). A prospective study of lipoprotein(a) and the risk of myocardial infarction. *Jour. Amer. Medical Assoc.* **270**, 2195-2199.

SEBER, G.A.F. (1977). *Linear Regression Analysis*, New York: Wiley.

SHAO, J. (1996. Resampling methods in sample survey, *Statistics*, **27**, 203-254.

SKINNER, C.J, HOLT, D. AND SMITH, T.M.F. (1989). *Analysis of Complex Surveys.* New York: Wiley.

THOMAS, D.R. AND RAO, J.N.K. (1987). Small-sample comparision of level and power for simple goodness-of-fit statistics under cluster sampling, *Jour. Amer. Statist. Assoc.* **82**, 630–636.

WILD, S.H., FORTMANN, S.P. AND MARCOVINA, S.M. (1997). A prospective case-control study of lipoprotein(a) levels and apo(a) size and risk of coronary heart disease in stanford five-city project participants. *Arteriosclerosis, Thrombosis and Vascular Biology* **17**, 239-245.

WOLTER, K. M. (1985). *Introduction to Variance Estimation*, New York: Springer-Verlag.

S.R. LEE and J.L. ELTINGE
DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION
TX 77843-3143.
e-mail : srlee@stat.tamu.edu / jeltinge@stat.tamu.edu