STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE DISTRIBUTIONS BASED ON INDUCED ORDER STATISTICS

by

Frank E. Harrell, Jr.

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1230

MAY 1979

STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE DISTRIBUTIONS BASED ON INDUCED ORDER STATISTICS

by

Frank E. Harrell, Jr.

A Dissertation submitted to the faculty of
the University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of
Biostatistics

Chapel Hill

1979

Approved by:

P K Sen

Adviser

D. Dale Williams

Reader

Clarence E Davis

Reader

FRANK EANES HARRELL, JR. Statistical Inference for Censored Multivariate Distributions based on Induced Order Statistics. (Under the direction of PRANAB KUMAR SEN.)

This research is concerned with methods of multivariate analysis applicable to life testing problems with concomitant variables observable only for failing units. With this setup the author attempts to provide:

1) estimates of parameters from censored bivariate normal observations begun by Watterson [1959],
2) the likelihood ratio test of independence based on censored bivariate normal observations,
3) methods for inference based on censored multivariate normal distributions especially by studying multiple and partial correlations,
4) parameter estimates and likelihood ratio tests of independence for other selected bivariate distributions, and
5) locally most powerful rank tests of independence for general censored bivariate distributions.

These methods also have applications in inference from biased samples resulting from a selection process.

ACKNOWLEDGMENTS

The author wishes to express much appreciation to his adviser, Professor P. K. Sen, who suggested the topic of this dissertation, and who was constantly available for counsel and guidance. The insight provided by Professor Sen was invaluable throughout the research process. The other members of the committee, Professors C. E. Davis, K. L. Lee, D. Quade, and O. D. Williams, also provided helpful suggestions and ideas which went into the dissertation.

The author also wishes to thank his wife, Marielouise Worden Harrell, for her patience and assistance in his graduate studies, and his parents, Mr. and Mrs. Frank E. Harrell, Sr. for their constant encouragement and understanding throughout his studies.

Finally, the author wishes to acknowledge the support provided by the Lipids Research Clinics Program through the National Heart, Lung and Blood Institute, contract NIH-NHLBI-71-2243 from the National Institutes of Health.

FEH

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND REVIEW OF LITERATURE

1.1. Introduction

In life testing studies, life length or time to failure of an item is the major response variable. Many parametric statistical methods are now available for analyzing the relationship between life length and a set of associated variables when life length measurements are censored in some way and the associated variables are observed for all n censored or uncensored units. The literature review included describes many of these methods as background for investigating the problem of concomitant data being available only for units that actually fail.

Selection problems can also be placed into this setting. For example, suppose that 1000 students take a college's entrance exam and the highest 100 scorers enter the college. Once there they take a second exam. The problem may be to estimate the population mean score of the second type of exam or the population correlation between the two exams given this biased sample, where the population of interest consists of all high school graduates.

1.2 Notation and Definitions

In most life testing studies, observations are accumulated sequentially in time and are censored, i.e., only a subset of the set of ordered life lengths is observed. Suppose that n items in an experiment have life lengths $Y_1,\ldots,Y_n$ and let $Y_{n,1} \leq Y_{n,2} \leq \cdots \leq Y_{n,n}$ be the order statistics corresponding to $Y_1,\ldots,Y_n$ with $Y_{n,j} \leq Y_{n,j+1} \leq \cdots \leq Y_{n,k}$ being observable, $1 \leq j \leq k \leq n$. If $j > 1$, $k = n$ we have censoring on the left; for $j = 1$, $k < n$ we have censoring on the right, the typical type for survival studies.

When j or k is a random variable, the censoring is said to be of type I; an example is an experiment with a pre-specified duration and k is the number of failures or deaths occuring during that time and $j = 1$. If k is pre-specified, the censoring is of type II and the duration of the experiment is thus a random variable. In multi-stage censoring schemes or staggered entry experiments (called progressive censoring by some authors - see for example Cohen [1965]), subjects are withdrawn from the study without having failed at each of a number of stages (or equivalently they enter the study at different times), but the statistical analysis is not performed until the end of the study. In true progressive censoring, statistical tests are made throughout the experiment to possibly terminate it early with rejection of the hypothesis of no effect of a treatment on life length.

Most published statistical methods as well as those proposed here are designed for analyzing type II right-censored data in which the smallest k out of n order statistics of life length are observed for fixed k. However, for both type I and type II censoring maximum likelihood estimation procedures are the same and asymptotically the estimators for either type will have equivalent properties. When k is a random variable, the distribution of derived estimators depends on the distribution of k and hence may be more complicated.

Let $X_1,\ldots,X_n$ denote concomitant variates corresponding to failure times $Y_1,\ldots,Y_n$ so that $(X_1,Y_1),\ldots,(X_n,Y_n)$ are independent and identically distributed random vectors (i.i.d.r.v.). Define

$$X_{n[i]} = X_j \quad \text{if} \quad Y_{n,i} = Y_j \quad \text{for} \quad i,j = 1,\ldots,n. \tag{1.2.1}$$

Bhattacharya [1974] has termed the $X_{n[i]}$ as the <u>induced order statistics</u> while David and Galambos [1974] have named these as the <u>concomitants of order statistics</u>. We will call $X_{n[1]},\ldots,X_{n[k]}$, $1 \le k \le n$ the first k of n induced order statistics. Note that these variates are not necessarily in any order. When concomitant data are observable only for failing units then only the first k of n induced order statistics are observable.

1.3. Methods for Analyzing Censored Survival Data When All Concomitant Data are Available.

Assume that the conditional distribution function (d.f.) of Y_i given $\underset{\sim}{X}_i$ is of the form $F_{\underset{\sim}{\beta}\underset{\sim}{x}_i}(y)$ with probability density function (p.d.f.) $f_{\underset{\sim}{\beta}\underset{\sim}{x}_i}(y)$, for $i = 1,\ldots,n$; here, the $\underset{\sim}{X}_i$ may be p (≥ 1)-vectors, $\underset{\sim}{\beta}$ is an unknown p-vector, and $\underset{\sim}{\beta}\underset{\sim}{X}_i$ is a vector product. The likelihood function of a type II right-censored sample (and given the $\underset{\sim}{X}_i$) is

$$\tilde{L}_{n,k} = \{\prod_{i=1}^{k} f_{\underset{\sim}{\beta}\underset{\sim}{X}_{n[i]}}(Y_{n,i})\}\{\prod_{i=k+1}^{n}[1-F_{\underset{\sim}{\beta}\underset{\sim}{X}_{n[i]}}(Y_{n,k})]\}. \tag{1.3.1}$$

Various authors have used (1.3.1) to draw statistical inference [on $\underset{\sim}{\beta}$ and other parameters associated with p.d.f. $f_{\underset{\sim}{\beta}\underset{\sim}{X}}(y)$] using the principle of maximum likelihood. See Zippin and Armitage [1966] for the exponential distribution, Glasser [1967] for the exponential hazard distribution, for example.

A good general description of the setup of the likelihood function is given by David and Moeschberger [1978] generalized to more than one cause of failure and multiple censoring points. Consider the case of type I multi-stage censoring in which an individual's failure is observed only if it occurs within some predetermined time interval, the length of which may vary from subject to subject.

In this case, subjects can enter the study at different times with the termination point being the same for all, so the times are adjusted to start at zero for each subject. Starting with n subjects, the ith subject is observed until censoring time B_i. The lifetime for a subject, Y_i , is known only if $Y_i \leq B_i$. The likelihood of the sample is then

$$\{\prod_j f_{\underset{\sim}{\beta}\underset{\sim}{X}_j}(Y_j)\} \{\prod_\ell [1- F_{\underset{\sim}{\beta}\underset{\sim}{X}_\ell}(B_\ell)]\} , \tag{1.3.2}$$

where the first product is over the known failures and the second is over the censored times. If the B_i are random as in random entry times to the study, the estimation procedure is the same but the properties of the estimators will depend on the distribution of the entry times. For type II censoring, it is assumed that all B_i are equal. A discussion of this also appears in Cohen [1965], aimed more at the case of multi-batch censoring without covariates.

Estimation of the parameter vector $\underset{\sim}{\beta}$ is usually carried out by maximizing the likelihood in (1.3.2) by an iterative process and associated tests of hypotheses are performed by computing likelihood ratios and using large sample theory pertaining to the distribution of -2 log L where L is the likelihood ratio. Halperin [1952] showed that the maximum likelohood estimate (MLE) of a single parameter with no covariates is in some sense optimal under type II right censoring with mild regularity conditions on the underlying density. He also gave an outline of the proof for the multiple parameter case, the MLE is weakly consistent, efficient, and asymptotically multi-normally distributed. An additional paper on this subject is by Silvey [1961], which includes a discussion of the properties of the likelihood for dependent random variables.

Maximum likelihood estimators can seldom be derived explicitly. Because of this the properties of the estimators are difficult to study. Saw [1961] was able to study the moments of the MLE of the mean (μ) and standard deviation (σ) derived from a type II right censored normal sample. He found the estimators to be badly biased for large proportion of censoring (small k/n). For $k/(n+1)=.2$, he derived

$$\mu - E(\hat{\mu}) = 5.538\sigma/(n+1) + 0(n+1)^{-2} \tag{1.3.3}$$

and the amount of bias is worse for smaller $k/(n+1)$.

An explicit and asymptotically efficient method for deriving unbiased estimators was given by Plackett [1958] and formalized by Chan [1967]. Chan's formulation is for a general location-scale distribution and <u>double</u> censoring. Here the sample consists of $Y_{n,j}$, $Y_{n,j+1},\ldots,Y_{n,k}$, $1 < j < k < n$ from the parent density $\theta_2^{-1}f((y-\theta_1)/\theta_2)$ where θ_1 and θ_2 are respectively the location and scale parameters. Let $Z_{n,i} = (Y_{n,i} - \theta_1)/\theta_2$ and $U_{n,i} = E(Z_{n,i})$. The log likelihood function is (Chan [1967])

$$L = \log[n!/(j-1)!(n-k)!] - (k-j+1)\log\theta_2 + \sum_{i=j}^{k} \log f(Z_{n,i})$$

$$+ (j-1)\log F(Z_{n,j}) + (n-k)\log[1-F(Z_{n,k})], \qquad (1.3.4)$$

where $F(z)$ is the d.f. corresponding to the p.d.f. $f(z)$. The strategy is to expand $\partial L/\partial\theta_r$, $r = 1,2$, in a Taylor series about $Z_{n,i} = u_{n,i}$, $j \leq i \leq k$, and then to ignore terms which converge in probability to zero. Thus, aside from these remaining terms of order greater than one, which converge in probability to zero,

$$n^{-1/2}(\partial L/\partial\theta_r) = n^{-1/2}(\partial L/\partial\theta_r)\Big|_{u_{n,j},\ldots,u_{n,k}}$$

$$+ n^{-1/2}\theta_2^{-1}L^{(r)} \qquad (1.3.5)$$

where

$$L^{(1)} = -(j-1)\,[(Y_{n,j}-\theta_1)/\theta_2 - u_{n,j}](f_j'/p_j - f_j^2/p_j^2)$$

$$- \sum_{i=j}^{k} [(Y_{n,i}-\theta_1)/\theta_2 - u_{n,i}](f_i''/f_i - f_i'^2/f_i^2) \qquad (1.3.6)$$

$$+ (n-k)\,[(Y_{n,k}-\theta_1)/\theta_2 - u_{n,k}](f_k'/q_k + f_k^2/q_k^2),$$

$$L^{(2)} = -(j-1)[(Y_{n,j}-\theta_1)/\theta_2 - u_{n,j}](f_j/p_j + u_{n,j}f'_j/p_j - u_{n,j}f_j^2/p_j^2)$$

$$- \sum_{i=j}^{k} [(Y_{n,i}-\theta_1)/\theta_2 - u_{n,i}](u_{n,i}f''_i/f_i - u_{n,i}f'^2_i/f_i^2 + f'_i/f_i)$$

$$+ (n-k)[(Y_{n,k}-\theta_1)/\theta_2 - u_{n,k}](f_k/q_k + u_{n,k}f'_k/q_k \qquad (1.3.7)$$

$$+ u_{n,k}f_k^2/q_k^2),$$

$$f_i = f(u_{n,i}),\ p_i = F(u_{n,i}),\ q_i = 1-p_i\ ,\ f'_i = \partial f(x)/\partial x|_{x=u_{n,i}} \qquad (1.3.8)$$

and so forth.

The first term of (1.3.5) also converges in probability to zero so the estimation procedure is to solve for $L^{(r)} = 0$, $r = 1,2$. Denote the estimators satisfying this requirement by θ_1^* and θ_2^*. The fact that (θ_1^*, θ_2^*) are unbiased follows from the fact that $E(L^{(r)}) = 0$. Chan also shows that estimators derived from replacing $u_{n,i}$ by $F^{-1}(i/(n+1))$ have the same asymptotic properties. The extimators (θ_1^*, θ_2^*) can be called linearized maximum likelihood estimators.

Saw [1962] developed a procedure similar to that of Chan for the multiple parameter case and for general likelihood functions in which the likelihood is expressed as a polynomial with a remainder term. This procedure can be used in many cases to derive simple and efficient moment estimators. Saw applied this to right censored normal samples (See also Saw [1959]) to derive simple and efficient estimators of the mean and variance. These estimators require computation of only two expected values as opposed to Chan's estimators which involve calculation of $u_{n,j},\ldots,u_{n,k}$.

Modified maximum likelihood techniques could perhaps be applied to the more complicated likelihood in (1.3.1) involving covariates. This technique might eliminate the need for iterative solutions and provide unbiased estimates of regression parameters. At any rate, there seems to be no rigorous proof of the optimality of MLE's derived from expressions such as (1.3.1) which involve covariates. Many authors in the literature who discuss survival analysis with covariates seem to assume that theorems concerning identically distributed observations apply.

There are other methods in the literature for analyzing censored survival data than the conventional parametric maximum likelihood technique. Cox [1972] developed a conditional likelihood model which could be called a mixed parametric/nonparametric approach. In his model, which allows for arbitrary right censoring, only the ranks of the failure times are used. The model assumes that for some (unspecified) monotonic transformation d(y) of the failure times, $d(Y)-\underset{\sim\sim}{\beta X}$ has an extreme value distribution. This implies that the ratio of hazard rates for observations i and j at each fixed failure time y is $\exp(\underset{\sim\sim}{\beta X_i}-\underset{\sim\sim}{\beta X_j})$. [The hazard rate is defined to be $f(y)/(1-F(y))$]. This mixed approach has been explored for general distributions by Kalbfleisch [1978].

Brown, Hollander, and Korwar [1974] developed a nonparametric test of association between arbitrarily right censored failure time and one uncensored concomitant variable. The test is based on Kendall's tau and is easily applied. The data for the test consist of n independent pairs $(X_1,Y_1),\ldots,(X_n,Y_n)$ where Y_i is known only if $Y_i \le B_i$.

Let $\delta_i = I\{Y_i \text{ uncensored}\} = I\{Y_i \le B_i\}$ and $Z_i = \min(Y_i,B_i)$, where $I\{A\}$ denotes the indicator function for the set A. Let

$$a_{ij} = \begin{cases} 1, & X_i > X_j \\ 0, & X_i = X_j \\ -1, & X_i < X_j \end{cases} \tag{1.3.9}$$

$$b_{ij} = \begin{cases} 1 & \text{if it can be inferred that } Y_i > Y_j \\ 0 & \text{if } Y_i = Y_j \text{ or if "uncertain"} \\ -1 & \text{if it can be inferred that } Y_i < Y_j \end{cases} \tag{1.3.10}$$

so that

$$b_{ij} = \begin{cases} 1, (\delta_j = 1 \text{ and } Z_i > Z_j) \text{ or } (\delta_i = 0, \delta_j = 1 \text{ and } Z_i = Z_j) \\ -1, (\delta_i = 1 \text{ and } Z_i < Z_j) \text{ or } (\delta_i = 1, \delta_j = 0 \text{ and } Z_i = Z_j) \\ 0, \text{ otherwise} \end{cases} \tag{1.3.11}$$

The test statistic for independence is

$$S = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\, b_{ij} \tag{1.3.12}$$

and the test can be performed by either finding the permutation distribution of S or by using a normal approximation to its null distribution.

Brown et al. [1974] proposed a modification of this test which uses the Kaplan-Meier [1958] empirical survival curve to gain more information in the "uncertain" case $b_{ij} = 0$. This test is more complicated and may make little improvement when much censoring is present. In fact, in the example in their paper, in which little censoring is present, the significance levels for the more complicated test are generally larger than for the simple test in (1.3.12). When X_i is a binary variable, (1.3.12) reduces to the modified two-sample test of Gehan [1965].

Majundar and Sen [1978] have considered generalized linear rank statistics for progressive censoring with staggered entry. Miller [1976] estimated parameters of a linear regression on failure times by means of weighted regression using weights constructed from the Kaplan-Meier curve. A general theory of rank analysis of covariance under progressive censoring when there is no staggered entry has been developed by Sen [1977]. The asymptotic theory of multiple likelihood ratio tests in progressive censoring has been studied by Sen [1976] for the single parameter case.

1.4. Framework for Survival Analysis with Incomplete Concomitant Data

All methods in section 1.2 assume that though $Y_{n,1},\ldots,Y_{n,k}$ are only observable, the entire set $(X_1,\ldots,X_n)$ [or equivalently $X_{n[1]},\ldots,X_{n[n]}$] is given prior to experimentation so that (1.3.1) is properly defined. In many other life testing problems, especially those arising in toxicological or other invasive or destructive studies, $X_{n[i]}$ is observable only when $Y_{n,i}$ is observable (for $i \leq n$), so that the second factor on the right hand side of (1.3.1) is no longer deterministic. This is particularly true if observing X_i necessitates the failure of the ith unit, so that for the surviving [n-k] units, the X_i are not observable.

As an example, suppose that an experiment is performed in which a toxic substance is injected in constant amounts at regular time intervals in mice until death occurs. At death, the amount of the toxic substance taken up by the kidneys is measured to test the hypothesis that death occurs when a certain amount is precipitated there. For mice alive at the end of the study, the toxic substance may have been dissipated by means other than by the kidneys' action. Suppose that for time or cost reasons, the experimenter, having started with 30 mice, had decided to stop the experiment, after the first 10 deaths occur. He may want to study the relationship between survival time and absorption of the toxic substance by the kidney, the latter being measureable only at the termination of life and removal of a kidney.

As another example, suppose that in an experiment designed to study the relationship between arteriosclerosis and life length, rhesus monkeys are fed an atherogenic diet. Autopsies are performed on the first 20 monkeys to die and the amount of fatty plague deposited on the walls of the aorta is measured. The experimenter may wish to quantify the strength of relationship between this measurement and life length. Other examples of this type arise in selection problems, particularly in educational testing.

The necessity of observing concomitant data for all individuals is not required if we make the assumption that (X_i,Y_i) has a multivariate distribution of known form. Suppose that (X_i,Y_i) has p.d.f. $f(x,y)$ and the conditional density of X_i given $Y_i = y_i$ is $g_{y_i}(X)$. By an appeal to Lemma 1 of Bhattacharya [1974], we conclude that

given $Y_{n,1},\ldots,Y_{n,k}$, $X_{n[1]},\ldots,X_{n[k]}$ are (conditionally) independently distributed and the conditional p.d.f. of $X_{n[j]}$ given the $Y_{n,i}$, $1 \le i \le k$ is given by

$$g_{y_{n,j}}(x) \quad \text{for } j=1,\ldots,k \text{ and every } 1 \le k \le n. \tag{1.4.1}$$

This result provides the basis for much of the statistical inference based on induced order statistics.

1.5. Censored Bivariate Normal Distribution

Watterson [1959] considered the case where (X_i,Y_i) have a bivariate normal distribution. Denote the p.d.f. of (X_i,Y_i) by $\phi_{\underset{\sim}{\theta}}(x,y)$ with $\underset{\sim}{\theta} = (\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$, μ_x,μ_y are the means, σ_x^2,σ_y^2

the variances, and ρ the correlation coefficient of X and Y. By (1.4.1), given $Y_{n,1},\ldots,Y_{n,k}$, $X_{n[1]},\ldots,X_{n[k]}$ are independently normally distributed with means

$$\mu_x + \rho\sigma_x(Y_{n,i} - \mu_y)/\sigma_y \tag{1.5.1}$$

and variances

$$\sigma_x^2(1-\rho^2) \tag{1.5.2}$$

for $j=1,\ldots,k$. Let then $\underset{\sim}{X} = (X_{n[1]},\ldots,X_{n[k]})'$, $\underset{\sim}{Y} = (Y_{n,1},\ldots,Y_{n,k})'$ and $\underset{\sim}{Z} = \sigma_y^{-1}(\underset{\sim}{Y} - \mu_y \underset{\sim}{1})$. The joint (conditional) p.d.f. of $\underset{\sim}{X}$ given $\underset{\sim}{Y}$ is then

$$[\sigma_x^2(1-\rho^2)2\pi]^{-k/2} \exp\{-\frac{1}{2\sigma_x^2(1-\rho^2)} [\underset{\sim}{X}-\mu_x \underset{\sim}{1}- \rho\sigma_x\underset{\sim}{Z}]'[\underset{\sim}{X}-\mu_x \underset{\sim}{1}-\rho\sigma_x\underset{\sim}{Z}]\}, \tag{1.5.3}$$

which is the k-variate normal density with mean vector

$$\mu_x \underset{\sim}{1} + \rho\sigma_x \underset{\sim}{Z} \tag{1.5.4}$$

and covariance matrix

$$\sigma_x^2(1-\rho^2)\underset{\sim}{I}_k \tag{1.5.5}$$

which implies

$$E(\underset{\sim}{X}\,\underset{\sim}{X}'|\underset{\sim}{Y}) = \sigma_x^2(1-\rho^2)\underset{\sim}{I}_k + [E(\underset{\sim}{X}|\underset{\sim}{Y})][E(\underset{\sim}{X}|\underset{\sim}{Y})]' \qquad (1.5.6)$$

Thus if we let

$$E\underset{\sim}{Z} = \underset{\sim}{u} = (u_{n,1},\ldots,u_{n,k}) \text{ and } E(\underset{\sim}{Z}-E\underset{\sim}{Z})(\underset{\sim}{Z}-E\underset{\sim}{Z})' = \underset{\sim}{V}=((\nu_{n,ij})), \qquad (1.5.7)$$

then, from (1.5.4), (1.5.5) and (1.5.6) we obtain

$$E(\underset{\sim}{X}) = \mu_x \underset{\sim}{1} + \rho\sigma_x \underset{\sim}{u} \text{ and } E(\underset{\sim}{X}-E\underset{\sim}{X})(\underset{\sim}{X}-E\underset{\sim}{X})' = \sigma_x^2(1-\rho^2)\underset{\sim}{I}_k+\rho^2\sigma_x^2\underset{\sim}{V}. \qquad (1.5.8)$$

For $k \leq n \leq 20$, $\underset{\sim}{u}$ and $\underset{\sim}{V}$ are tabulated in Sarhan and Greenberg [1962, pp. 193-205] and several approximations for these moments for higher values of n are in David [1970, pp. 65-7]. Harter [1961] suggests that a good approximation for $u_{n,i}$ is

$$u_{n,i} \doteq \Phi_1^{-1}((i - \alpha_n)/(n+1 - 2\alpha_n)),$$

$$\alpha_n = 3/8, \quad 1 \leq n \leq 50, \quad \alpha_n = .4, \quad n > 50, \qquad (1.5.9)$$

where Φ_1 denotes the standard univariate normal d.f. The $u_{n,i}$ can also be approximated with high accuracy by numerical integration or by using the power series of Saw [1958] for select $k/(n+1)$ not too near 0 or 1. The result of Saw [1959, section 6.] may be used to interpolate the power series for values of $k/(n+1)$ not tabled. From (1.5.8) we obtain

$$E(X_{n[i]} - EX_{n[i]})(X_{n[j]} - EX_{n[j]}) = \rho^2\sigma_x^2\nu_{n,ij}\ , \quad i \neq j$$

$$= \sigma_x^2(1-\rho^2) + \rho^2\sigma_x^2\nu_{n,ii}\ , \quad i = j \qquad (1.5.10)$$

and using (1.5.4),

$$E(X_{n[i]} - EX_{n[i]})(Y_{n,j} - EY_{n,j}) = \rho\sigma_x\sigma_y\nu_{n,ij}$$

$$= \nu_{n,ij}\ \mathrm{Cov}(X_1, Y_1). \qquad (1.5.11)$$

All of these results can be found in Watterson [1959].

From (1.5.7) and the definition of $\underset{\sim}{Z}$,

$$E(\underset{\sim}{Y}) = \mu_y \underset{\sim}{1} + \sigma_y \underset{\sim}{u} \quad \text{and} \quad E(\underset{\sim}{Y} - \mu_y \underset{\sim}{1})(\underset{\sim}{Y} - \mu_y \underset{\sim}{1})' = \sigma_y^2 \underset{\sim}{V}. \qquad (1.5.12)$$

Let $\underset{\sim}{\alpha}$ denote a $k \times 1$ vector of constants. Then

$$E(\underset{\sim}{\alpha}'\underset{\sim}{Y}) = \mu_y \underset{\sim}{\alpha}' \underset{\sim}{1} + \sigma_y \underset{\sim}{\alpha}' \underset{\sim}{u}\ , \quad V(\underset{\sim}{\alpha}'\underset{\sim}{Y}) = \sigma_y^2 \underset{\sim}{\alpha}' \underset{\sim}{V}\underset{\sim}{\alpha} \qquad (1.5.13)$$

from which the best linear unbiased estimators (BLUE's) of μ_y and

σ_y can be determined. However, if $\underset{\sim}{\alpha}'\underset{\sim}{\alpha}$ is minimized instead of $\underset{\sim}{\alpha}'\underset{\sim}{V}\underset{\sim}{\alpha}$, as in Gupta [1952], the estimators are much simpler. The result is

$$\tilde{\mu}_y = \underset{\sim}{\tilde{\alpha}}' \underset{\sim}{Y}\ ,\ \underset{\sim}{\tilde{\alpha}} = k^{-1} - (\underset{\sim}{u}-\bar{u}_k\underset{\sim}{1})\bar{u}_k / \sum_{m=1}^{k} (u_{n,i}-\bar{u}_k)^2 \qquad (1.5.14)$$

$$\tilde{\sigma}_y = \underset{\sim}{\tilde{\beta}}' \underset{\sim}{Y}\ ,\ \underset{\sim}{\tilde{\beta}} = (\underset{\sim}{u}-u_k\ \underset{\sim}{1}) / \sum_{m=1}^{k} (u_{n,i}-\bar{u}_k)^2 \qquad (1.5.15)$$

where

$$\bar{u}_k = k^{-1} \sum_{i=1}^{k} u_{n,i}\ . \qquad (1.5.16)$$

For censoring on either side, Sarhan and Greenberg [1962, pp. 208-11] computed the minimum efficiency for the simpler estimators with respect to the BLUE's to be over 84% for both the mean and standard deviation for $n \leq 15$ and in general the efficiencies are well over 90%.

Now consider linear estimation using $\underset{\sim}{X}$:

$$E(\underset{\sim}{\alpha}'\underset{\sim}{X}) = \mu_x \underset{\sim}{\alpha}' \underset{\sim}{1} + \rho\sigma_x \underset{\sim}{\alpha}'\underset{\sim}{u} \qquad (1.5.17)$$

$$V(\underset{\sim}{\alpha}'\underset{\sim}{X}) = \sigma_x^2[\underset{\sim}{\alpha}'\underset{\sim}{\alpha} + \rho^2\underset{\sim}{\alpha}'(\underset{\sim}{V}-\underset{\sim}{I}_k)\underset{\sim}{\alpha}]. \qquad (1.5.18)$$

To get an unbiased estimator of μ_x we require $\underset{\sim}{\alpha}'\ \underset{\sim}{1} = 1,\ \ \underset{\sim}{\alpha}'\ \underset{\sim}{u} = 0.$

Minimizing the variance is a problem since it depends on the unknown parameter ρ. If we minimize for $\rho = 0$, we have the estimators

$$\mu_x = \tilde{\alpha}'X \quad \text{and} \quad \rho\tilde{\sigma}_x = \tilde{\beta}'X \tag{1.5.19}$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are defined in (1.5.14) and (1.5.15) respectively. Alternatively, minimization can be done for $\rho = 1$ yielding the same vectors as in the BLUE's of μ_y and σ_y. For $n = 10$ and all possible ways of left and/or right censoring, Watterson [1959] compared the relative efficiencies of the two for the worst possible value of $\bar{\rho}^2$ (0 or 1) and found the simpler estimators in (1.5.19) to be often better than the others and never much worse. Note that from (1.5.17), no linear estimator in X exists for ρ or σ_x. Estimation of ρ and σ_x will be the subject of much of the work in later chapters. Estimation of μ_x and σ_x will perhaps be of more interest in selection problems while estimation of ρ and testing $\rho = 0$ will be of interest in both selection problems and life testing studies.

We note that from (1.5.3) and (1.5.7) that the basic assumption of bivariate normality can be roughly checked by examining the conditional normality of X by usual regression methods and by examining a plot of $Y_{n,i}$ vs. $u_{n,i}$, $i = 1,\ldots,k$ for linearity.

David and Galambos [1974] also discuss this setup. They give asymptotic results for the expected rank of $X_{n[i]}$ in $X_1,\ldots,X_n$

for $i/n \to \lambda$ $(0 < \lambda < 1)$ as $n \to \infty$. They also show that for fixed k, $X_{n[1]} - E(X_{n[1]}),\ldots,X_{n[k]} - E(X_{n[k]})$ are asymptotically independently normally distributed with mean 0 and variance $\sigma_x^2(1-\rho^2)$ as $n \to \infty$ for $|\rho| < 1$. For our purposes, however, we will need $k=pn$ with $n \to \infty$, $0 < p \leq 1$, p fixed.

1.6. Summary

Chapter I provides the background necessary for analyzing censored data when concomitant information is available only as induced order statistics for uncensored observations. In chapters to follow, estimators for the remaining parameters, ρ and σ_x, of the bivariate normal distribution will be developed and a test of independence will be studied. These methods will be extended for the multivariate normal distribution thereby providing many tools for statistical inference for censored data that are parallel to those commonly used for complete data. Next, some other bivariate distributions will be considered. Finally, some nonparametric tests of independence will be studied and compared to parametric tests by both empirical power calculations and asymptotic results.

CHAPTER II

ESTIMATION OF PARAMETERS OF CENSORED BIVARIATE NORMAL DISTRIBUTIONS AND TEST OF INDEPENDENCE BASED ON INDUCED ORDER STATISTICS

2.1. Introduction.

Suppose that (X_i, Y_i) follows a bivariate normal distribution with parameters μ_x , μ_y , σ_x^2 , σ_y^2 , and ρ. Estimators of μ_x and $\rho\sigma_x$ using $X_{n[1]}, \ldots, X_{n[k]}$ have been developed by Watterson [1959] while the estimation of μ_y and σ_y using the order statistics $Y_{n,1} \cdots Y_{n,k}$ can be made as in Sarhan and Greenberg [1962]. However, the estimation of σ_x and ρ poses certain difficulties and constitutes a major objective of this chapter. In deriving estimators of σ_x and ρ we also derive an estimator of μ_x using both order statistics and induced order statistics which is slightly better than Watterson's. We will also show that the classical Pearsonian correlation coefficient has desirable properties as a test statistic for testing $H_0: \rho = 0$ and we will estimate its power function.

2.2. Estimation of σ_x^2 using only Induced Order Statistics.

Sometimes one may be interested in estimating population parameters from censored data even when the ordered variates, $\underset{\sim}{Y}$, are not observable. This problem arises often in selection processes. For example, suppose in a class of 100 students, the 20 tallest are determined by having the students order themselves by increasing height. These 20 students are asked to do high-jumps to determine their maximum jumping capability. Later, the athletic instructor wishes to estimate the population variability of jumping capacity not just for the tallest students, using only the observed jumps and the ranks of students' heights out of 100.

As we have seen in section 1.5, the population standard deviation of X_i , σ_x , is not estimable by any linear function of $\underset{\sim}{X}$. It is natural to try a quadratic estimator of the form $\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$. The variance of such an estimator would depend on the unknown parameter ρ so the variance cannot be minimized for all ρ. Following the Watterson approach, we might minimize the variance of $\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}$ when $\rho = 0$, which amounts to minimizing $\mathrm{tr}\underset{\sim}{A}^2$ subject to $E(\underset{\sim}{X}'\underset{\sim}{A}\underset{\sim}{X}) = \sigma_x^2$. The authors have found by simulations that this estimator, which requires the solution of k simultaneous equations, has little to gain over a much simpler estimator presented below. Let

$$\tilde{\sigma}_x^2 = \sum_{i=1}^{k} c_i(X_{n[i]} - \tilde{\mu}_x)^2 \tag{2.2.1}$$

where $\tilde{\mu}_x$ is Watterson's estimator, $\sum_{i=1}^{k} \tilde{\alpha}_i X_{n[i]}$, defined in (1.5.19) and the c_i are real constants. Now

$$E(X_{n[i]} - \tilde{\mu}_x)^2 = V(X_{n[i]}) + V(\tilde{\mu}_x) + \rho^2\sigma_x^2 u_{n,i}^2 + 2E\{X_{n[i]}(\mu_x - \tilde{\mu}_x)\}$$

and

$$V(X_{n[i]}) = \sigma_x^2(1-\rho^2 + \rho^2\nu_{n,ii}),\ V(\tilde{\mu}_x) = \sigma_x^2\ \{(1-\rho^2)\tilde{\underset{\sim}{\alpha}}'\tilde{\underset{\sim}{\alpha}} + \rho^2\tilde{\underset{\sim}{\alpha}}'V\underset{\sim}{\alpha}\},$$

$$\tilde{\underset{\sim}{\alpha}}'\tilde{\underset{\sim}{\alpha}} = k^{-1} + \bar{u}_k \ / \ \sum_m (u_{n,m} - \bar{u}_k)^2$$

where $u_{n,i}$, $\nu_{n,ii}$, and $\bar{u}_k$ are defined in (1.5.7) and (1.5.16).

Also,

$$E(X_{n[i]}\tilde{\mu}_x) = E(\sum_{\substack{j=1 \\ j\neq i}}^{k} \tilde{\alpha}_j X_{n[i]} X_{n[j]}) + \tilde{\alpha}_i\ E(X_{n[i]}^2)$$

$$= \sum_{j\neq i} \tilde{\alpha}_j \{\rho^2\sigma_x^2(\nu_{n,ij} + u_{n,i}u_{n,j}) + \mu_x^2 + \rho\mu_x\,\sigma_x\ (u_{n,i}+u_{n,j})\}$$

$$+\ \tilde{\alpha}_i\ E(X_{n[i]}^2)$$

$$= \rho^2\sigma_x^2 \sum_{j=1}^{k} \tilde{\alpha}_j \nu_{n,ij} + \mu_x^2 + \rho\mu_x\sigma_x\ u_{n,i} + \tilde{\alpha}_i\sigma_x^2(1-\rho^2)$$

and

$$E[X_{n[i]}(\mu_x - \tilde{\mu}_x)] = -\ \rho^2\sigma_x^2 \sum_{j=1}^{k} \tilde{\alpha}_j \nu_{n,ij} - \tilde{\alpha}_i\sigma_x^2(1-\rho^2).$$

Therefore,

$$E(X_{n[i]} - \tilde{\mu}_x)^2 = \sigma_x^2(a_i + \rho b_i) \tag{2.2.2}$$

where

$$a_i = (k-1)/k + \bar{u}_k \{1 + 2(u_{n,i} - \bar{u}_n)\} / \sum_{m=1}^{k} (u_{n,m} - \bar{u}_k)^2$$

$$b_i = - a_i - 2 \sum_{j=1}^{k} \tilde{\alpha}_j \nu_{n,ij} + \sum_{r=1}^{k} \sum_{s=1}^{k} \tilde{\alpha}_r \tilde{\alpha}_s \nu_{n,rs} + \nu_{n,ii} + u_{n,i}^2 \tag{2.2.3}$$

for $i=1,\ldots,k$. Let $\underset{\sim}{a} = (a_1,\ldots,a_k)'$, $\underset{\sim}{b} = (b_1,\ldots,b_k)'$, $\underset{\sim}{c} = (c_1,\ldots,c_k)'$. Thus, $E(\tilde{\sigma}_x^2) = \sigma_x^2(\underset{\sim}{a}'\underset{\sim}{c} + \rho \underset{\sim}{b}'\underset{\sim}{c})$ and for an unbiased estimator, we require $\underset{\sim}{a}'\underset{\sim}{c} = 1$ and $\underset{\sim}{b}'\underset{\sim}{c} = 0$ and, motivated by Watterson, we minimize $\underset{\sim}{c}'\underset{\sim}{c}$ which minimizes $V(\tilde{\sigma}_x^2)$ when $\rho = 0$, considering $\tilde{\mu}_x$ given. The result is

$$c_i = \{b_i \underset{\sim}{a}'\underset{\sim}{b} - a_i \underset{\sim}{b}'\underset{\sim}{b}\} / \{(\underset{\sim}{a}'\underset{\sim}{b})^2 - (\underset{\sim}{a}'\underset{\sim}{a})(\underset{\sim}{b}'\underset{\sim}{b})\}, \; 1 \leq i \leq k. \tag{2.2.4}$$

An estimator of ρ can also be obtained by considering the Watterson estimator of $\rho\sigma_x$ (based on $\underset{\sim}{X}$ alone) and dividing the same by the square root of $\tilde{\sigma}_x^2$ in (2.2.1). However, such an estimator does not depend on $\underset{\sim}{Y}$, and hence, when the latter is given, may not be a very efficient one. For this reason, in the next section, we proceed to study the efficiency of estimators of $\underset{\sim}{\theta} = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ based on $(\underset{\sim}{X}, \underset{\sim}{Y})$.

2.3 Cramèr-Rao Lower Bounds for Variances of Unbiased Estimators

Denote the parent bivariate normal density of (X_i, Y_i) by $\phi_{\underset{\sim}{\theta}}(x,y)$ where $\underset{\sim}{\theta} = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, μ_x, μ_y stand for the means, σ_x^2, σ_y^2 for the variances, and ρ for the correlation coefficient of X_i and Y_i. By (1.5.3) the joint (conditional) pdf of $\underset{\sim}{X}$ given $\underset{\sim}{Y}$ is

$$\prod_{i=1}^{k} \{\phi_1((x_i - \mu_x - \rho\sigma_x\sigma_y^{-1}(Y_{n,i} - \mu_y))/\sigma_x\sqrt{1-\rho^2})/\sigma_x\sqrt{1-\rho^2}\} \tag{2.3.1}$$

$$= \prod_{i=1}^{k} \{\phi_{\underset{\sim}{\theta}}(x_i, Y_{n,i})/\phi_1((Y_{n,i} - \mu_y)/\sigma_y)/\sigma_y\}$$

where ϕ_1 is the standard univariate normal density. The joint pdf for the first k of n order statistics is

$$n^{[k]} \prod_{i=1}^{k} \phi_1((y_i - \mu_y)/\sigma_y)[1 - \Phi_1((y_k - \mu_y)/\sigma_y)]^{n-k}\, \sigma_y^{-k}, \tag{2.3.2}$$

$$y_1 \le y_2 \le \cdots \le y_k$$

where Φ_1 is the standard univariate normal df and $n^{[k]} = n\cdots(n-k+1)$. Multiplying (2.3.1) and (2.3.2), we obtain that the joint pdf of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ is equal to

$$n^{[k]}\{\prod_{i=1}^{k} \phi_{\underset{\sim}{\theta}}(x_i, y_i)\}\, [1 - \Phi_1((y_k - \mu_y)/\sigma_y)]^{n-k} \tag{2.3.3}$$

$$y_1 \le y_2 \cdots \le y_k\,.$$

Let $L_{n,k}$ denote the log likelihood function based on (2.3.3) aside from constant terms. Then letting $\underset{\sim}{T} = (\underset{\sim}{X} - \mu_x \underset{\sim}{1})/\sigma_x$, $Z = (\underset{\sim}{Y} - \mu_y \underset{\sim}{1})/\sigma_y$, $Z_{n,k} = (Y_{n,k} - \mu_y)/\sigma_y$ we have

$$L_{n,k} = (n-k)\ \log[1-\Phi_1(Z_{n,k})] - k\log\sigma_x - k\ \log\ \sigma_y - (k/2)\log(1-\rho^2)$$
$$- \{(1-\rho^2)^{-1}/2\}\{\underset{\sim}{T}'\underset{\sim}{T} - 2\rho\ \underset{\sim}{T}'\underset{\sim}{Z} + \underset{\sim}{Z}'\underset{\sim}{Z}\}. \tag{2.3.4}$$

Then we have

$$(\partial/\partial\mu_x)L_{n,k} = \underset{\sim}{1}'\ (\underset{\sim}{T} - \rho\underset{\sim}{Z})/\sigma_x(1-\rho^2)$$

$$(2.3.5)$$

$$(\partial/\partial\sigma_x)L_{n,k} = -k/\sigma_x + (\underset{\sim}{T}'\underset{\sim}{T} - \rho\ T'Z)/\sigma_x(1-\rho^2)$$

$$(\partial/\partial\rho)\ L_{n,k} = (1-\rho^2)^{-2}[k\rho(1-\rho^2) - \rho\underset{\sim}{T}'\underset{\sim}{T} - \rho\underset{\sim}{Z}'\underset{\sim}{Z} + (1+\rho^2)\underset{\sim}{T}'\underset{\sim}{Z}]$$

$$(\partial/\partial\mu_y)L_{n,k} = (n-k)\sigma_y^{-1}\phi_1(Z_{n,k})/[1-\Phi_1(Z_{n,k})]$$
$$-(1-\rho^2)^{-1}\sigma_y^{-1}\ \underset{\sim}{1}'(\rho\underset{\sim}{T}-\underset{\sim}{Z})$$

$$(\partial/\partial\sigma_y)L_{n,k} = (n-k)\sigma_y^{-1}\ Z_{n,k}\phi_1(Z_{n,k})/[1-\Phi_1(Z_{n,k})]$$
$$- \sigma_y^{-1}(1-\rho^2)^{-1}(\rho\underset{\sim}{T}'\underset{\sim}{Z} - \underset{\sim}{Z}'\underset{\sim}{Z}) - k\ \sigma_y^{-1}.$$

With notations in (1.5.7), we let $\underset{\sim}{u} = (u_{n,1},\ldots,u_{n,k})'$ and $\underset{\sim}{w} = (w_1,\ldots,w_k)'$ where $w_i = \nu_{n,ii} + u_{n,i}^2$. Also let

$$\psi(n,k;\ a,b,c,d) = E\{\phi_1(Z_{n,k})^a \Phi_1(Z_{n,k})^{-b} [1-\Phi_1(Z_{n,k})]^{-c} Z_{n,k}^d\} \quad (2.3.6)$$

where a,b,c,d are non-negative numbers and the pdf of $Z_{n,k}$ is

$$[n^{[k]}/(k-1)!]\Phi_1(z)^{k-1}[1-\Phi_1(z)]^{n-k}\phi_1(z),\ -\infty < z < \infty \quad (2.3.7)$$

for $k = 1,\ldots,n$. Noting that from (1.5.8)

$$\begin{aligned} &E(\underset{\sim}{T}) = \rho\underset{\sim}{u},\quad E(\underset{\sim}{T}'\underset{\sim}{T}) = k(1-\rho^2) + \rho^2\, \underset{\sim}{1}'\, \underset{\sim}{w}\,, \\ &E(\underset{\sim}{T}'\underset{\sim}{Z}) = \rho\underset{\sim}{1}'\, \underset{\sim}{w},\quad E(\underset{\sim}{Z}'\underset{\sim}{Z}) = \underset{\sim}{1}'\, \underset{\sim}{w}\,, \end{aligned} \quad (2.3.8)$$

we find the following results after differentiating each side of each equation in (2.3.5) with respect to $\mu_x,\mu_y,\sigma_x,\sigma_y$, and ρ and taking expected values of the negations of the resulting second partial derivatives:

$$(\partial^2/\partial\mu_x^2)L_{n,k} = -k/\sigma_x^2(1-\rho^2)$$

$$I_{11} = E\{\ -(\partial^2/\partial\mu_x^2)L_{n,k}\} = \ k/\sigma_x^2(1-\rho^2) \quad (2.3.9)$$

$$(\partial^2/\partial\mu_x\partial\sigma_x)L_{n,k} = -\underset{\sim}{1}'\ (2\ \underset{\sim}{T} - \rho\,\underset{\sim}{Z})/\sigma_x^2(1-\rho^2)$$

$$I_{13} = E\{\ -(\partial^2/\partial\mu_x\partial\sigma_x)L_{n,k}\} = \ \rho\,\underset{\sim}{1}'\underset{\sim}{u}/\sigma_x^2\ (1-\rho^2) \quad (2.3.10)$$

$$(\partial^2/\partial\mu_x\partial\rho)L_{n,k} = \underset{\sim}{1}'\{2\rho\underset{\sim}{T} - (1+\rho^2)\underset{\sim}{Z}\}/\sigma_x(1-\rho^2)^2$$

$$I_{15} = E\{-(\partial^2/\partial\mu_x\partial\rho)\} = \underset{\sim}{1}'\underset{\sim}{u}/\sigma_x(1-\rho^2) \tag{2.3.11}$$

$$(\partial^2/\partial\mu_x\partial\mu_y)L_{n,k} = k\rho/\sigma_x\sigma_y(1-\rho^2)$$

$$I_{12} = E\{-(\partial^2/\partial\mu_x\partial\mu_y)\} = -k\rho/\sigma_x\sigma_y(1-\rho^2) \tag{2.3.12}$$

$$(\partial^2/\partial\mu_x\partial\sigma_y)L_{n,k} = \rho\underset{\sim}{1}'\underset{\sim}{Z}/\sigma_x\sigma_y(1-\rho^2)$$

$$I_{14} = E\{-(\partial^2/\partial\mu_x\partial\sigma_y)L_{n,k}\} = -\rho\underset{\sim}{1}'\underset{\sim}{u}/\sigma_x\sigma_y(1-\rho^2) \tag{2.3.13}$$

$$(\partial^2/\partial\sigma_x^2)L_{n,k} = k/\sigma_x^2 - (3\underset{\sim}{T}'\underset{\sim}{T} - 2\rho\underset{\sim}{T}'\underset{\sim}{Z})/\sigma_x^2(1-\rho^2)$$

$$I_{33} = E\{-(\partial^2/\partial\sigma_x^2)L_{n,k}\} = \{2k(1-\rho^2)+\rho^2\underset{\sim}{1}'\underset{\sim}{w}\}/\sigma_x^2(1-\rho^2) \tag{2.3.14}$$

$$(\partial^2/\partial\sigma_x\partial\mu_y)L_{n,k} = \rho\underset{\sim}{1}'\underset{\sim}{T}/\sigma_x\sigma_y(1-\rho^2)$$

$$I_{23} = E\{-(\partial^2/\partial\sigma_x\partial\mu_y)L_{n,k}\} = -\rho^2\underset{\sim}{1}'\underset{\sim}{u}/\sigma_x\sigma_y(1-\rho^2) \tag{2.3.15}$$

$$(\partial^2/\partial\sigma_x\partial\rho)L_{n,k} = \{2\rho\underset{\sim}{T}'\underset{\sim}{T} - (1+\rho^2)\underset{\sim}{T}'\underset{\sim}{Z}\}/\sigma_x(1-\rho^2)^2$$

$$I_{35} = E\{-(\partial^2/\partial\sigma_x\partial\rho)L_{n,k}\} = \rho(\underset{\sim}{1}'\underset{\sim}{w} - 2k)/\sigma_x(1-\rho^2) \tag{2.3.16}$$

$$(\partial^2/\partial\sigma_x\partial\sigma_y)L_{n,k} = \rho\ \underset{\sim}{T}'\underset{\sim}{Z}/\sigma_x\sigma_y(1-\rho^2)$$

$$I_{34} = E\ \{-(\partial^2/\partial\sigma_x\partial\sigma_y)L_{n,k}\} = -\ \rho^2\ \underset{\sim}{1}'\ \underset{\sim}{w}/\sigma_x\sigma_y(1-\rho^2) \tag{2.3.17}$$

$$(\partial^2/\partial\rho^2)L_{n,k} = \{k(1+\rho^2)(1-\rho^2) - (1+3\rho^2)(\underset{\sim}{T}'\underset{\sim}{T} + \underset{\sim}{Z}'\underset{\sim}{Z})$$
$$+ 2\rho(3+\rho^2)\underset{\sim}{T}'\underset{\sim}{Z}\}/(1-\rho^2)^3$$

$$I_{55} = E\{\ -(\partial^2/\partial\rho^2)L_{n,k}\} = \{2k\rho^2 + (1-\rho^2)\underset{\sim}{1}'\ \underset{\sim}{w}\}/(1-\rho^2)^2 \tag{2.3.18}$$

$$(\partial^2/\partial\rho\partial\mu_y)L_{n,k} = \underset{\sim}{1}'\{2\rho\underset{\sim}{Z} - (1+\rho^2)\underset{\sim}{T}\}/\sigma_y(1-\rho^2)$$

$$I_{52} = E\{\ -(\partial^2/\partial\rho\partial\mu_y)L_{n,k}\} = -\ \rho\,\underset{\sim}{1}'\ \underset{\sim}{u}/\sigma_y(1-\rho^2) \tag{2.3.19}$$

$$(\partial^2/\partial\rho\partial\sigma_y)L_{n,k} = \{2\rho\underset{\sim}{Z}'\underset{\sim}{Z} - (1+\rho^2)\underset{\sim}{T}'\underset{\sim}{Z}\}/\sigma_y(1-\rho^2)^2$$

$$I_{54} = E\{\ -\ (\partial^2/\partial\rho\partial\sigma_y)L_{n,k}\} = -\ \ \rho\,\underset{\sim}{1}'\ \underset{\sim}{w}/\sigma_y(1-\rho^2) \tag{2.3.20}$$

$$(\partial^2/\partial\mu_y^2)L_{n,k} = -(n-k)\sigma_y^{-2}\phi_1(Z_{n,k})[1-\Phi_1(Z_{n,k})]^{-1}$$
$$\{\phi_1(Z_{n,k})[1-\Phi_1(Z_{n,k})]^{-1} - Z_{n,k}\} - k/\sigma_y^2(1-\rho^2)$$

$$I_{22} = E\ \{-(\partial^2/\partial\mu_y^2)L_{n,k}\} = \{(n-k)[\psi(n,k;\ 2,0,2,0)-\psi(n,k;\ 1,0,1,1)]$$
$$+ k(1-\rho^2)^{-1}\}/\sigma_y^2 \tag{2.3.21}$$

$$(\partial^2/\partial\mu_y\partial\sigma_y)L_{n,k} = -(n-k)\phi_1(Z_{n,k})[1-\Phi_1(Z_{n,k})]^{-1}\{1-Z^2_{n,k} + \phi_1(Z_{n,k})[1-\Phi_1(Z_{n,k})]^{-1}Z_{n,k}\}/\sigma^2_y - \underset{\sim}{1}'(\rho\underset{\sim}{T}-2\underset{\sim}{Z})/\sigma^2_y(1-\rho^2)$$

$$I_{24} = E\{-(\partial^2/\partial\mu_y\partial\sigma_y)\}=\{(n-k)[\psi(n,k;1,0,1,0)-\psi(n,k;1,0,1,2) + \psi(n,k;2,0,2,1)] + \underset{\sim}{1}'\underset{\sim}{u}(2-\rho^2)/(1-\rho^2)\}/\sigma^2_y \qquad (2.3.22)$$

$$(\partial^2/\partial\sigma^2_y)L_{n,k} = -(n-k)Z_{n,k}\phi_1(Z_{n,k})[1-\Phi_1(Z_{n,k})]^{-1}\{2-Z^2_{n,k} + Z_{n,k}\phi_1(Z_{n,k})[1-\Phi_1(Z_{n,k})]^{-1}\}/\sigma^2_y + (2\rho\underset{\sim}{T}'\underset{\sim}{Z}-3\,\underset{\sim}{Z}'\underset{\sim}{Z})/\sigma^2_y(1-\rho^2) + k/\sigma^2_y$$

$$I_{44} = E\{-(\partial^2/\partial\sigma^2_y)L_{n,k}\} = \{(n-k)[2\psi(n,k;1,0,1,1) - (n,k;1,0,1,3) + \psi(n,k;2,0,2,2)] + \underset{\sim}{1}'\,\underset{\sim}{w}\,(3-2\rho^2)/(1-\rho^2) - k\}/\sigma^2_y \qquad (2.3.23)$$

and where $I_{j\ell} = I_{\ell j}$ for every $j,\ell = 1,\ldots,5$.

Des Raj (1953), who discussed maximum likelihood estimation from bivariate samples when Y is truncated instead of censored, developed equations similar to these. The expected values of the second partial derivatives are different in that case.

The ψ-functions, which include all moments of single order statistics, can be evaluated for given n,k by numerical integration to almost any desired precision (12 significant digits were used for the computations in this paper). To complete the evaluation of the Fisher information matrix $\underset{\sim}{I} = ((I_{j\ell}))$, we need to evaluate $\underset{\sim}{1}'\underset{\sim}{u}$ and $\underset{\sim}{1}'\underset{\sim}{w}$. For this as in Saw [1958], we note that given $Z_{n,k}$ and disregarding the ordering of $(Z_{n,1},\ldots,Z_{n,k-1})$, the joint distribution of $(Z_{n,1},\ldots,Z_{n,k-1})$ is reducible to that of k-1 independent random variables from a univariate normal d.f., truncated from above by $Z_{n,k}$. Denote $(Z_{n,1},\ldots,Z_{n,k-1})$ in random order by $(S_1,\ldots,S_{k-1})$, so that for $1 \le i \le k-1$,

$$E(S_i|Z_{n,k}) = -\phi_1(Z_{n,k})/\Phi_1(Z_{n,k}), E(S_i^2|Z_{n,k}) = 1-Z_{n,k}\phi_1(Z_{n,k})/\Phi_1(Z_{n,k}) \tag{2.3.24}$$

and, in general, for $r \ge 0$,

$$E(S_i^{r+2}|Z_{n,k}) = (r+1)E(S_i^r|Z_{n,k}) - Z_{n,k}^{r+1}\phi_1(Z_{n,k})/\Phi_1(Z_{n,k}) . \tag{2.3.25}$$

Hence

$$\begin{aligned}\underset{\sim}{1}'\underset{\sim}{u} = \sum_{i=1}^{k} E(Z_{n,i}) &= E(Z_{n,k}) + \sum_{i=1}^{k-1} E(S_i) \\ &= E(Z_{n,k}) + \sum_{i=1}^{k-1} E\{E(S_i|Z_{n,k})\} \\ &= E(Z_{n,k}) + \sum_{i=1}^{k-1} E\{-\phi_1(Z_{n,k})/\Phi_1(Z_{n,k})\} \\ &= \psi(n,k;0,0,0,1) - (k-1)\psi(n,k;1,1,0,0)\end{aligned} \tag{2.3.26}$$

$$\underset{\sim}{1}'\underset{\sim}{w} = \sum_{i=1}^{k} E(Z_{n,i}^2) = E(Z_{n,k}^2) + \sum_{i=1}^{k-1} E(S_i^2)$$

$$= E(Z_{n,k}^2) + \sum_{i=1}^{k-1} E\{E(S_i^2|Z_{n,k})\}$$

$$= E(Z_{n,k}^2) + \sum_{i=1}^{k-1} E\{1-Z_{n,k}\phi_1(Z_{n,k})/\Phi_1(Z_{n,k})\}$$

$$= \psi(n,k;0,0,0,2) + (k-1)[1-\psi(n,k;1,1,0,1)]. \tag{2.3.27}$$

Note that in Saw's notation, $\psi(n,k;\ a,a,o,b)$ is $\psi(k/(n+1),n:a,b)$. Let A_{22}, A_{24}, and A_{44} denote respectively the linear combinations of ψ-functions in (2.3.21), (2.3.22) and (2.3.23). For three selected sample sizes, the values for use in determining the information matrix, $\underset{\sim}{I}$ are:

n	k	A_{22}	A_{24}	A_{44}	$\underset{\sim}{1}'\underset{\sim}{u}$	$\underset{\sim}{1}'\underset{\sim}{w}$
20	10	.618616	.746412	.0471597	-7.67491	10.0000
100	20	.412067	-.000905918	-.276115	-27.7132	43.0867
1000	50	.189672	-.203353	.157445	-102.906	218.888

The Cramèr-Rao lower bounds for variances of unbiased estimators are presented in the following table for three sample sizes, five values of ρ and $\sigma_x = \sigma_y = 1$. These bounds are the diagonal elements of $\underset{\sim}{I}^{-1}$ and also represent approximate variances of the maximum likelihood estimators, $\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho}$ (i.e., solutions to

(2.3.5)) as predicted by asymptotic theory.

TABLE 2.1

CRAMÈR-RAO LOWER BOUNDS FOR VARIANCES

n	k	ρ	$V[\hat{\mu}_x]$	$V[\hat{\mu}_y]$	$V[\hat{\sigma}_x]$	$V[\hat{\sigma}_y]$	$V[\hat{\rho}]$
20	10	0	.2433	.0761	.0500	.0601	.2433
		.1	.2417	.0761	.0514	.0601	.2372
		.3	.2283	.0761	.0618	.0601	.1916
		.5	.2015	.0761	.0775	.0601	.1181
		.9	.1078	.0761	.0787	.0601	.0049
100	20	0	.4598	.0567	.0250	.0345	.2134
		.1	.4557	.0567	.0266	.0345	.2077
		.3	.4235	.0567	.0385	.0345	.1653
		.5	.3590	.0567	.0562	.0345	.0984
		.9	.1333	.0567	.0564	.0345	.0032
1000	50	0	.6170	.0508	.0100	.0155	.1410
		.1	.6114	.0508	.0112	.0155	.1370
		.3	.5661	.0508	.0200	.0155	.1081
		.5	.4755	.0508	.0330	.0155	.0630
		.9	.1584	.0508	.0322	.0155	.0017

It is interesting to note that for a large proportion of censoring (small $\frac{k}{n}$) μ_x cannot be estimated well from induced order statistics, yet σ_x can be estimated with more accuracy than σ_y for small ρ. Also, lower bounds for variance of estimators of μ_y and σ_y are independent of ρ as one would expect; μ_y and σ_y can be estimated efficiently with a small number of order statistics whether or not induced order statistics are observable.

The difficulty in estimating μ_x stems from ρ being unknown, so the selection process on $\underset{\sim}{Y}$ causes an unknown location bias in the sample of induced order statistics. For example, if ρ were known to be zero, μ_x would be estimated by the sample mean with variance $.02\ \sigma_x^2$ at $n=1000$, $k=50$, instead of $.6170\ \sigma_x^2$.

2.4. Maximum Likelihood Estimation

An iterative method may be used to estimate $\underset{\sim}{\theta}$ by using (2.3.5). Des Raj [1953] set up the implicit solutions to the likelihood equations for the case of double truncation on Y from a bivariate normal sample which is similar to our case relating to censoring. Assuming the theorem outlined by Halperin[1952] to be valid for our bivariate case, the maximum likelihood estimator (MLE) will be consistent, asymptotically normally distributed and asymptotically efficient. Note that since $Y_{n,1},\ldots,Y_{n,k}$ are not independent, knowledge of the asymptotic properties of the MLE may demand additional regularity conditions [viz. Sen [1976]]. Nevertheless, for bivariate normal d.f.'s, optimal properties hold under quite general setups.

The complexity of the solutions to (2.3.5) makes the study of the properties of the MLE difficult for any finite sample size. It will be shown in a more general setting in section 4.2, however, that the MLE of μ_y, σ_y is the usual MLE based on $Y_{n,1},\ldots,Y_{n,k}$ alone. Nonetheless, the results obtained by Saw [1961] for estimating μ_y and σ_y in this case implies that the MLE of $\underset{\sim}{\theta}$ will have undesirable properties for small k or n. In particular, the MLE may have considerable bias (relative to its standard error). The linearization of the log-likelihood function of $Y_{n,1},\ldots,Y_{n,k}$ described in Chan [1967], which results in simple and efficient estimators of μ_y and σ_y, will solve this problem.

There are other good estimators of μ_y and σ_y besides the ones considered by Chan. For instance, Saw [1959] developed an unbiased estimator of μ_y which is a linear combination of $Y_{n,k}$ and $(k-1)^{-1}\sum_{i=1}^{k-1} Y_{n,i}$ and has asymptotic efficiency (a function of $k/(n+1)$) $\geq 94\%$ for $k/(n+1)$ as low as .25. He also considered a linear combination of $\sum_{i=1}^{k-1}(Y_{n,i}-Y_{n,k})^2$ and $\{\sum_{i=1}^{k-1}(Y_{n,i}-Y_{n,k})\}^2$ as an unbiased estimator of σ_y^2 which is of 100% asymptotic efficiency for all $k/(n+1)$. Linear unbiased estimators of σ_y have also been developed (for example see Gupta [1952]).

Since for the estimation of (μ_y,σ_y), the induced order statistics do not contribute any information, it seems natural to estimate (μ_y,σ_y) by some efficient procedure based only on $\underset{\sim}{Y} = (Y_{n,1},\ldots,Y_{n,k})$. Denote the estimators by (μ_y^*,σ_y^*) and then consider only the first three equations in (2.3.5) wherein μ_y,σ_y are replaced by μ_y^*,σ_y^*.

Equating these functions to 0, we get three equations in three unknown parameters (μ_x, σ_x, ρ) and denote these solutions $(\mu_x^*, \sigma_x^*, \rho^*)$ as the modified MLE. For this purpose, we write

$$\bar{X} = k^{-1} \sum_{i=1}^{k} X_{n[i]},\ \bar{Y} = k^{-1} \sum_{i=1}^{k} Y_{n,i},\ S_x^2 = k^{-1} \sum_{i=1}^{k} (X_{n[i]} - \bar{X})^2,$$
$$S_y^2 = k^{-1} \sum_{i=1}^{k} (Y_{n,i} - \bar{Y})^2 \text{ and } S_{xy} = k^{-1} \sum_{i=1}^{k} (X_{n[i]} - \bar{X})(Y_{n,i} - \bar{Y}). \tag{2.4.1}$$

Equating $(\partial/\partial\mu_x)L_{n,k}$ in (2.3.5) to 0 we get

$$(\bar{X} - \mu_x^*)/\sigma_x^* = \rho^*(\bar{Y} - \mu_y^*)/\sigma_y^*. \tag{2.4.2}$$

Equating $(\partial/\partial\sigma_x)L_{n,k}$ to 0, we get

$$\sigma_x^{*2} = S_x^2 + \sigma_y^{*2}(\bar{X}-\mu_x^*)^2/(\bar{Y}-\mu_y^*)^2 - S_{xy}(\bar{X}-\mu_x^*)/(\bar{Y}-\mu_y^*). \tag{2.4.3}$$

Combining (2.4.2) and (2.4.3) we have

$$\rho^* S_{xy}/\sigma_x^*\sigma_y^* - \rho^{*2} = S_x^2/\sigma_x^{*2} - 1. \tag{2.4.4}$$

Equating $(\partial/\partial\rho)L_{n,k}$ to 0,

$$\rho^*(1-\rho^{*2}) - \rho^*[\{S_x^2 + (\bar{X}-\mu_x^*)^2\}/\sigma_x^{*2} + \{S_y^2 + (\bar{Y}-\mu_y^*)^2\}/\sigma_y^{*2}]$$

$$+ (1+ \rho^{*2})(\sigma_x^* \sigma_y^*)^{-1} \{S_{xy} + (\bar{X}-\mu_x^*)(\bar{Y}-\mu_y^*)\} = 0. \qquad (2.4.5)$$

Using (2.4.2) this simplifies to

$$\rho^*(1-\rho^{*2}) - \rho^*(S_x^2/\sigma_x^{*2} + S_y^2/\sigma_y^{*2}) + (1-\rho^{*2}) S_{xy}/\sigma_x^* \sigma_y^*$$

$$= \rho^*(\rho^* S_{xy}/\sigma_x^* \sigma_y^* - \rho^{*2}) + S_{xy}/\sigma_x^* \sigma_y^*$$

$$- \rho^*(S_x^2/\sigma_x^{*2} + S_y^2/\sigma_y^{*2}) + \rho^* = 0. \qquad (2.4.6)$$

Substituting the identity in (2.4.4) into (2.4.6),

$$\rho^* = \sigma_y^* S_{xy}/\sigma_x^* S_y^2. \qquad (2.4.7)$$

The similarity of ρ^* to the Pearson product moment correlation should be noted, for (2.4.7) can be re-written

$$\rho^* = (S_{xy}/S_x S_y) \; [(S_x/\sigma_x^*)(\sigma_y^*/S_y)]. \qquad (2.4.8)$$

From (2.4.4) and (2.4.7) we get

$$\sigma_x^{*2} = S_x^2 + (S_{xy}^2/S_y^2) \; (\sigma_y^{*2}/S_y^2 - 1). \qquad (2.4.9)$$

Substituting (2.4.7) into (2.4.2),

$$\mu_x^* = \bar{X} + (S_{xy}/S_y^2)(\mu_y^* - \bar{Y}). \tag{2.4.10}$$

Thus, having estimated μ_y, σ_y by μ_y^*, σ_y^*, we may estimate μ_x, σ_x and ρ by (2.4.10), (2.4.9), and (2.4.7), respectively. One nice feature of these estimators is that as $k/n \to 1$ (i.e., the amount of censoring decreases), the estimators approach the ordinary MLE, which are known to be efficient for complete samples of any size. Again, if μ_y^*, σ_y^* are MLE based on $Y_{n,1}, \ldots, Y_{n,k}$ the entire set of estimators is MLE.

Since $E(\underset{\sim}{X}|\underset{\sim}{Y}) = \mu_x \underset{\sim}{1} + \rho\sigma_x(\underset{\sim}{Y} - \mu_y \underset{\sim}{1})/\sigma_y$ it follows that $E(\bar{X}|\underset{\sim}{Y}) = \mu_x + \rho\sigma_x(\bar{Y} - \mu_y)/\sigma_y$ and

$$E(S_{xy}|\underset{\sim}{Y}) = E(k^{-1} \sum_{i=1}^{k} X_{n[i]}(Y_{n,i} - \bar{Y})|\underset{\sim}{Y})$$

$$= k^{-1} \sum_{i=1}^{k} [\mu_x + \rho\sigma_x(Y_{n,i} - \mu_y)/\sigma_y](Y_{n,i} - \bar{Y})$$

$$= \rho\sigma_x S_y^2/\sigma_y . \tag{2.4.11}$$

And so from (2.4.10),

$$E(\mu_x^*|\underset{\sim}{Y}) = \mu_x + \rho\sigma_x(\mu_y^* - \mu_y)\sigma_y . \tag{2.4.12}$$

Therefore, μ_x^* is an unbiased estimator of μ_x if μ_y^* is an unbiased estimator of μ_y. Since $V(\underset{\sim}{X}|\underset{\sim}{Y}) = \sigma_x^2(1-\rho^2)\underset{\sim}{I}_k$,

$$V(S_{xy}|\underset{\sim}{Y}) = \sum_{i=1}^{k} [(Y_{n,i}-\overline{Y})^2/k]^2 \, \sigma_x^2(1-\rho^2) = \sigma_x^2(1-\rho^2)S_y^2/k. \tag{2.4.13}$$

Combining (2.4.13) and (2.4.11) we find

$$E(S_{xy}^2|\underset{\sim}{Y}) = V(S_{xy}|\underset{\sim}{Y}) + E(S_{xy}|\underset{\sim}{Y})^2 = \sigma_x^2 S_y^2\{(1-\rho^2)/k + \rho^2 S_y^2/\sigma_y^2\} \tag{2.4.14}$$

It can also be shown that $E(S_x^2|\underset{\sim}{Y}) = \sigma_x^2\{k^{-1}(k-1)(1-\rho^2)+\rho^2 S_y^2/\sigma_y^2\}$. Combining this with (2.4.14) and (2.4.9) we have

$$E(\sigma_x^{*2}) = \sigma_x^2 \, \{1 + \rho^2[E(\sigma_y^{*2})/\sigma_y^2 - 1] + k^{-1}(1-\rho^2) \, E(\sigma_y^{*2}/S_y^2)-2]\} \tag{2.4.15}$$

and if σ_y^{*2} is unbiased for σ_y^2,

$$E(\sigma_x^{*2}) = \sigma_x^2 \, \{1 + k^{-1}(1-\rho^2)[E(\sigma_y^{*2}/S_y^2)-2]\}. \tag{2.4.16}$$

It can also be shown that

$$V(\mu_x^*) = \sigma_x^2(1-\rho^2)[1+E\{(\mu_y^*-\overline{Y})^2/S_y^2\}]/k + \rho^2\sigma_x^2 V(\mu_y^*)/\sigma_y^2. \tag{2.4.17}$$

Note that from (2.4.11) if σ_y^{*2} is unbiased for σ_y^2 then an unbiased estimator for the population covariance of X and Y, $\rho\sigma_x\sigma_y$, is given by

$$\sigma_y^{*2} \; S_{xy}/S_y^2 . \tag{2.4.18}$$

In order to estimate the expected values of σ_x^* and ρ^* as well as the variances of μ_x^*, σ_x^*, and ρ^*, for each of three different combinations of (n,k) and five different values of ρ, 500 random sets of $(\underset{\sim}{X},\underset{\sim}{Y})$ were generated and the statistics were evaluated. μ_y^*, σ_y^* were taken as the estimators due to Gupta (1952) in (1.5.14) and (1.5.15). Also, for comparison, the variance of Watterson's estimator of μ_x, $\tilde{\mu}_x$ defined by (1.5.19) was estimated. For μ_y^* and $\tilde{\mu}_x$ the $u_{n,i}$ in (1.5.14) were approximated by (1.5.9). For simplicity, we have taken $\sigma_x = \sigma_y = 1$. We have also estimated the <u>trace efficiency</u> of the modified MLE, which is the ratio of the sum of the five Cramèr-Rao lower bounds from Table 2.1 to the sum of the estimated variances of the five modified MLE.

Table 2.2. Estimated moments of modified maximum likelihood estimators

n	k	ρ	$V[\mu_x^*]$	$V[\tilde{\mu}_x]$	$V[\mu_y^*]$	$E[\sigma_x^*]$	$V[\sigma_x^*]$	$V[\sigma_y^*]$	$E[\rho^*]$	$V[\rho^*]$	Trace Efficiency
20	10	0	.281	.297	.074	1.021	.077	.078	-.020	.236	.902
		.1	.287	.293	.092	.998	.074	.088	.060	.213	.884
		.3	.281	.293	.082	1.043	.079	.079	.275	.196	.862
		.5	.265	.276	.087	.998	.102	.086	.449	.172	.749
		.9	.123	.126	.093	1.014	.106	.083	.881	.024	.764
100	20	0	.495	.508	.076	1.038	.042	.045	-.044	.169	.955
		.1	.548	.564	.087	1.059	.047	.051	.080	.176	.859
		.3	.487	.507	.090	1.054	.050	.049	.256	.152	.868
		.5	.400	.420	.081	1.020	.057	.046	.442	.115	.865
		.9	.202	.209	.087	1.014	.080	.048	.885	.009	.667
1000	50	0	.663	.685	.078	1.045	.019	.022	-.016	.114	.931
		.1	.601	.616	.091	1.044	.018	.025	.092	.105	.983
		.3	.624	.648	.094	1.048	.024	.024	.274	.103	.875
		.5	.477	.508	.091	1.030	.032	.025	.462	.066	.923
		.9	.187	.188	.082	.987	.038	.023	.887	.003	.777

From Table 2.2 we see that μ_x^* is consistently slightly better than $\tilde{\mu}_x$ although a small part of this may have to do with the approximation used for the $u_{n,i}$. The efficiency of μ_x^* by simulation is between 66% and 102% according to the bounds in Table 2.1. As expected, μ_x^* has large variance for large censoring proportion. On the other hand, σ_x is estimated as well as σ_y in almost every case and has lower variance in some cases. Also, σ_x^* appears to have small bias and is 53-103% efficient. The efficiency of ρ^* is between 69% and 130% for the cases studied except when $\rho = .9$. The bias in ρ^* is very low except for moderate ρ ($\rho = .3$ or .5 resulted in maximum estimated bias of .06). Note that $V(\mu_x^*)$ would be even smaller if for example Saw's [1959] estimator of μ_y were used.

2.5. Test of H_0: $\rho = 0$

2.5.1. Likelihood Ratio Test

From the results of the previous section, if (μ_y^*, σ_y^*) are MLE of (μ_y, σ_y), they are MLE independent of any restriction on ρ. Thus the modified MLE of section 2.4 may be incorporated to prescribe a likelihood ratio test for H_0: $\rho = 0$. As before, $\mu_y^*, \sigma_y^*; \mu_x^*, \sigma_x^*, \rho^*$ are the estimators considered in section 2.4 except that now we require μ_y^*, σ_y^* to be true MLE. Over the parameter space restricted by $\rho = 0$, the parallel estimators are $\mu_y^*, \sigma_y^*; \tilde{\mu}_y^*, \tilde{\sigma}_x^*$, 0 where

$$\tilde{\mu}_x^* = \overline{X} \text{ and } \tilde{\sigma}_x^{*2} = S_x^2 \text{ are defined by (2.4.1)} \tag{2.5.1.1}$$

From the definition of $L_{n,k}$ in (2.3.4), the log ratio of the maximum likelihood under the null hypothesis and the maximum likelihood over the whole parameter space, multiplied by -2, results in a likelihood ratio statistic of

$$\begin{aligned}\chi = {} & k \log [\tilde{\sigma}_x^{*2}/\sigma_x^{*2}(1-\rho^{*2})] + \tilde{\sigma}_x^{*-2} \Sigma(X_{n[i]} - \tilde{\mu}_x^*)^2 \\ & + \sigma_y^{*-2} \Sigma(Y_{n,i} - \mu_y^*)^2 \\ & - (1-\rho^*)^{-1} \Sigma\{\sigma_x^{*-2}(X_{n[i]}-\mu_x^*)^2 - 2\rho^* \sigma_x^{*-1}\sigma_y^{*-1}(X_{n[i]}-\mu_x^*)(Y_{n,i}-\mu_y^*) \\ & + \sigma_y^{*-2}(Y_{n,i} - \mu_y^*)^2\}\end{aligned} \tag{2.5.1.2}$$

where all summations are from 1 to k. We now use the relation

$$\Sigma(Y_{n,i} - \mu_y^*)^2 = k\, S_y^2 + k(\overline{Y} - \mu_y^*)^2 \tag{2.5.1.3}$$

and further, by (2.4.10),

$$\Sigma(X_{n[i]} - \mu_x^*)^2 = k\, S_x^2 + k\, S_{xy}^2(\overline{Y} - \mu_y^*)^2/S_y^4 \tag{2.5.1.4}$$

$$\Sigma(X_{n[i]} - \mu_x^*)(Y_{n,i} - \mu_y^*) = k\, S_{xy} + k\, S_{xy}(\overline{Y} - \mu_y^*)^2/S_y^2 \tag{2.5.1.5}$$

and using (2.4.4) and (2.4.7),

$$\sigma_y^{*2}(1-\rho^{*2}) = (S_x^2 S_y^2 - S_{xy}^2)/S_y^2 \tag{2.5.1.6}$$

so the first term in (2.5.1.2) is

$$-k \log (1-r^{*2}) \text{ where } r^* = S_{xy}/S_x S_y \tag{2.5.1.7}$$

and S_x, S_y, S_{xy} are defined by (2.4.1). Thus r^* is the ordinary product moment correlation of the set $(X_{n[i]}, Y_{n,i})$, $i=1,\ldots,k$. The remaining terms of χ become

$$k + \sigma_y^{*-2} \Sigma(Y_{n,i} - \mu_y^*)^2 - S_y^2(S_x^2 S_y^2 - S_{xy}^2)^{-1} \Sigma(X_{n[i]} - \mu_x^*)^2$$

$$+ 2\, S_{xy}\{\sigma_x^{*2}(1-\rho^{*2}) S_y^2\}^{-1} \Sigma(X_{n[i]} - \mu_x^*)(Y_{n,i} - \mu_y^*)$$

$$- (1-\rho^{*2})^{-1} \sigma_y^{*-2} \Sigma(Y_{n,i} - \mu_y^*)^2.$$

Using (2.5.1.3) through (2.5.1.6), we get, on dividing by k,

$$1 + \sigma_y^{*-2} \{S_y^2 + (\bar{Y} - \mu_y^*)^2\} - (S_x^2 S_y^2 - S_{xy}^2)^{-1}\{S_x^2 S_y^2 + S_{xy}^2 (\bar{Y} - \mu_y^*)^2/S_y^2\}$$

$$+ 2\, S_{xy}(S_x^2 S_y^2 - S_{xy}^2)^{-1}\{S_{xy} + S_{xy}(\bar{Y} - \mu_y^*)^2/S_y^2\}$$

$$-(S_x^2S_y^4 - S_y^2S_{xy}^2)^{-1}(S_x^2S_y^4 - S_y^2S_{xy}^2 + \sigma_y^{*2}S_{xy}^2)\sigma_y^{*^{-2}}\{S_y^2 + (\bar{Y} - \mu_y^{*2})\}$$

which after simplification is seen to be zero. Hence the likelihood ratio statistic is given by (2.4.1.7) which is identical to the likelihood ratio statistic for a complete sample of size k. This suggests that an "optimum" test of independence using censored data would still be the same test as if the data were not censored and that we may use r^* (or some function of r^*) as a test statistic. As with a complete sample it will be advantageous to consider as a test statistic

$$T^* = (k-2)r^{*2}/(1-r^{*2}) = S_{xy}^2/\{(k-2)^{-1}[S_x^2S_y^2 - S_{xy}^2]\}. \tag{2.5.1.8}$$

Note that this test is identical to that one would obtain from the usual conditional regression test based on the conditional distribution in (2.3.1).

2.5.2. Distribution of Test Statistic under Censoring

A statistic for testing the hypothesis that X and Y are independent is T^* in (2.5.1.8). Define

$$U^* = k\, S_y^2/\sigma_y^2 = \sigma_y^{-2} \sum_{i=1}^{k} (Y_{n,i} - \bar{Y})^2. \tag{2.5.2.1}$$

Let F_Δ (t;a,b) be the non-central F d.f. with degrees of freedom

(DF) (a,b) and non-centrality parameter Δ and $G_{n,k}(u) = P\{U^* \leq u\}$ for $u \geq 0$. Finally, let

$$F_\rho^*(t;a,b) = \int_0^\infty F_{\rho^2 u/(1-\rho^2)}(t;a,b)\,d\,G_{n,k}(u), \quad t \geq 0 . \tag{2.5.2.2}$$

Then, we have the following

Theorem 2.5.1.1:
For every $t \geq 0$ and given U^*, $P\{T^* \leq t | U^*\} = F_{\rho^2(1-\rho^2)^{-1}U^*}(t;1,k-2)$.

(2.5.2.3)

Hence $P\{T^* \leq t\} = F_\rho^*(t;\, 1,k-2)$, $t \geq 0$. Under $H_0: \rho = 0$, $F_\rho^*(t;1,k-2) = F_0(t;1,k-2)$ is the central F d.f. with DF (1,k-2).

Proof: From (1.5.3), given $\underset{\sim}{Y}$, $X_{n[1]}, \ldots, X_{n[k]}$ are (conditionally) independently normally distributed with means $\mu_x + \rho\sigma_x\sigma_y^{-1}(Y_{n,i} - \mu_y)$, $1 \leq i \leq k$ and a common variance $\sigma_x^2(1-\rho^2)$. Hence $k\,S_{xy}$ is conditionally normally distributed with mean

$$(\underset{\sim}{Y} - \overline{Y}\,\underset{\sim}{1})'\{\mu_x\,\underset{\sim}{1} + \rho\sigma_x\sigma_y^{-1}(\underset{\sim}{Y} - \mu_y\,\underset{\sim}{1})\} = k\,\rho\sigma_x S_y^2/\sigma_y = \rho\sigma_x\sigma_y U^*$$

and variance

$$\sigma_x^2(1-\rho^2)(\underset{\sim}{Y} - \overline{Y}\,\underset{\sim}{1})'\,\underset{\sim}{I}_k(\underset{\sim}{Y} - \overline{Y}\,\underset{\sim}{1}) = \sigma_x^2\sigma_y^2(1-\rho^2)U^*.$$

Therefore

$$k^2\, S_{xy}^2/\{\sigma_x^2\sigma_y^2(1-\rho^2)U^*\} \sim \chi_1^2(\rho^2 U^*(1-\rho^2)^{-1}), \tag{2.5.2.4}$$

where $\chi_p^2(\Delta)$ stands for the non-central chi-square d.f. with p DF and non-centrality parameter Δ. Further,

$$k^2(S_x^2 S_y^2 - S_{xy}^2)/\{\sigma_x^2\sigma_y^2(1-\rho^2)U^*\}$$

$$= [\sigma_y^2 U^* \underset{\sim}{X}'(\underset{\sim}{I}_k - k^{-1}\, \underset{\sim}{1}\, \underset{\sim}{1}')\underset{\sim}{X} - \{\underset{\sim}{X}'(\underset{\sim}{Y} - \overline{Y}\, \underset{\sim}{1})\}^2]/[\sigma_x^2\sigma_y^2(1-\rho^2)U^*]$$

$$= \underset{\sim}{X}'\ [\sigma_y^2 U^*(\underset{\sim}{I}_k - k^{-1}\underset{\sim}{1}\, \underset{\sim}{1}') - (\underset{\sim}{Y} - \overline{Y}\, \underset{\sim}{1})(\underset{\sim}{Y} - \overline{Y}\, \underset{\sim}{1})']\ \underset{\sim}{X}/[\sigma_x^2\sigma_y^2(1-\rho^2)U^*]$$

$$= \underset{\sim}{U}'\ \underset{\sim}{A}\ \underset{\sim}{U} \tag{2.5.2.5}$$

where $$\underset{\sim}{U} = \{\sigma_x^2\sigma_y^2(1-\rho^2)U^*\}^{-1/2}\ \underset{\sim}{X} \tag{2.5.2.6}$$

$$\underset{\sim}{A} = \sigma_y^2 U^*(\underset{\sim}{I}_k - k^{-1}\ \underset{\sim}{1}\ \underset{\sim}{1}') = (\underset{\sim}{Y} - \overline{Y}\ \underset{\sim}{1})(\underset{\sim}{Y} - \overline{Y}\ \underset{\sim}{1})' \tag{2.5.2.7}$$

and given $\underset{\sim}{Y}$, $\underset{\sim}{U}$ is conditionally normally distributed with mean vector

$$\underset{\sim}{\gamma} = \{\sigma_x^2\sigma_y^2(1-\rho^2)U^*\}^{-1/2}\{\mu_x\, \underset{\sim}{1} + \rho\sigma_x\sigma_y^{-1}(\underset{\sim}{Y} - \mu_y\, \underset{\sim}{1})\} \tag{2.5.2.8}$$

and dispersion matrix $k^{-1}S_y^{-2}\underset{\sim}{I}_k$. Further, it can be shown that $k^{-1}S_y^{-2}\underset{\sim}{A}$ is an idempotent matrix and that the rank of $\underset{\sim}{A} = k-2$. The non-centrality parameter, $\underset{\sim}{\gamma}' A \underset{\sim}{\gamma}$, is zero. Thus given $\underset{\sim}{Y}$,

$$k^2(S_x^2S_y^2 - S_{xy}^2)/\{\sigma_x^2\sigma_y^2(1-\rho^2)U^*\} \sim \chi^2_{k-2}(0), \tag{2.5.2.9}$$

which implies its unconditional distribution is also $\chi^2_{k-2}(0)$. Finally, noting that $(\underset{\sim}{Y}-\overline{Y}\,\underset{\sim}{1})'\,\underset{\sim}{A} = 0$, we obtain from (2.5.2.5) and (2.5.2.6) that given $\underset{\sim}{Y}$, S_{xy}^2 and $\{S_x^2S_y^2 - S_{xy}^2\}$ are conditionally independent. Hence (2.5.2.3) follows from (2.5.1.8), (2.5.2.4), and (2.5.2.9), and the fact that this conditional d.f. depends on $\underset{\sim}{Y}$ through U^* alone. Finally, by (2.5.2.3), $P(T^* \leq t) = E\{P(T^* \leq t | U^*) = F_\rho^*(t;\ 1,k-2)$. Under H_0: $\rho = 0$, this is equal to $F_0(t;\ 1,k-2)$ for all U^* and hence, $F_\rho^* = F_0$. Q.E.D.

From Theorem 2.5.2.1, T^* has the same comditional distribution as the statistic from an uncensored sample and under H_0: $\rho = 0$, it has the classical (variance ratio) distribution with DF (1,k-2) and the test can be made without any difficulty. Let F_α^* be the upper 100 α% point of $F_0(t;\ 1,k-2)$ i.e., $F_0(F_\alpha^*;\ 1,\ k-2) = 1 - \alpha$. Then the power of the test based on T^* (or r^*) and corresponding to the level of significance $\alpha\,(0 < \alpha < 1)$ is given by

$$1 - F_\rho^*(F_\alpha^*\ ;\ 1,\ k-2)\ . \tag{2.5.2.10}$$

In general, for $k < n$, $G_{n,k}$ the d.f. of U^* is quite complicated, and hence, analytical solutions for (2.5.2.10) are difficult to obtain. However, one can obtain one or two moment approximations for it by computing the moments of U^* by the formulae of Saw [1958]. Unfortunately these formulae contain additional errors not reported in the Corrigenda to Saw [1958]. Let $\psi_{ab} = \psi(n,k;\ a,a,0,b)$, defined by (2.3.6). Then, noting that $U^* = (\underset{\sim}{Z}-\overline{Z}\ \underset{\sim}{1})'\ (\underset{\sim}{Z}-\overline{Z}\ \underset{\sim}{1})$ where $\underset{\sim}{Z} = (Z_{n,1},\ldots,Z_{n,k})'$, $\overline{Z} = k^{-1}\sum_{i=1}^{k} Z_{n,i}$, with $Z_{n,i}=(Y_{n,i}-\mu_y)/\sigma_y$, $1 \le i \le k$ relate to the standard normal d.f., we find after deriving Saw's formulae that

$$E(U^*) = k^{-1}(k-1)[(k-1)-(k-3)\psi_{11}+\psi_{02}-(k-2)\psi_{20}] = kE(S_y^2/\sigma_y^2)$$

and that the coefficient of ψ_{13} in Saw's expression (3.6) concerning $E(U^{*2})$ should be $-(3k^2 - 14k + 15)$ instead of $(5k^2 - 10k + 1)$. These findings have been confirmed by Professor Saw in a personal communication. Note that for a complete sample $(k=n)$, $E(U^*)=k-1$, $V(U^*)=2(k-1)$.

Letting $\zeta = \rho^2(1-\rho^2)^{-1}$ and using the properties of the F-distribution,

$$E(T^*|U^*) = (1 + \zeta\ U^*)(k-2)/(k-4) \tag{2.5.2.11}$$

$$E(T^{*2}|U^*) = (3 + 6\zeta\ U^* + \zeta^2U^{*2})(k-2)^2/(k-4)(k-6) \tag{2.5.2.12}$$

which imply that

$$E(T^*) = \{1 + \zeta E(U^*)\}\ (k-2)/(k-4) \tag{2.5.2.13}$$

$$E(T^{*2}) = \{3 + 6\ \zeta E(U^*) + \zeta^2 E(U^{*2})\}(k-2)^2/(k-4)(k-6). \tag{2.5.2.14}$$

For the one moment approximation, we suppose that $T^* \sim F_{1,k-2}(\lambda)$ where $F_{a,b}(\Delta)$ has the d.f. $F_\Delta(t;\ a,b)$ defined after (2.5.2.1). Thus we require $\lambda = \zeta E(U^*)$ and replace the distribution of T^* with $F_{1,k-2}(\zeta E(U^*))$. For the two moment approximation, we replace the distribution of T^* by $aF_{b,k-2}(0)$ and equate its moments to (2.5.2.13), (2.5.2.14). Thus

$$E(T^*) = a(k-2)/(k-4) \equiv \{1 + \zeta E(U^*)\}(k-2)/(k-4) \quad \text{so}$$

$$a = 1 + \zeta E(U^*) \tag{2.5.2.15}$$

$$E(T^*) = a^2(k-2)^2(b+2)/b(k-4)(k-6) \quad \text{which implies}$$

$$b = 2a^2/\{\zeta^2 V(U^*) + 4\ \zeta E(U^*) + 2\}. \tag{2.5.2.16}$$

Note that when $\rho = 0$, $a = b = 1$ and $\lambda = 0$ so that both approximations give the correct null d.f. Also, we limit ourselves to $k > 4$

for the one-moment approximation or $k > 6$ for the two moment approximation.

For the different n,k the moments of U^* are as follows:

n	k	$E(U^*)$	$V(U^*)$
20	10	3.43246	3.87797
100	20	4.22029	3.60135
1000	50	6.79179	4.41258

The resulting (a,b) values are:

	n=20	k=10				n=100	k=20			
ρ	0	.1	.3	.5	..9	0	.1	.3	.5	.9
a	1	1.035	1.339	2.144	15.633	1	1.043	1.417	2.407	18.992
b	1	1.001	1.057	1.312	3.731	1	1.002	1.085	1.443	5.174

		n=1000		k=50	
ρ	0	.1	.3	.5	.9
a	1	1.069	1.672	3.264	29.954
b	1	1.004	1.182	1.845	9.063

For each n,k,ρ the power of the r^*-test at the 5% level under censoring was estimated by generating 2000 $k \times 1$ random $\underset{\sim}{Y}$ vectors. The exact upper 5% critical values were used and the power was calculated by averaging 2000 conditional non-central F probabilities to estimate the unconditional distribution of U^* as in (2.5.2.2). In order to compare the power of the test using censored data to that using a complete bivariate sample, the complete sample size n^* required to achieve the same simulated power using the Fisher-Yates Z-test is given.

TABLE 2.3

ESTIMATED POWER OF 5% r^*-TEST FOR CENSORED DATA

n	k	ρ	Power from One Moment Approximation	Power from Two Moment Approximation	Simulated Power	n^*
20	10	.1	.0531	.0531	.0532	6
		.3	.0810	.0802	.0814	6
		.5	.1570	.1522	.1581	6
		.9	.9160	.8331	.8256	7
100	20	.1	.0544	.0544	.0544	7
		.3	.0939	.0921	.0942	7
		.5	.2025	.1929	.2026	8
		.9	.9796	.9468	.9349	9
1000	50	.1	.0576	.0575	.0576	10
		.3	.1266	.1219	.1273	10
		.5	.3140	.2967	.3140	11
		.9	.9995	.9985	.9956	13

The two-moment approximation is better for $\rho = .9$, but the one-moment approximation is better overall in estimating the power function. This is fortunate because the one-moment formula is easier to calculate since only the first moment of U^* need be computed. Note that a complete sample of about one-fifth the size (k) of the observed censored sample would achieve the same power when $n = 1000$ $k = 50$ in which a high proportion of observations are censored. However, because of considerations of cost of failures or controlling the duration of the experiment, this type of sacrifice may be necessary.

CHAPTER III

STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE NORMAL DISTRIBUTIONS BASED ON INDUCED ORDER STATISTICS

3.1. Introduction.

Let $(\underset{\sim}{X}_1, Y_1), \ldots, (\underset{\sim}{X}_n, Y_n)$ be independent and identically distributed random (p+1)-vectors where $\underset{\sim}{X}_i$ is a p-vector. There is no difficulty in (1.2.1) in defining $\underset{\sim}{X}_{n[i]}$ as a $1 \times p$ induced order statistic where $\underset{\sim}{X}_{n[i]} = \underset{\sim}{X}_j$ if $Y_{n,i} = Y_j$ for $i,j=1,\ldots,n$. In this chapter we will extend the results of Chapter 2 by supposing that $(\underset{\sim}{X}_i, Y_i)$ has a (p+1)-variate normal d.f. with density $\phi_{\underset{\sim}{\theta}}(x,y)$ where $\underset{\sim}{\theta}$ denotes the mean, variance, and covariance parameters. Maximum likelihood estimation as in section 2.4 will be used to estimate $\underset{\sim}{\theta}$ and a likelihood ratio test of H_0: $\underset{\sim}{X}$ and Y are independent will be developed as in section 2.5.1. Other tools of statistical inference commonly used in complete data problems, namely tests of partial and canonical correlation, will be developed for censored multivariate normal data.

3.2. Maximum Likelihood Estimation

By (1.4.1), the joint (conditional) pdf of

$\underset{\sim}{X} = (X'_{n[1]}, X'_{n[2]}, \ldots, X'_{n[k]})'$ given $\underset{\sim}{Y} = (Y_{n,1}, \ldots, Y_{n,k})'$ is

$$\prod_{i=1}^{k} \phi_{y_{n,i}}(\underset{\sim}{x}_i) = \prod_{i=1}^{k} \{\phi_{\underset{\sim}{\theta}}(\underset{\sim}{x}_i, Y_{n,i})/\phi_1((Y_{n,i}-\mu_y)/\sigma_y)/\sigma_y\} \tag{3.2.1}$$

where $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ and $\phi_{y_{n,i}}(\underset{\sim}{x}_i)$ is the conditional density of $\underset{\sim}{X}_i$ given $Y_{n,i}$. Multiplying (3.2.1) by (2.3.2) we obtain the joint pdf of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$,

$$n^{[k]}\{\prod_{i=1}^{k} \phi_{\underset{\sim}{\theta}}(\underset{\sim}{x}_i, y_i)\}\ [1- \Phi_1((y_k - \mu_y)/\sigma_y)]^{n-k}\ , \tag{3.2.2}$$

$$y_1 \le y_2 \le \cdots \le y_k$$

where $n^{[k]}$ is defined in (2.3.2). The log likelihood function based on (3.2.2) is, aside from constant terms,

$$L_{n,k} = (n-k)\log\ [1- \Phi_1((Y_{n,k} - \mu_y)/\sigma_y)] - (k/2)\log |\underset{\sim}{V}| \tag{3.2.3}$$

$$- \frac{1}{2} \sum_{i=1}^{k} (\underset{\sim}{U}_i - \underset{\sim}{\mu})'\ \underset{\sim}{V}^{-1}\ (\underset{\sim}{U}_i - \underset{\sim}{\mu})$$

where $\underset{\sim}{U}_i = (\underset{\sim}{X}_{n[i]}, Y_{n,i})$, $\underset{\sim}{V}$ is the dispersion matrix of $(\underset{\sim}{X}_i, Y_i)$, and $\underset{\sim}{\mu} = (\underset{\sim}{\mu}_x, \mu_y)$ where $\underset{\sim}{\mu}_x = E(\underset{\sim}{X}_i)$, $\mu_y = E(Y_i)$. Partition $\underset{\sim}{V}$ as

$$\underset{\sim}{V} = \begin{bmatrix} \underset{\sim}{V}_{xx}\ , & \underset{\sim}{V}_{xy} \\ \underset{\sim}{V}'_{xy}\ , & \sigma_y^2 \end{bmatrix} \tag{3.2.4}$$

where V_{xx} is the $p \times p$ dispersion matrix of X_i, V_{xy} is the $p \times 1$ covariance vector of X_i and Y_i, and σ_y^2 is the (scaler) variance of Y_i. Following section 2.4 we estimate μ_y and σ_y^2 using the vector Y only; denote these estimators by μ_y^* and σ_y^{2*} respectively. For simplicity, the * will be dropped for the present. So we must estimate μ_x, V_{xx}, and V_{xy} and we consider the non-fixed terms of $L_{n,k}$, when multiplied by -2 to be:

$$k \log |V| + \sum_{i=1}^{k} (U_i - \mu)' V^{-1} (U_i - \mu) . \tag{3.2.5}$$

Now define

$$\bar{X} = k^{-1} \sum_{i=1}^{k} X_{n[i]}, \quad \bar{Y} = k^{-1} \sum_{i=1}^{k} Y_{n,i}, \quad \bar{U} = (\bar{X},\bar{Y})$$

$$S_{xx} = k^{-1} \sum_{i=1}^{k} (X_{n[i]}-\bar{X})(X_{n[i]}-\bar{X})', \quad S_{xy} = k^{-1} \sum_{i=1}^{k} (X_{n[i]}-\bar{X})(Y_{n,i}-\bar{Y})$$

$$S_{yy} = k^{-1} \sum_{i=1}^{k} (Y_{n,i} - \bar{Y})^2 \tag{3.2.6}$$

$$S = \begin{bmatrix} S_{xx} & S_{xy} \\ S'_{xy} & S_{yy} \end{bmatrix}$$

After dividing by k, (3.2.5) can be written

$$\log |V| + \operatorname{tr} V^{-1} S + (\mu - \bar{U})' V^{-1} (\mu - \bar{U}) \tag{3.2.7}$$

We find that $|\underset{\sim}{V}| = \sigma_y^2 \, |\underset{\sim}{A}^{-1}|$ and

$$\underset{\sim}{V}^{-1} = \begin{bmatrix} \underset{\sim}{A} & -\sigma_y^{-2} \underset{\sim}{A} \underset{\sim}{V}_{xy} \\ -\sigma_y^{-2} \underset{\sim}{V}'_{xy} \underset{\sim}{A} & \sigma_y^{-2} + \sigma_y^{-4} \underset{\sim}{V}'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} \end{bmatrix} \tag{3.2.8}$$

$$\text{tr}\, \underset{\sim}{V}^{-1} \underset{\sim}{S} = \sigma_y^{-2} S_{yy} - 2\sigma_y^{-2} \underset{\sim}{S}'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} + \sigma_y^{-4} S_{yy} \underset{\sim}{V}'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} + \text{tr}\, \underset{\sim}{A} \underset{\sim}{S}_{xx} \tag{3.2.9}$$

$$(\underset{\sim}{\mu}-\underset{\sim}{\overline{U}})' \underset{\sim}{V}^{-1} (\underset{\sim}{\mu}-\underset{\sim}{\overline{U}}) = (\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})' \underset{\sim}{A} (\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}}) - 2\sigma_y^{-2} (\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})' \underset{\sim}{A} \underset{\sim}{V}_{xy} (\mu_y-\overline{Y})$$

$$+ \sigma_y^{-2} (\mu_y-\overline{Y})^2 + \sigma_y^{-4} (\mu_y-\overline{Y}) \underset{\sim}{V}'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} (\mu_y-\overline{Y}) \tag{3.2.10}$$

where the symmetric matrix $\underset{\sim}{A}$ is defined to be

$$(\underset{\sim}{V}_{xx} - \sigma_y^{-2} \underset{\sim}{V}_{xy} \underset{\sim}{V}'_{xy})^{-1}. \tag{3.2.11}$$

Therefore, ignoring fixed terms, (3.2.7) implies

$$L_{n,k} \alpha L = \log|\underset{\sim}{A}^{-1}| - 2\sigma_y^{-2} \underset{\sim}{S}'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} + \sigma_y^{-4} S_{yy} V'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} + \text{tr}\, \underset{\sim}{A} \underset{\sim}{S}_{xx}$$

$$+ (\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})' \underset{\sim}{A} (\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}}) - 2\sigma_y^{-2} (\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})' \underset{\sim}{A} \underset{\sim}{V}_{xy} (\mu_y-\overline{Y})$$

$$+ \sigma_y^{-4} (\mu_y-\overline{Y})^2 \underset{\sim}{V}'_{xy} \underset{\sim}{A} \underset{\sim}{V}_{xy} \tag{3.2.12}$$

$$\partial L/\partial \underset{\sim}{\mu}_x = 2\, \underset{\sim}{A}(\underset{\sim}{\mu}_x - \underset{\sim}{\bar{X}}) - 2\, \sigma_y^{-2} \underset{\sim}{A}\, \underset{\sim}{V}_{xy}(\mu_y - \bar{Y}) \tag{3.2.13}$$

so

$$\underset{\sim}{\mu}_x - \underset{\sim}{\bar{X}} = \sigma_y^{-2}\, \underset{\sim}{V}_{xy}(\mu_y - \bar{Y}). \tag{3.2.14}$$

Let v_{ij} denote an element of $\underset{\sim}{V}_{xx}$, $i,j = 1,\ldots,p$. Also let $\underset{\sim}{\Delta}_{ij} = \partial \underset{\sim}{V}_{xx}/\partial v_{ij}$. $\underset{\sim}{\Delta}_{ij}$ is all zeros except for elements (i,j) and (k,i), which are unity. Using

$$\partial \underset{\sim}{A}/\partial v_{ij} = -\underset{\sim}{A}\, \underset{\sim}{\Delta}_{ij}\, \underset{\sim}{A} \tag{3.2.15}$$

$$\partial \log |\underset{\sim}{A}^{-1}|/\partial v_{ij} = \mathrm{tr}\, \underset{\sim}{A}\, \underset{\sim}{\Delta}_{ij} \tag{3.2.16}$$

we have

$$\begin{aligned}
\partial L/\partial v_{ij} = {} & \mathrm{tr}\, \underset{\sim}{A}\, \underset{\sim}{\Delta}_{ij} + 2\sigma_y^{-2} \underset{\sim}{S}'_{xy} \underset{\sim}{A}\underset{\sim}{\Delta}_{ij}\underset{\sim}{A}\, \underset{\sim}{V}_{xy} - \sigma_y^{-4} S_{yy} \underset{\sim}{V}'_{xy} \underset{\sim}{A}\underset{\sim}{\Delta}_{ij}\underset{\sim}{A}\, \underset{\sim}{V}_{xy} \\
& - \mathrm{tr}\, \underset{\sim}{A}\, \underset{\sim}{\Delta}_{ij}\underset{\sim}{A}\, \underset{\sim}{S}_{xx} - (\underset{\sim}{\mu}_x - \underset{\sim}{\bar{X}})'\underset{\sim}{A}\, \underset{\sim}{\Delta}_{ij}\underset{\sim}{A}\, (\underset{\sim}{\mu}_x - \underset{\sim}{\bar{X}}) \\
& + 2\sigma_y^{-2}(\underset{\sim}{\mu}_x - \underset{\sim}{\bar{X}})'\underset{\sim}{A}\, \underset{\sim}{\Delta}_{ij}\underset{\sim}{A}\, \underset{\sim}{V}_{xy}(\mu_y - \bar{Y}) - \sigma_y^{-4}(\mu_y - \bar{Y})^2 \underset{\sim}{V}'_{xy}\underset{\sim}{A}\underset{\sim}{\Delta}_{ij}\underset{\sim}{A}\underset{\sim}{V}_{xy} \\
= {} & \mathrm{tr}[\{\underset{\sim}{A} + 2\sigma_y^{-2}\underset{\sim}{A}\underset{\sim}{V}_{xy}\underset{\sim}{S}'_{xy}\underset{\sim}{A} - \sigma_y^{-4} S_{yy}\underset{\sim}{A}\, \underset{\sim}{V}_{xy}\, \underset{\sim}{V}'_{xy}\, \underset{\sim}{A} - \underset{\sim}{A}\underset{\sim}{S}_{xx}\underset{\sim}{A}
\end{aligned}$$

$$- \underset{\sim}{A}(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})'\underset{\sim}{A} + 2\sigma_y^{-2}(\mu_y-\overline{Y})\underset{\sim}{A}\underset{\sim}{V}_{xy}(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})'\underset{\sim}{A}$$

$$- \sigma_y^{-4}(\mu_y-\overline{Y})^2\underset{\sim}{A}\ \underset{\sim}{V}_{xy}\ \underset{\sim}{V}'_{xy}\ \underset{\sim}{A}\}\ \underset{\sim}{\Delta}_{ij}]$$

$$= \mathrm{tr}(\underset{\sim}{D}\ \underset{\sim}{\Delta}_{ij}) \tag{3.2.17}$$

Since this trace is zero for every $i,j = 1,\ldots,p$ we conclude by the Lemma of Smith [1978] that $\underset{\sim}{D}$ must be zero. Using (3.2.14) this implies

$$\underset{\sim}{A} + 2\sigma_y^{-2}\ \underset{\sim}{A}\ \underset{\sim}{V}_{xy}\ \underset{\sim}{S}'_{xy}\ \underset{\sim}{A} - \sigma_y^{-4}S_{yy}\underset{\sim}{A}\ \underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\underset{\sim}{A} - \underset{\sim}{A}\ \underset{\sim}{S}_{xx}\underset{\sim}{A} = \underset{\sim}{0}_{p\times p} \tag{3.2.18}$$

Postmultiplying by $\sigma_y^4\ \underset{\sim}{A}^{-1}$ yields

$$\sigma_y^4\ \underset{\sim}{I}_p + \underset{\sim}{A}\ [2\sigma_y^2\ \underset{\sim}{V}_{xy}\ \underset{\sim}{S}'_{xy} - S_{yy}\ \underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy} - \sigma_y^4\ \underset{\sim}{S}_{xx}] = \underset{\sim}{0} \tag{3.2.19}$$

or

$$\sigma_y^4\ [S_{yy}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy} + \sigma_y^4\underset{\sim}{S}_{xx} - 2\ \sigma_y^2\ \underset{\sim}{V}_{xy}\underset{\sim}{S}'_{xy}]^{-1} = \underset{\sim}{A} \tag{3.2.20}$$

$$\sigma_y^{-4}[S_{yy}\ \underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy} + \sigma_y^4\underset{\sim}{S}_{xx} - 2\ \sigma_y^2\underset{\sim}{V}_{xy}\underset{\sim}{S}'_{xy}] = \underset{\sim}{A}^{-1} =$$

$$\underset{\sim}{V}_{xx} - \sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\ . \tag{3.2.21}$$

Therefore,

$$\underset{\sim}{V}_{xx} = \underset{\sim}{S}_{xx} - 2\,\sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{S}'_{xy} + \sigma_y^{-4}(S_{yy} + \sigma_y^2)\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\,. \tag{3.2.22}$$

Next we need to find the derivative of (3.2.12) with respect to $\underset{\sim}{V}_{xy}$.

We have

$$\partial \log|\underset{\sim}{A}^{-1}|/\partial \underset{\sim}{V}_{xy} = (\partial/\partial \underset{\sim}{V}_{xy})\,\mathrm{tr}\,\{\underset{\sim}{B}(-\sigma_y^{-2})\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\}\Big|_{\underset{\sim}{B}=\underset{\sim}{A}}$$

$$= -\sigma_y^{-2}(\partial/\partial \underset{\sim}{V}_{xy})\underset{\sim}{V}'_{xy}\,\underset{\sim}{B}\,\underset{\sim}{V}'_{xy}\Big|_{\underset{\sim}{B}=\underset{\sim}{A}} = -2\sigma_y^{-2}\underset{\sim}{A}\,\underset{\sim}{V}_{xy} \tag{3.2.23}$$

$$(\partial/\partial \underset{\sim}{V}_{xy})\underset{\sim}{S}'_{xy}\underset{\sim}{A}\underset{\sim}{V}_{xy} = (\partial/\partial \underset{\sim}{V}_{xy})\underset{\sim}{S}'_{xy}(-B)(-\sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy})BQ\Big|_{\underset{\sim}{B}=\underset{\sim}{A},\,\underset{\sim}{Q}=\underset{\sim}{V}_{xy}}$$

$$+ (\partial/\partial \underset{\sim}{V}_{xy})\underset{\sim}{S}'_{xy}\underset{\sim}{B}\underset{\sim}{V}_{xy}\Big|_{\underset{\sim}{B}=\underset{\sim}{A}}$$

$$= (\partial/\partial \underset{\sim}{V}_{xy})\sigma_y^{-2}\underset{\sim}{V}'_{xy}\underset{\sim}{B}\underset{\sim}{Q}\underset{\sim}{S}'_{xy}\underset{\sim}{B}\underset{\sim}{V}_{xy}\Big|_{\underset{\sim}{B}=\underset{\sim}{A},\,\underset{\sim}{Q}=\underset{\sim}{V}_{xy}} + \underset{\sim}{A}\,\underset{\sim}{S}_{xy}$$

$$= 2\,\sigma_y^{-2}\underset{\sim}{A}\,\underset{\sim}{V}_{xy}\,\underset{\sim}{S}'_{xy}\,\underset{\sim}{A}\,\underset{\sim}{V}_{xy} + \underset{\sim}{A}\,\underset{\sim}{S}_{xy} \tag{3.2.24}$$

Similarly

$$(\partial/\partial \underset{\sim}{V}_{xy})\underset{\sim}{V}'_{xy}\,\underset{\sim}{A}\,\underset{\sim}{V}_{xy} = 2\,\sigma_y^{-2}\,\underset{\sim}{A}\,\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\,\underset{\sim}{A}\,\underset{\sim}{V}_{xy} + 2\,\underset{\sim}{A}\,\underset{\sim}{V}_{xy} \tag{3.2.25}$$

$$(\partial/\partial \underset{\sim}{V}_{xy})\;\mathrm{tr}\,\underset{\sim}{A}\,\underset{\sim}{S}_{xx} = 2\,\sigma_y^{-2}\underset{\sim}{A}\,\underset{\sim}{S}_{xx}\,\underset{\sim}{A}\,\underset{\sim}{V}_{xy} \tag{3.2.26}$$

$$(\partial/\partial \underset{\sim}{V}_{xy})(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})'\underset{\sim}{A}(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}}) = 2\ \sigma_y^{-2}\underset{\sim}{A}(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})'\underset{\sim}{A}\ \underset{\sim}{V}_{xy}$$

$$= 2\ \sigma_y^{-6}(\mu_y-\overline{Y})^2\underset{\sim}{A}\ \underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\underset{\sim}{A}\ \underset{\sim}{V}_{xy} \qquad (3.2.27)$$

using (3.2.14).

$$(\partial/\partial \underset{\sim}{V}_{xy})(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})'\underset{\sim}{A}\ \underset{\sim}{V}_{xy} = 2\ \sigma_y^{-2}\underset{\sim}{A}\ \underset{\sim}{V}_{xy}(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})'\underset{\sim}{A}\ \underset{\sim}{V}_{xy} + A(\underset{\sim}{\mu}_x-\underset{\sim}{\overline{X}})$$

$$= 2\ \sigma_y^{-4}(\mu_y-\overline{Y})\underset{\sim}{A}\ \underset{\sim}{V}_{xy}\ \underset{\sim}{V}'_{xy}\ \underset{\sim}{A}\ \underset{\sim}{V}_{xy}$$

$$+ \sigma_y^{-2}(\mu_y-\overline{Y})\underset{\sim}{A}\ \underset{\sim}{V}_{xy} \qquad (3.2.28)$$

using (3.2.14). Combining (3.2.23) through (3.2.28) and premultiplying by $\frac{1}{2}\ \sigma_y^6\underset{\sim}{A}^{-1}$,

$$\partial L/\partial \underset{\sim}{V}_{xy}\ \alpha\ (\sigma_y^2 S_{yy} - \sigma_y^4)\underset{\sim}{V}_{xy} - \sigma_y^4\underset{\sim}{S}_{xy} \qquad (3.2.29)$$

$$+ [\sigma_y^4\underset{\sim}{S}_{xx} + S_{yy}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy} - 2\sigma_y^2\underset{\sim}{V}_{xy}\underset{\sim}{S}'_{xy}]\underset{\sim}{A}\ \underset{\sim}{V}_{xy} = \underset{\sim}{0}\,.$$

By (3.2.20) the expression in brackets is $\sigma_y^4\underset{\sim}{A}^{-1}$. Solving for $\underset{\sim}{V}_{xy}$ and adding the * notation to denote the estimators we have

$$\underset{\sim}{V}^*_{xy} = \sigma_y^{*2}S_{yy}^{-1}\ \underset{\sim}{S}_{xy} \qquad (3.2.30)$$

Substituting (3.2.30) into (3.2.14),

$$\underset{\sim}{\mu}_x^* = \underset{\sim}{\overline{X}} + (\mu_y^* - \overline{Y}) S_{yy}^{-1} \underset{\sim}{S}_{xy} . \qquad (3.2.31)$$

Substituting (3.2.30) into (3.2.22),

$$\underset{\sim}{V}_{xx}^* = \underset{\sim}{S}_{xx} + S_{yy}^{-2} (\sigma_y^{*2} - S_{yy}) \underset{\sim}{S}_{xy} \underset{\sim}{S}_{xy}' . \qquad (3.2.32)$$

These modified MLE's are direct extensions of those for the bivariate case given respectively in (2.4.18), (2.4.10), and (2.4.9).

Using (3.2.1),

$$E(\underset{\sim}{X}_{n[i]} | Y_{n,i}) = \underset{\sim}{\mu}_x + \sigma_y^{-2} (Y_{n,i} - \mu_y) \underset{\sim}{V}_{xy} \qquad (3.2.33)$$

which implies

$$E(\underset{\sim}{X} | \underset{\sim}{Y}) = \underset{\sim}{\mu}_x + \sigma_y^{-2} (\overline{Y} - \mu_y) \underset{\sim}{V}_{xy} \qquad (3.2.34)$$

and

$$E(\underset{\sim}{S}_{xy} | \underset{\sim}{Y}) = k^{-1} E[\sum_{i=1}^{k} \underset{\sim}{X}_{n[i]} (Y_{n,i} - \overline{Y}) Y_{n,i}]$$

$$= \sigma_y^{-2} S_{yy} \underset{\sim}{V}_{xy} . \qquad (3.2.35)$$

So from (3.2.31), (3.2.34), and (3.2.35),

$$E(\underset{\sim}{\mu}_x^* | \underset{\sim}{Y}) = \underset{\sim}{\mu}_x + \sigma_y^{-2}(\mu_y^* - \mu_y)\underset{\sim}{V}_{xy} \tag{3.2.36}$$

so $\underset{\sim}{\mu}_x^*$ is unbiased for $\underset{\sim}{\mu}_x$ if μ_y^* is unbiased for μ_y. From (3.2.30) and (3.2.35),

$$E(\underset{\sim}{V}_{xy}^* | \underset{\sim}{Y}) = \sigma_y^{*2}\sigma_y^{-2} \underset{\sim}{V}_{xy} \tag{3.2.37}$$

so $\underset{\sim}{V}_{xy}^*$ is unbiased for $\underset{\sim}{V}_{xy}$ if σ_y^{*2} is unbiased for σ_y^2.

By the invariance property of maximum likelihood estimators we may take as the modified MLE of the vector of correlations between $\underset{\sim}{X}$ and Y to be

$$\underset{\sim}{C}_{xy}^* = \sigma_y^* S_{yy}^{-1} \operatorname{Diag}(\underset{\sim}{V}_{xx}^*)^{-1/2} \underset{\sim}{S}_{xy} \tag{3.2.38}$$

using (3.2.30) and (3.2.32).

3.3. Inference about Independence of $\underset{\sim}{X}$ and Y

3.3.1. Estimation of Multiple Correlation Coefficient

With the setup in (3.2.4), the square of the population multiple correlation coefficient between $\underset{\sim}{X}_i$ and Y_i is

$$\rho^2 = \sigma_y^{-2} \underset{\sim}{V}_{xy}' \underset{\sim}{V}_{xx}^{-1} \underset{\sim}{V}_{xy}. \tag{3.3.1.1}$$

The modified maximum likelihood estimator of ρ^2 is

$$\rho^{2*} = \sigma_y^{*-2} \underset{\sim}{V}_{xy}^{*'} \underset{\sim}{V}_{xx}^{*-1} \underset{\sim}{V}_{xy}^{*} , \tag{3.3.1.2}$$

where σ_y^* is defined in section 2.4 and $\underset{\sim}{V}_{xy}^*$ and $\underset{\sim}{V}_{xx}^*$ are defined respectively in (3.2.30) and (3.2.32). Expanding $\underset{\sim}{V}_{xy}^*$ and $\underset{\sim}{V}_{xx}^*$ this is

$$\rho^{2*} = \sigma_y^{*2} S_{yy}^{-2} \underset{\sim}{S}_{xy}' \, [\underset{\sim}{S}_{xx} + S_{yy}^{-2}(\sigma_y^{*2} - S_{yy}) \underset{\sim}{S}_{xy} \underset{\sim}{S}_{xy}']^{-1} \underset{\sim}{S}_{xy} . \tag{3.3.1.3}$$

Now using for example, Rao [1965, p. 29] if $\underset{\sim}{P}$ is a non-singular matrix , $\underset{\sim}{W}$ a column vector, s a scaler, and $M = \underset{\sim}{W}' \underset{\sim}{P}^{-1} \underset{\sim}{W}$,

$$(\underset{\sim}{P} + s \underset{\sim}{W} \underset{\sim}{W}')^{-1} = \underset{\sim}{P}^{-1} - s(1 + sM)^{-1} \underset{\sim}{P}^{-1} \underset{\sim}{W} \underset{\sim}{W}' \underset{\sim}{P}^{-1} \tag{3.3.1.4}$$

which implies further that

$$\underset{\sim}{W}' (\underset{\sim}{P} + s \underset{\sim}{W} \underset{\sim}{W}')^{-1} \underset{\sim}{W} = M/(1 + s M) . \tag{3.3.1.5}$$

Using this, (3.3.1.3) becomes

$$\rho^{2*} = \sigma_y^{*2} \underset{\sim}{S}_{xy}' \underset{\sim}{S}_{xx}^{-1} \underset{\sim}{S}_{xy} \, \{S_{yy}^2 + (\sigma_y^{*2} - S_{yy}) \underset{\sim}{S}_{xy}' \underset{\sim}{S}_{xx}^{-1} \underset{\sim}{S}_{xy}\}^{-1} \tag{3.3.1.6}$$

or, with $c = S_{yy}\,\sigma_y^{*-2}$,

$$\rho^{2*} = R^{2*}\{c^2 + (1-c)R^{2*}\}^{-1} \tag{3.3.1.7}$$

where R^{2^*} is the ordinary sample squared multiple correlation coefficient defined by

$$R^{2^*} = S_{\sim yy}^{-1}\, S'_{\sim xy}\, S_{\sim xx}^{-1}\, S_{\sim xy}. \tag{3.3.1.8}$$

Note that for a complete sample, c approaches unity since in that case both S_{yy} and σ_y^{*2} are consistent estimates of the same quantity.

3.3.2. Likelihood Ratio Test of Independence

As in section (2.5.1), unrestricted parameter estimates for the (p+1)-variate censored normal distribution are

$$\underset{\sim}{\theta}^* = (\mu_y^*\ ,\ \sigma_y^*\ ;\ \underset{\sim}{\mu}_x^*\ ,\ V_{\sim xx}^*\ ,\ V_{\sim xy}^*). \tag{3.3.2.1}$$

$\underset{\sim}{X}_i$ and Y_i are independent if and only if $V_{\sim xy} = \underset{\sim}{0}$. The estimators over the parameter space restricted by $V_{\sim xy} = \underset{\sim}{0}$ are

$$\tilde{\underset{\sim}{\theta}}^* = (\mu_y^*\ ,\ \sigma_y^*;\ \tilde{\mu}_x^*\ ,\ \tilde{V}_{xx}^*\ ,\ \underset{\sim}{0}) \tag{3.3.2.2}$$

where $\tilde{\underset{\sim}{\mu}}_x^* = \overline{\underset{\sim}{X}}$ and $\tilde{\underset{\sim}{V}}_{xx}^* = \underset{\sim}{S}_{xx}$ as defined by (3.2.6). (3.3.2.3)

Now under H_0: $\underset{\sim}{V}_{xy} = \underset{\sim}{0}$, $\underset{\sim}{A}$ in (3.2.11) evaluated at $\underset{\sim}{\theta} = \tilde{\underset{\sim}{\theta}}^*$ becomes S_{xx}^{-1}. Evaluating $\underset{\sim}{A}$ at $\underset{\sim}{\theta}^*$ using (3.2.30) through (3.2.32) we obtain

$$\underset{\sim}{A}^* = (\underset{\sim}{S}_{xx} - S_{yy}^{-1} \underset{\sim}{S}_{xy} \underset{\sim}{S}_{xy}')^{-1} \tag{3.3.2.4}$$

with the quantities involved defined in (3.2.6). Using these results along with (3.2.12), the maximum log likelihood under H_0 is proportional to

$$p + \log|\underset{\sim}{S}_{xx}| \tag{3.3.2.5}$$

and the unrestricted maximum is

$$\log |\underset{\sim}{A}^{*^{-1}}| - S_{yy}^{-1} \underset{\sim}{S}_{xy}' \underset{\sim}{A}^* \underset{\sim}{S}_{xy} + \mathrm{tr}\, \underset{\sim}{A}^* \underset{\sim}{S}_{xx}. \tag{3.3.2.6}$$

By for example Rao [1965, p. 29],

$$|\underset{\sim}{A}^{*^{-1}}| = |\underset{\sim}{S}_{xx}|(1 - S_{yy}^{-1} \underset{\sim}{S}_{xy}' \underset{\sim}{S}_{xx}^{-1} \underset{\sim}{S}_{xy}). \tag{3.3.2.7}$$

We obtain using (3.3.1.4) that

$$\mathrm{tr}\ (\underset{\sim}{P} + s\ \underset{\sim}{W}\ \underset{\sim}{W}')^{-1}\underset{\sim}{P} = p - s\ M/(1 + s\ M) \tag{3.3.2.8}$$

with the notation as in section 3.3.1 where $\underset{\sim}{P}$ is a non-singular $p \times p$ matrix and $M = \underset{\sim}{W}'\ \underset{\sim}{P}^{-1}\ \underset{\sim}{W}$. So (3.3.2.7), (3.3.1.5), and (3.3.2.8) imply that (3.3.2.6) equals

$$\log |\underset{\sim}{S}'_{xx}| + \log (1 - S_{yy}^{-1}\ \underset{\sim}{S}'_{xy}\ \underset{\sim}{S}_{xx}^{-1}\ \underset{\sim}{S}_{xy}) + p\,. \tag{3.3.2.9}$$

Subtracting the restricted maximum, we obtain as the likelihood ratio criterion

$$\log (1 - R^{*2}),\quad R^{*2} = S_{yy}^{-1}\ \underset{\sim}{S}'_{xy}\ \underset{\sim}{S}_{xx}^{-1}\ \underset{\sim}{S}_{xy} \tag{3.3.2.10}$$

and R^{*2} is the ordinary squared multiple correlation coefficient. Thus the test for independence in multivariate normal complete samples is optimal for testing independence with censored multivariate normal samples in the sense that this test statistic is the likelihood ratio criterion.

3.3.3. Distribution of Test Statistic under Censoring

We take as a test statistic for testing H_0: $\underset{\sim}{V}_{xy} = \underset{\sim}{0}$

$$T^* = \{R^{*2}/p\}/\{(1-R^{*2})/(k-1-p)\}, \tag{3.3.3.1}$$

where R^{*2} is defined in (3.3.1.8) and p is the dimension of $\underset{\sim}{X}_{n[i]}$. Let $F_\Delta(t;\, a,b)$ denote the non-central F d.f. as in section 2.5.2. Let ρ denote the population multiple correlation coefficient of $\underset{\sim}{X}$ and Y in (3.3.1.1) and

$$U^* = k\,\sigma_y^{-2}\,S_{yy} = \sigma_y^{-2}\sum_{i=1}^{k}(Y_{n,i} - \overline{Y})^2 \tag{3.3.3.2}$$

$$\lambda = U^*\,\rho^2/(1-\rho)^2 \tag{3.3.3.3}$$

$$G_{n,k}(u) = P\{U^* \leq u\}, \quad u \geq 0 \tag{3.3.3.4}$$

Then with $F_\rho^*(t;\, a,b)$ defined in (2.5.2.2) we have the following

<u>Theorem 3.3.3.1</u>:

For every $t \geq 0$

$$P\{T^* \leq t \mid U^*\} = F_\lambda(t;\, p,\, k-1-p), \quad t \geq 0. \tag{3.3.3.5}$$

Hence $P\{T^* \leq t\} = F_\rho^*(t;\, p,\, k-1-p)$, $t \geq 0$. <u>Under H_0: $\underset{\sim}{V}_{xy} = \underset{\sim}{0}$</u>, $F^*(t;\, p,\, k-1-p) = F_0(t;\, p,\, k-1-p)$ <u>is the central</u> F <u>d.f. with DF</u> $(p,\, k-1-p)$.

<u>Proof</u>: By (3.3.1.5) with $\underset{\sim}{P} = \underset{\sim}{S}_{xx}$, $\underset{\sim}{W} = \underset{\sim}{S}_{xy}$, $s = S_{yy}^{-1}$,

$$R^{*2}/(1-R^{*2}) = S_{yy}^{-1}\, \underset{\sim}{S}'_{xy}(\underset{\sim}{S}_{xx} - S_{yy}^{-1}\underset{\sim}{S}_{xy}\underset{\sim}{S}'_{xy})^{-1}\underset{\sim}{S}_{xy}. \qquad (3.3.3.6)$$

From (3.2.1), (3.2.4), and (3.2.33), conditional on $Y_{n,i}$, $\underset{\sim}{X}_{n[i]}$ has a p-variate normal distribution with mean

$$\underset{\sim}{\mu}_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)\underset{\sim}{V}_{xy} \qquad (3.3.3.7)$$

and variance matrix

$$\underset{\sim}{V}_{xx} - \sigma_y^{-2}\, \underset{\sim}{V}_{xy}\, \underset{\sim}{V}'_{xy} \qquad (3.3.3.8)$$

and the $X_{n[i]}$, $1 \le i \le k$ are independent. So $\underset{\sim}{S}_{xy}$ is conditionally p-variate normal with mean

$$k^{-1} \sum_{i=1}^{k} \{\mu_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)\underset{\sim}{V}_{xy}\}(Y_{n,i} - \bar{Y}) = \sigma_y^{-2} S_{yy}\underset{\sim}{V}_{xy} \qquad (3.3.3.9)$$

and variance

$$k^{-2} \sum_{i=1}^{k} \{\underset{\sim}{V}_{xx} - \sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}\}(Y_{n,i} - Y)^2 = k^{-1} S_{yy}(\underset{\sim}{V}_{xx} - \sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}). \qquad (3.3.3.10)$$

Let $\underset{\sim}{L}$ be any non-zero p-vector and $Q_i = \underset{\sim}{X}_{n[i]}\underset{\sim}{L}$, $1 \le i \le k$. Conditional on $Y_{n,i}$, Q_i is univariate normal with

$$E(Q_i) = \underset{\sim}{L}'\underset{\sim}{\mu}_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)\underset{\sim}{L}'\, \underset{\sim}{V}_{xy} \qquad (3.3.3.11)$$

$$V(Q_i) = L'(\underset{\sim}{V}_{xx} - \sigma_y^{-2} \underset{\sim}{V}_{xy} \underset{\sim}{V}'_{xy}) \underset{\sim}{L} \tag{3.3.3.12}$$

and $Q_1, \ldots, Q_k$ are independent. Define $\underset{\sim}{Q} = (Q_1, \ldots, Q_k)'$. Conditional on $\underset{\sim}{Y}$, $\underset{\sim}{Q}$ is k-variate normal with moments

$$E(\underset{\sim}{Q}) = \underset{\sim}{L}' \ \underset{\sim}{\mu}_x \underset{\sim}{1} + \sigma_y^{-2} \underset{\sim}{L}' \ \underset{\sim}{V}_{xy} (\underset{\sim}{Y} - \mu_y \underset{\sim}{1}) \tag{3.3.3.13}$$

$$V(\underset{\sim}{Q}) = \underset{\sim}{L}' \ (\underset{\sim}{V}_{xx} - \sigma_y^{-2} \underset{\sim}{V}_{xy} \ \underset{\sim}{V}'_{xy}) \underset{\sim}{L} \ \underset{\sim}{I}_k \ . \tag{3.3.3.14}$$

Let $\underset{\sim}{H} = \underset{\sim}{I}_k - k^{-1} \ \underset{\sim}{1} \ \underset{\sim}{1}'$ so that $\underset{\sim}{H}^2 = \underset{\sim}{H}$. Then $\underset{\sim}{S}_{xx} = k^{-1} \underset{\sim}{X}' \underset{\sim}{H} \underset{\sim}{X}$ and $\underset{\sim}{S}_{xy} = k^{-1} \underset{\sim}{X}' \underset{\sim}{H} \underset{\sim}{Y}$ with $\underset{\sim}{X}_{k \times p}$, $\underset{\sim}{Y}_{k \times 1}$ defined in section 3.2. Note that $\underset{\sim}{Q} = \underset{\sim}{X} \ \underset{\sim}{L}$. Let $\underset{\sim}{E} = k(\underset{\sim}{S}_{xx} - S_{yy}^{-1} \underset{\sim}{S}_{xy} \underset{\sim}{S}'_{xy})$. Then

$$\underset{\sim}{L}' \ \underset{\sim}{E} \ \underset{\sim}{L} = \underset{\sim}{L}' \ \underset{\sim}{X}' [\ \underset{\sim}{H} - k^{-1} S_{yy}^{-1} \ \underset{\sim}{H} \underset{\sim}{Y} \underset{\sim}{Y}' \underset{\sim}{H}] \underset{\sim}{X} \ \underset{\sim}{L} = \underset{\sim}{Q}' \ \underset{\sim}{J} \ \underset{\sim}{Q}. \tag{3.3.3.15}$$

Using (3.3.3.14) and the fact that $\underset{\sim}{J}$ is idempotent,

$$[\underset{\sim}{J} \ V \ (\underset{\sim}{Q})]^2 = [\underset{\sim}{L}' \ (\underset{\sim}{V}_{xx} - \sigma_y^{-2} \underset{\sim}{V}_{xy} \ \underset{\sim}{V}'_{xy}) \underset{\sim}{L}]^2 \ \underset{\sim}{J}. \tag{3.3.3.16}$$

The noncentrality parameter of $\underset{\sim}{L}' \ \underset{\sim}{E} \ \underset{\sim}{L}$ is $\underset{\sim}{\mu}' \ \underset{\sim}{J} \ \underset{\sim}{\mu}$ where $\underset{\sim}{\mu}$ is given by (3.3.3.13). This is seen to be zero. Since rank $(\underset{\sim}{J}) = k-2$ it follows that

$$\underset{\sim}{L}' \ \underset{\sim}{E} \ \underset{\sim}{L} \sim L'(\underset{\sim}{V}_{xx} - \sigma_y^{-2} \ \underset{\sim}{V}_{xy} \ \underset{\sim}{V}'_{xy}) \underset{\sim}{L} \cdot \chi^2_{k-2} \ . \tag{3.3.3.17}$$

Since this is true for every $\underset{\sim}{L}$ we have by Rao [1965, p. 452]

$$k(\underset{\sim}{S}_{xx} - S_{yy}^{-1} \underset{\sim}{S}_{xy} \ \underset{\sim}{S}'_{xy}) \sim W_p(k-2, \ \underset{\sim}{V}_{xx} - \sigma_y^{-2} \ \underset{\sim}{V}_{xy} \ \underset{\sim}{V}'_{xy}) \tag{3.3.3.18}$$

where $W_p(m, \underset{\sim}{\Sigma})$ denotes the central Wishart distribution of dimension p, m degrees of freedom, and matrix parameter $\underset{\sim}{\Sigma}$. Since this distribution does not depend on $\underset{\sim}{Y}$, the unconditional distribution of $\underset{\sim}{E}$ is the central Wishart distribution.

Now $\underset{\sim}{L}' \underset{\sim}{S}_{xy} = k^{-1}\underset{\sim}{L}' \underset{\sim}{X}' \underset{\sim}{H} \underset{\sim}{Y} = k^{-1} \underset{\sim}{Q}'(\underset{\sim}{H} \underset{\sim}{Y})$. Based on (3.3.3.13) and (3.3.3.14), $\underset{\sim}{L}' \underset{\sim}{S}_{xy}$ and $\underset{\sim}{L}' \underset{\sim}{E} \underset{\sim}{L}$ are conditionally independent since

$$\underset{\sim}{Y}'\underset{\sim}{H} \{\underset{\sim}{L}' (\underset{\sim}{V}_{xx} - \sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy})\underset{\sim}{L}\}\underset{\sim}{I}_k (\underset{\sim}{H} - k^{-1}S_{yy}^{-1} \underset{\sim}{H} \underset{\sim}{Y}\underset{\sim}{Y}' \underset{\sim}{H}) = \underset{\sim}{0}.$$

Since the conditional independence holds for any $\underset{\sim}{L}$, $\underset{\sim}{E}$ and $\underset{\sim}{S}_{xy}$ are conditionally independent by Rac [1965, p. 453]. Finally, from (3.3.3.9) and (3.3.3.10), $k^{1/2} \underset{\sim}{S}_{xy}$ is conditionally p-variate normal with mean and variance given respectively by

$$k^{1/2} \sigma_y^{-2}S_{yy}\underset{\sim}{V}_{xy} \quad \text{and} \quad S_{yy}(\underset{\sim}{V}_{xx} - \sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}). \tag{3.3.3.19}$$

We now invoke the theorem on the distribution of Hotelling's T^2 statistic (see Rao [1965, p. 458]). In Rao's notation we let $c = S_{yy}^{-1}$, $\underset{\sim}{d} = \underset{\sim}{S}_{xy}$, $\underset{\sim}{S} = \underset{\sim}{S}_{xx} - S_{yy}^{-1} \underset{\sim}{S}_{xy} \underset{\sim}{S}'_{xy}$, $\underset{\sim}{\Sigma} = \underset{\sim}{V}_{xx} - \sigma_y^{-2}\underset{\sim}{V}_{xy}\underset{\sim}{V}'_{xy}$, $\underset{\sim}{\delta} = k^{1/2} \sigma_y^{-2}S_{yy} \underset{\sim}{V}_{xy}$. Then

$$(k-1-p)\ c\ \underset{\sim}{d}' \underset{\sim}{S}^{-1} \underset{\sim}{d}/p = T^*$$

has conditional d.f. $F_\lambda(t;\ p,\ k-1-p)$ with non-centrality parameter

$$\lambda = c\, \underset{\sim}{\delta}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta} = S_{yy}^{-1}k\sigma_y^{-4}S_{yy}^2\underset{\sim}{V}_{xy}'\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{V}_{xy} = k\, S_{yy}\sigma_y^{-4}\underset{\sim}{V}_{xy}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{V}_{xy} \;.$$

Using (3.3.1.5),

$$\lambda = U^* \rho^2/(1-\rho^2) \tag{3.3.3.20}$$

with ρ^2 and U^* defined respectively in (3.3.1.1) and (3.3.3.2). Equation (3.3.3.5) follows from the fact that this conditional d.f. depends on $\underset{\sim}{Y}$ only through U^*. Finally by (3.3.3.5), $P(T^* \leq t) = E\{P(T^* \leq t|U^*)\} = F_\rho^*(t;\, p,\, k-1-p)$. Under H_0: $\underset{\sim}{V}_{xy} = \underset{\sim}{0}$ this is equal to $F_0(t;\, p,\, k-1-p)$ for all U^* and in that case $F_\rho^* = F_0$. Q.E.D.

From Theorem 3.3.3.1, T^* has the same conditional distribution as the statistic from an uncensored sample and under H_0: $\underset{\sim}{V}_{xy} = \underset{\sim}{0}$ it has the classical F distribution with DF (p, k-1-p). Note that other authors, e.g. Anderson [1958, pp. 89-96], obtain the distribution of the sample multiple correlation coefficient by conditioning on $\underset{\sim}{X}_i$. In our case this would not lead to a normal conditional distribution of Y_i so instead we have had to condition on Y_i. The complicated unconditional non-null distribution of the complete sample multiple correlation coefficient can be found in Anderson [1958, pp. 93-5]. As in section 2.5.2, exact power calculations for this test are difficult to obtain. Here we may also take a one-moment approximation to the distribution of T^* using (3.3.3.20) by taking

$$\lambda = \rho^2/(1-\rho^2)\ E(U^*). \tag{3.3.3.21}$$

$E(U^*)$ can be calculated as in section 2.5.1. The estimated power of the test of size α is

$$1 - F_\lambda(F_\alpha^*;\ p,\ k-1-p) \tag{3.3.3.22}$$

where $F_0(F_\alpha^*;\ p,\ k-1-p) = 1-\alpha$. This provides a much simpler method for higher moment power approximations in the complete sample case ($n=k$) than does the result obtained by conditioning on $\underset{\sim}{X}_i$ (see Anderson [1958, p. 95]) since the non-centrality parameter is a function only of the univariate Y_i , $1 \le i \le n$.

3.4. Inference about Independence of $\underset{\sim}{X}^{(2)}$ and Y given $\underset{\sim}{X}^{(1)}$

3.4.1. Estimation of Partial Correlation Coefficients

We make the following partitions:

$$\underset{\sim}{X}_{n[i]} = (\underset{\sim}{X}^{(1)}_{n[i]1\times p_1}\ \underset{\sim}{X}^{(2)}_{n[i]1\times p_2})\ ,\quad p_1 + p_2 = p$$

$$\underset{\sim}{V}_{xx} = \begin{bmatrix} \underset{\sim}{V}_{11}\ p_1\times p_1 & \underset{\sim}{V}_{12}\ \ p_1\times p_2 \\ \underset{\sim}{V}'_{12} & \underset{\sim}{V}_{22}\ \ p_2\times p_2 \end{bmatrix} \quad \underset{\sim}{V}_{xy} = \begin{bmatrix} \underset{\sim}{V}_{1y}\ p_1\times 1 \\ \underset{\sim}{V}_{2y}\ p_2\times 1 \end{bmatrix} \tag{3.4.1.1}$$

and similarly for $\underset{\sim}{V}^*_{xx}$, $\underset{\sim}{V}^*_{xy}$, $\underset{\sim}{S}_{xx}$, and $\underset{\sim}{S}_{xy}$.

Let $d = S_{yy}^{-2}(\sigma_y^{*^2} - S_{yy})$. (3.4.1.2)

Then from (3.2.32) and (3.2.30),

$$\underset{\sim}{V}_{xx}^{*} = \begin{bmatrix} \underset{\sim}{S}_{11} + d\, \underset{\sim}{S}_{1y}\underset{\sim}{S}_{1y}' & \underset{\sim}{S}_{12} + d\, \underset{\sim}{S}_{1y}\underset{\sim}{S}_{2y}' \\ \underset{\sim}{S}_{12}' + d\, \underset{\sim}{S}_{2y}\underset{\sim}{S}_{1y}' & \underset{\sim}{S}_{22} + d\, \underset{\sim}{S}_{2y}\underset{\sim}{S}_{2y}' \end{bmatrix} \tag{3.4.1.3}$$

$$V_{xy}^{*} = \sigma_y^{*^2} S_{yy}^{-1} \begin{bmatrix} \underset{\sim}{S}_{1y} \\ \underset{\sim}{S}_{2y} \end{bmatrix}. \tag{3.4.1.4}$$

The conditional covariance matrix of $(\underset{\sim}{X}_i^{(2)}, Y_i)$ given $\underset{\sim}{X}_i^{(1)}$ is given by

$$\underset{\sim}{W} = \begin{bmatrix} \underset{\sim}{V}_{22} - \underset{\sim}{V}_{12}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12} & \underset{\sim}{V}_{2y} - \underset{\sim}{V}_{12}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{1y} \\ \underset{\sim}{V}_{2y} - \underset{\sim}{V}_{1y}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12} & \sigma_y^2 - \underset{\sim}{V}_{1y}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{1y} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{W}_{22} & \underset{\sim}{W}_{2y} \\ \underset{\sim}{W}_{2y}' & \underset{\sim}{W}_{yy} \end{bmatrix}. \tag{3.4.1.5}$$

Using (3.4.1.3) and (3.3.1.4),

$$\underset{\sim}{V}_{11}^{*^{-1}} = \underset{\sim}{S}_{11}^{-1} - d(1+d\, \underset{\sim}{S}_{1y}'\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{1y})^{-1}\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{1y}\underset{\sim}{S}_{1y}'\underset{\sim}{S}_{11}^{-1} \tag{3.4.1.6}$$

which along with (3.4.1.3) and (3.4.1.4) provides the modified MLE of $\underset{\sim}{W}$,

$$W^*_{yy} = \sigma^{*2}_y (1 - S^{-1}_{yy} \underset{\sim}{S}'_{1y} \underset{\sim}{S}^{-1}_{11} \underset{\sim}{S}_{1y})/(1+d \underset{\sim}{S}'_{1y} \underset{\sim}{S}^{-1}_{11} \underset{\sim}{S}_{1y}) \qquad (3.4.1.7)$$

$$\underset{\sim}{W}^*_{22} = \underset{\sim}{S}_{22} - \underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{12} + d(1+d\underset{\sim}{S}'_{1y}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})^{-1}\cdot$$

$$(\underset{\sim}{S}_{2y}-\underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})(\underset{\sim}{S}_{2y}-\underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})' \qquad (3.4.1.8)$$

$$\underset{\sim}{W}^*_{2y} = \sigma^{*2}_y S^{-1}_{yy}(\underset{\sim}{S}_{2y}-\underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})/(1+d\underset{\sim}{S}'_{1y}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y}). \qquad (3.4.1.9)$$

Thus the modified MLE of the partial correlation vector of $\underset{\sim}{X}^{(2)}$ and Y given $\underset{\sim}{X}^{(1)}$ is

$$\underset{\sim}{C}^*_{x^{(2)}y|x^{(1)}} = \sigma^*_y S^{-1}_{yy}(1-\underset{\sim}{S}^{-1}_{yy}\underset{\sim}{S}'_{1y}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})^{-1/2}(1+d\underset{\sim}{S}'_{1y}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})^{-1/2}\cdot \qquad (3.4.1.10)$$

$$\mathrm{Diag}(W^*_{22})^{-1/2}(\underset{\sim}{S}_{2y}-\underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y}).$$

The multiple (or "multiple partial") correlation between $\underset{\sim}{X}^{(2)}$ and Y given $\underset{\sim}{X}^{(1)}$ is defined in an analogous manner to the multiple correlation between $\underset{\sim}{X}$ and Y (3.3.1.1) except that the covariance matrix is replaced by $\underset{\sim}{W}$. This reduces to

$$\rho^2_{y2|1} = (\rho^2 - \rho^2_{y1})/(1 - \rho^2_{y1}) \qquad (3.4.1.11)$$

where ρ is the multiple correlation coefficient between $\underset{\sim}{X}$ and Y defined by (3.3.1.1) and ρ_{y1} is the multiple correlation coefficient between $\underset{\sim}{X}^{(1)}$ and Y. Using (3.3.1.7) the modified MLE of ρ_{y1}^2 is

$$\rho_{y1}^{2*} = R_{y1}^{2*} \{c^2 + (1-c)R_{y1}^{2*}\}^{-1} \tag{3.4.1.12}$$

where R_{y1}^{*} is the ordinary sample multiple correlation coefficient between $\underset{\sim}{X}^{(1)}$ and Y ,

$$R_{y1}^{2*} = S_{yy}^{-1} \underset{\sim}{S}_{1y}' \underset{\sim}{S}_{11}^{-1} \underset{\sim}{S}_{1y} . \tag{3.4.1.13}$$

Combining (3.4.1.12) and (3.3.1.7),

$$\rho_{y2|1}^{2*} = (R^{2*} - R_{y1}^{2*})/\{(c-R_{y1}^{2*})[c+(c^{-1}-1)R^{2*}]\} \tag{3.4.1.14}$$

where as before $c = S_{yy} \sigma_y^{*-2}$ is a "correction factor for censoring".

3.4.2. Distribution of Test Statistic under Censoring.

A statistic for testing the independence of $\underset{\sim}{X}^{(2)}$ and Y given $\underset{\sim}{X}^{(1)}$, $H_0: \underset{\sim}{V}_{2y} - \underset{\sim}{V}_{12}' \underset{\sim}{V}_{11}^{-1} \underset{\sim}{V}_{1y} = \underset{\sim}{0}$, or equivalently $H_0: \rho_{y2|1}^2 = 0$ is

$$S^* = \{R_{y2|1}^{*2}/p_2\}/\{(1-R_{y2|1}^{*2})/(k-1-p)\} \tag{3.4.2.1}$$

with the sample squared multiple partial correlation coefficient defined by

$$R^{*2}_{y2|1} = (R^{*2} - R^{*2}_{y1})/(1-R^{*2}_{y1}) \tag{3.4.2.2}$$

and its components defined in (3.3.1.8) and (3.4.1.13). Let ρ^2_{y1} and $\rho^2_{y2|1}$ denote the population correlations defined in section 3.4.1 and

$$P^* = k\,\sigma_y^{-2}\, S_{yy}(1-R^{*2}_{y1})(1-\rho^2_{y1})^{-1} \tag{3.4.2.3}$$

$$\lambda = P^*\, \rho^2_{y2|1}/(1-\rho^2_{y2|1})$$

$$G_{n,k}(u) = P\{P^* \leq u\}, \quad u \geq 0. \tag{3.4.2.4}$$

Then with $F_\Delta(t;a,b)$ denoting the non-central F d.f. as before and $F^*_\rho(t;\, a,b)$ defined in (2.5.2.2), we have the following

Theorem 3.4.2.1:

For every $t \geq 0$ and given P^*

$$P\{S^* \leq t | P^*\} = F_\lambda(t;\, p_2\,,\, k-1-p), \quad t \geq 0. \tag{3.4.2.5}$$

Hence $P\{S^* \leq t\} = F^*_{\rho_{y2|1}}(t;\, p_2\,,\, k-1-p)$, $t \geq 0$. Under H_0: $\rho_{y2|1} = 0$, $F^*_{y2|1}(t;\, p_2\,,\, k-1-p) = F_0(t;\, p_2\,,\, k-1-p)$ is the central F d.f. with DF $(p_2\,,\, k-1-p)$.

Proof: Equation (3.3.1.5) and the matrix partitions given in (3.4.1.5) lead us to the relation

$$R^{*2}_{y2|1}/(1-R^{*2}_{y2|1}) = S^{-1}_{yy}(1-R^{*2}_{y1})^{-1}\ \underset{\sim}{w}'\ \underset{\sim}{A}\ \underset{\sim}{w}\ , \tag{3.4.2.6}$$

where

$$\underset{\sim}{w} = \underset{\sim}{S}_{2y} - \underset{\sim}{S}'_{12}\ \underset{\sim}{S}^{-1}_{11}\ \underset{\sim}{S}_{1y}\ , \text{ and}$$

$$\underset{\sim}{A} = \underset{\sim}{S}_{22} - \underset{\sim}{S}'_{12}\ \underset{\sim}{S}^{-1}_{11}\ \underset{\sim}{S}_{12} -$$

$$S^{-1}_{yy}(1-R^{*2}_{y1})^{-1}(\underset{\sim}{S}_{2y}-\underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})(\underset{\sim}{S}_{2y}-\underset{\sim}{S}'_{12}\underset{\sim}{S}^{-1}_{11}\underset{\sim}{S}_{1y})'. \tag{3.4.2.7}$$

Making partitions as in section 3.4.1 and letting

$$\underset{\sim}{\mu}_x = \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}$$

$$\underset{\sim}{X}^{(1)} = (\underset{\sim}{X}^{(1)'}_{n[1]}\ ,\ldots,\ \underset{\sim}{X}^{(1)'}_{n[k]})',\ \underset{\sim}{X}^{(2)} = (\underset{\sim}{X}^{(2)'}_{n[1]}\ ,\ldots,X^{(2)'}_{n[k]})'$$

$$\underset{\sim}{\overline{X}}^{(1)} = k^{-1}\sum_{i=1}^{k}\underset{\sim}{X}^{(1)}_{n[i]}\ ,\quad \underset{\sim}{\overline{X}}^{(2)} = k^{-1}\sum_{i=1}^{k}\underset{\sim}{X}^{(2)}_{n[i]} \tag{3.4.2.8}$$

$$\underset{\sim}{H} = \underset{\sim}{I}_k - k^{-1}\ \underset{\sim}{1}\ \underset{\sim}{1}'$$

then we have

$$\underset{\sim}{S}_{11} = k^{-1}\underset{\sim}{X}^{(1)'}\,\underset{\sim}{H}\,\underset{\sim}{X}^{(1)}\ ,\ \underset{\sim}{S}_{12} = k^{-1}\,\underset{\sim}{X}^{(1)'}\underset{\sim}{H}\,\underset{\sim}{X}^{(2)}$$

$$\underset{\sim}{S}_{22} = k^{-1}\underset{\sim}{X}^{(2)'}\,\underset{\sim}{H}\,\underset{\sim}{X}^{(2)}\ ,\ \underset{\sim}{S}_{2y} = k^{-1}\,\underset{\sim}{X}^{(2)'}\,\underset{\sim}{H}\,\underset{\sim}{Y}\ . \qquad (3.4.2.9)$$

Now conditional on $Y_{n,i}$, $\underset{\sim}{X}'_{n[i]}$ is p-variate normal with

$$E(X'_{n[i]}|Y_{n,i}) = \begin{pmatrix}\underset{\sim}{\mu}_1\\ \underset{\sim}{\mu}_2\end{pmatrix} + \sigma_y^{-2}(Y_{n,i}-\mu_y)\begin{pmatrix}\underset{\sim}{V}_{1y}\\ \underset{\sim}{V}_{2y}\end{pmatrix}$$

$$V(\underset{\sim}{X}'_{n[i]}|Y_{n,i}) = \begin{pmatrix}\underset{\sim}{V}_{11}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{1y} & \underset{\sim}{V}_{12}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{2y}\\ \underset{\sim}{V}'_{12}-\sigma_y^{-2}\underset{\sim}{V}_{2y}\underset{\sim}{V}'_{1y} & \underset{\sim}{V}_{22}-\sigma_y^{-2}\underset{\sim}{V}_{2y}\underset{\sim}{V}'_{2y}\end{pmatrix}. \qquad (3.4.2.10)$$

So conditional on $Y_{n,i}$ and $\underset{\sim}{X}^{(1)}_{n[i]}$, $\underset{\sim}{X}^{(2)'}_{n[i]}$ is p_2-variate normal with

$$E(\underset{\sim}{X}^{(2)'}_{n[i]}|Y_{n,i}\ ,\underset{\sim}{X}^{(2)}_{n[i]}) = \underset{\sim}{\mu}_2 + \sigma_y^{-2}(Y_{n,i}-\mu_y)\underset{\sim}{V}_{2y} + (\underset{\sim}{V}'_{12}-\sigma_y^{-2}\underset{\sim}{V}_{2y}\underset{\sim}{V}'_{1y})\cdot$$

$$(\underset{\sim}{V}_{11}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{1y})^{-1}[\underset{\sim}{X}^{(1)'}_{n[i]} - (\underset{\sim}{\mu}_1+\sigma_y^{-2}(Y_{n,i}-\mu_y)\underset{\sim}{V}_{1y})] \qquad (3.4.2.11)$$

and

$$V(\underset{\sim}{X}^{(2)'}_{n[i]}|Y_{n,i}\ ,\ \underset{\sim}{X}^{(1)}_{n[i]}) = (\underset{\sim}{V}_{22}-\sigma_y^{-2}\underset{\sim}{V}_{2y}\underset{\sim}{V}'_{2y}) - (\underset{\sim}{V}'_{12}-\sigma_y^{-2}\underset{\sim}{V}_{2y}\underset{\sim}{V}'_{1y})\cdot$$

$$(\underset{\sim}{V}_{11}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{1y})^{-1}(\underset{\sim}{V}_{12}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{2y})$$

which, using (3.3.1.4) is

$$\underset{\sim}{\Sigma} = (\underset{\sim}{V}_{22}-\underset{\sim}{V}'_{12}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12}) - \sigma_y^{-2}(1-\rho_{y1}^2)^{-1}(\underset{\sim}{V}_{2y}-\underset{\sim}{V}'_{12}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{1y})(\underset{\sim}{V}_{2y}-\underset{\sim}{V}'_{12}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{1y})' \tag{3.4.2.12}$$

and $X_{n[i]}^{(2)}$, $1 \leq i \leq k$, are conditionally independent. We have

$$\underset{\sim}{w} = k^{-1} \sum_{i=1}^{k} \underset{\sim}{X}_{n[i]}^{(2)'}\{(Y_{n,i}-\overline{Y}) - (\underset{\sim}{X}_{n[i]}^{(1)}-\overline{\underset{\sim}{X}}^{(1)})\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{1y}\}$$

$$E(\underset{\sim}{w}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y}) = \sigma_y^{-2}S_{yy}(1-R_{y1}^{*2})(1-\rho_{y1}^2)^{-1}(\underset{\sim}{V}_{2y}-\underset{\sim}{V}'_{12}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{1y}) \tag{3.4.2.13}$$

and because of the conditional independence of the $\underset{\sim}{X}_{n[i]}^{(2)}$,

$$V(\underset{\sim}{w}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y}) = k^{-2} \sum_{i=1}^{k} (Y_{n,i}-\overline{Y})-(\underset{\sim}{X}_{n[i]}^{(1)} - \overline{\underset{\sim}{X}}^{(1)})\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{1y}\}^2 \underset{\sim}{\Sigma}$$

$$= k^{-1} S_{yy}(1-R_{y1}^{*2})\underset{\sim}{\Sigma} \ . \tag{3.4.2.14}$$

Let $\underset{\sim}{L}$ be any non-zero p_2-vector and $Q_i = \underset{\sim}{X}_{n[i]}^{(2)} \underset{\sim}{L}$, $1 \leq i \leq k$. Conditional on $Y_{n,i}$, $\underset{\sim}{X}_{n[i]}^{(1)}$, Q_i is univariate normal and

$$E(Q_i|Y_{n,i} , \underset{\sim}{X}_{n[i]}^{(1)}) = \underset{\sim}{L}' \underset{\sim}{\mu}_2 + \sigma_y^{-2}(Y_{n,i}- \mu_y)\underset{\sim}{L}' \underset{\sim}{V}_{2y} +$$

$$[\underset{\sim}{X}_{n[i]}^{(1)} - (\mu_1 + \sigma_y^{-2}(Y_{n,i}- \mu_y)\underset{\sim}{V}_{1y})' (\underset{\sim}{V}_{11}- \sigma_y^{-2} \underset{\sim}{V}_{1y}\underset{\sim}{V}'_{1y})^{-1} \cdot$$

$$(\underset{\sim}{V}_{12} - \sigma_y^{-2} \underset{\sim}{V}_{1y}\underset{\sim}{V}'_{2y})\underset{\sim}{L} \tag{3.4.2.15}$$

$$V(Q_i | Y_{n,i}\ ,\ \underset{\sim}{X}^{(1)}_{n[i]}\ ) = \underset{\sim}{L}'\ \underset{\sim}{\Sigma}\ \underset{\sim}{L} \tag{3.4.2.15}$$

and $Q_1,\ldots,Q_k$ are conditionally independent. Let $\underset{\sim}{Q} = (Q_1,\ldots,Q_k)' = \underset{\sim}{X}^{(2)}\underset{\sim}{L}$. It follows that $\underset{\sim}{Q}$ is k-variate normal with

$$E(\underset{\sim}{Q}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y}) = (\underset{\sim}{L}'\underset{\sim}{\mu}_2)\underset{\sim}{1} + \sigma_y^{-2}(\underset{\sim}{L}'\underset{\sim}{V}_{2y})(\underset{\sim}{Y}-\mu_y\ \underset{\sim}{1}) +$$

$$+ [X^{(1)}-(\underset{\sim}{\mu}_1+\sigma_y^{-2}(Y_{n,i}-\mu_y)\underset{\sim}{V}_{1y})']\ (\underset{\sim}{V}_{11}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{1y})^{-1}(\underset{\sim}{V}_{12}-\sigma_y^{-2}\underset{\sim}{V}_{1y}\underset{\sim}{V}'_{2y})\underset{\sim}{L}$$

$$V(\underset{\sim}{Q}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y}) = \underset{\sim}{L}'\ \underset{\sim}{\Sigma}\underset{\sim}{L}\quad \underset{\sim}{I}_k\ . \tag{3.4.2.16}$$

Let $\underset{\sim}{E} = k\ \underset{\sim}{A}$. Then

$$\underset{\sim}{L}'\underset{\sim}{E}\ \underset{\sim}{L} = k^{-1}\underset{\sim}{L}'\underset{\sim}{X}^{(2)'}[k\ \underset{\sim}{H}-\underset{\sim}{H}\underset{\sim}{X}^{(1)}\underset{\sim}{S}_{11}^{-1}\underset{\sim}{X}^{(1)'}\underset{\sim}{H}-S_{yy}^{-1}(1-R_{y1}^{*2})^{-1}(\underset{\sim}{H}\ \underset{\sim}{Y}\ \underset{\sim}{Y}'\underset{\sim}{H}$$

$$-\ \underset{\sim}{H}\ \underset{\sim}{Y}\ \underset{\sim}{S}'_{1y}\underset{\sim}{S}_{11}^{-1}\underset{\sim}{X}^{(1)'}\underset{\sim}{H}-\underset{\sim}{H}\ \underset{\sim}{X}^{(1)}\underset{\sim}{S}_{11}^{-1}\ \underset{\sim}{S}_{1y}\ \underset{\sim}{Y}'\underset{\sim}{H}$$

$$+\ \underset{\sim}{H}\ \underset{\sim}{X}^{(1)}\underset{\sim}{S}_{11}^{-1}\ \underset{\sim}{S}_{1y}\underset{\sim}{S}'_{1y}\underset{\sim}{S}_{11}^{-1}\ \underset{\sim}{X}^{(1)'}\underset{\sim}{H})]\ \underset{\sim}{X}^{(2)}\ \underset{\sim}{L}$$

$$= (\underset{\sim}{X}^{(2)}\underset{\sim}{L})'\ \underset{\sim}{J}\ (\underset{\sim}{X}^{(2)}\underset{\sim}{L}) = \underset{\sim}{Q}'\ \underset{\sim}{J}\ \underset{\sim}{Q}. \tag{3.4.2.17}$$

It can be verified that $\underset{\sim}{J}$ is idempotent. Therefore,

$$[\underset{\sim}{J}\ V(\underset{\sim}{Q}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y})]^2 = (\underset{\sim}{L}'\ \underset{\sim}{\Sigma}\ \underset{\sim}{L})^2\ \underset{\sim}{J}. \tag{3.4.2.18}$$

The non-centrality parameter, $\underset{\sim}{\mu}'\ \underset{\sim}{J}\ \underset{\sim}{\mu}$ with $\underset{\sim}{\mu} = E(\underset{\sim}{Q}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y})$ can be shown to be zero. Note that rank $(\underset{\sim}{J}) = k-2-p_1$. It follows that conditionally

$$\underset{\sim}{L}'\ \underset{\sim}{E}\ \underset{\sim}{L} \sim \underset{\sim}{L}'\ \underset{\sim}{\Sigma}\ \underset{\sim}{L} \cdot \chi^2_{k-2-p_1}. \tag{3.4.2.19}$$

Since this is true for every $\underset{\sim}{L}$,

$$\underset{\sim}{E} = k\ \dot{\underset{\sim}{A}} \sim W_{p_2}(k-2-p_1\ ,\ \underset{\sim}{\Sigma}) \tag{3.4.2.20}$$

where as before $W_{p_2}(\cdot)$ denotes the central Wishart distribution. We note in passing that since the parameters of this conditional distribution do not depend on $\underset{\sim}{X}^{(1)}$ or $\underset{\sim}{Y}$, $\underset{\sim}{E}$ also has an unconditional Wishart distribution. This result is also a direct consequence of (3.3.3.18) and properties of the Wishart distribution. However, for our purposes we needed to verify that the conditional distribution (given $\underset{\sim}{Y}$ <u>and</u> $\underset{\sim}{X}^{(1)}$ is also a Wishart distribution.

To demonstrate that $\underset{\sim}{w}$ and $\underset{\sim}{E}$ are conditionally independent we see that

$$\underset{\sim}{L}'\underset{\sim}{w} = k^{-1}\underset{\sim}{L}'\underset{\sim}{X}^{(2)\prime}(\underset{\sim}{H}\underset{\sim}{Y}-\underset{\sim}{H}\underset{\sim}{X}^{(1)}\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{1y}) = k^{-1}\underset{\sim}{Q}'(\underset{\sim}{H}\underset{\sim}{Y}-\underset{\sim}{H}\underset{\sim}{X}^{(1)}\underset{\sim}{S}_{11}^{-1}\underset{\sim}{S}_{1y})$$

and

$$(\underset{\sim}{Y}'\underset{\sim}{H} - \underset{\sim}{S}_{1y}'\underset{\sim}{S}_{11}^{-1}\underset{\sim}{X}^{(1)'}\underset{\sim}{H})(\underset{\sim}{L}'\underset{\sim}{\Sigma}\ \underset{\sim}{L})\underset{\sim}{I}_k\ \underset{\sim}{J} = 0. \qquad (3.4.2.21)$$

Then since $\underset{\sim}{L}'\underset{\sim}{w}$ and $\underset{\sim}{L}'\underset{\sim}{E}\ \underset{\sim}{L}$ are conditionally independent for every $\underset{\sim}{L}$, $\underset{\sim}{w}$ and $\underset{\sim}{E}$ are conditionally independent. From (3.4.2.16), $\underset{\sim}{d} = k^{1/2}\ \underset{\sim}{w}$ is conditionally p_2-variate normal with

$$\underset{\sim}{\delta} = E(\underset{\sim}{d}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y}) = k^{1/2}\sigma_y^{-2}S_{yy}(1-R_{y1}^{*2})(1-\rho_{y1}^2)^{-1}(\underset{\sim}{V}_{2y}-\underset{\sim}{V}_{12}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{1y}) \qquad (3.4.2.22)$$

$$V(\underset{\sim}{d}|\underset{\sim}{X}^{(1)},\underset{\sim}{Y}) = S_{yy}(1-R_{y1}^{*2})\underset{\sim}{\Sigma}\ .$$

We again use Rao's Theorem concerning Hotelling's T^2 distribution [1965, p. 458] with $\underset{\sim}{S} = \underset{\sim}{E}$, $c = S_{yy}^{-1}(1-R_{y1}^{*2})^{-1}$, and $\underset{\sim}{\delta}$ and $\underset{\sim}{d}$ given as above. Thus

$$S_{yy}^{-1}(1-R_{y1}^{*2})^{-1}\ \underset{\sim}{w}'\ \underset{\sim}{A}^{-1}\ \underset{\sim}{w}\ (k-1-p)/p_2 = S^*$$

has conditional d.f. $F_\lambda(t;\ p_2,\ k-1-p)$ with $\lambda = c\ \underset{\sim}{\delta}'\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta}$ given by (3.4.2.3) using (3.3.1.5). Equation (3.4.2.5) follows from the fact that the conditional d.f. depends on $\underset{\sim}{Y}$ and $\underset{\sim}{X}^{(1)}$ only through P^*. This implies $P(S^* \leq t) = E\{P(S^* < t|P^*)\} = F^*_{\rho y2|1}(t;\ p_2\ ,\ k-1-p)$. Under H_0: $\rho^2_{y2|1} = 0$, $F^*_{\rho y2|1}$ is equal to $F_0(t;\ p_2\ ,\ k-1-p)$ for all P^* and in that case $F^*_{\rho y2|1} = F_0$. Q.E.D.

From Theorem 3.4.2.1, S^* has the classical null distribution. Power approximation to the S^* test will be more difficult than in the preceeding sections. For a complete sample some simplification may be made, for in that case R^{*2}_{y1} is a consistent estimator of ρ^2_{y1} in (3.4.2.3.).

3.5. Inference about Independence of $\underset{\sim}{X}^{(1)}$ and $(\underset{\sim}{X}^{(2)}, Y)$

3.5.1. Estimation of Canonical Correlations

Let the random variable $\underset{\sim}{X}$ be partitioned as in section 3.4.1. We wish to test H_0: $\underset{\sim}{X}^{(1)}$ is independent of $(\underset{\sim}{X}^{(2)}, Y)$ where Y may be subject to censoring. Let the covariance matrix of $(\underset{\sim}{X}^{(1)'}, \underset{\sim}{X}^{(2)'}, Y)$ be partitioned as

$$\begin{pmatrix} \underset{\sim}{V}_{11} & \underset{\sim}{V}_{12} & \underset{\sim}{V}_{1y} \\ \underset{\sim}{V}'_{12} & \underset{\sim}{V}_{22} & \underset{\sim}{V}_{2y} \\ \underset{\sim}{V}'_{1y} & \underset{\sim}{V}_{2y} & \sigma^2_y \end{pmatrix} \tag{3.5.1}$$

The hypothesis of interest is then H_0: $(\underset{\sim}{V}_{12}\ \underset{\sim}{V}_{1y}) = \underset{\sim}{0}$. Let $\underset{\sim}{T} = (\underset{\sim}{X}^{(2)}, Y)$ and

$$V(\underset{\sim}{T}) = \begin{pmatrix} \underset{\sim}{V}_{22} & \underset{\sim}{V}_{2y} \\ \underset{\sim}{V}_{2y} & \sigma^2_y \end{pmatrix} = \underset{\sim}{\Sigma}_T. \tag{3.5.1.2}$$

Also let $\underset{\sim}{\Sigma}_{1T} = (\underset{\sim}{V}_{12} \ \ \underset{\sim}{V}_{1y})$ and V_{11}^{*} , $\underset{\sim}{\Sigma}_{T}^{*}$, $\underset{\sim}{\Sigma}_{1T}^{*}$ denote modified MLEs of the corresponding parameters as given in section 3.2. Population canonical correlations, the largest of which is the maximum correlation between any linear combination of $\underset{\sim}{T}$ with any linear combination of $\underset{\sim}{X}^{(1)}$, can be estimated by finding eigenvalues of

$$\underset{\sim}{V}_{11}^{*-1} \ \underset{\sim}{\Sigma}_{1T}^{*} \ \underset{\sim}{\Sigma}_{T}^{*-1} \ \underset{\sim}{\Sigma}_{1T}^{*\prime} . \qquad (3.5.1.3)$$

3.5.2. <u>Null Distribution of a Test Statistic</u>

From (3.4.2.10) conditional on $Y_{n,i}$, $X_{n[i]}^{(1)\prime}$ is p_1-variate normal with

$$\begin{aligned} E(X_{n[i]}^{(1)\prime} | Y_{n,i}) &= \underset{\sim}{\mu}_1 + \sigma_y^{-2}(Y_{n,i} - \mu_y)\underset{\sim}{V}_{1y} \\ V(X_{n[i]}^{(1)\prime} | Y_{n,i}) &= \underset{\sim}{V}_{11} - \sigma_y^{-2} \ \underset{\sim}{V}_{1y}\underset{\sim}{V}_{1y}' . \end{aligned} \qquad (3.5.2.1)$$

Under H_0: $\underset{\sim}{\Sigma}_{1T} = \underset{\sim}{0}$, $\underset{\sim}{X}_{n[i]}^{(1)\prime}$, $1 \leq i \leq k$ is (unconditionally) normal with parameters $\underset{\sim}{\mu}_1$ and $\underset{\sim}{V}_{11}$ and $\underset{\sim}{X}_{n[i]}^{(1)}$ is independent of $\underset{\sim}{X}_{n[i]}^{(2)}$ as well as of $Y_{n,i}$. Thus by Anderson [1958 , p. 324] , eigenvalues of

$$\underset{\sim}{S}_{11}^{-1} \ \underset{\sim}{S}_{1T} \ \underset{\sim}{S}_{T}^{-1} \ \underset{\sim}{S}_{1T}' \ , \ \underset{\sim}{S}_{1T} = (\underset{\sim}{S}_{12} \ \underset{\sim}{S}_{1y}) \qquad (3.5.2.2)$$

have the classical null distribution and the test of H_0: $\underset{\sim}{\Sigma}_{1T} = \underset{\sim}{0}$

can be carried out without difficulty. The matrices involved are ordinary sample sums of products matrices. The non-null distribution of the eigenvalues will be exceedingly complex.

CHAPTER IV

STATISTICAL INFERENCE FOR OTHER MULTIVARIATE DISTRIBUTIONS

4.1. Introduction.

Suppose that (X_i, Y_i) has a bivariate d.f. such that Y_i has marginal density $g_{\underset{\sim}{\lambda}}(y)$ with at least the first two moments existing where $\underset{\sim}{\lambda}$ is a $1 \times \ell$ vector. Suppose also that, conditional on Y_i, X_i has a normal distribution with linear regression and homoscedasticity. Define

$$E(Y_i) = \mu_y \,, \quad V(Y_i) = \sigma_y^2 \,, \quad E(X_i) = \mu_x \,, \quad V(X_i) = \sigma_x^2$$
$$E(X_i|Y_i) = \mu_x + \beta(Y_i - \mu_y), \quad V(X_i|Y_i) = \delta^2. \tag{4.1.1}$$

It follows that

$$\mathrm{Cov}(X_i,Y_i) = \beta\sigma_y^2 \,, \quad \mathrm{Corr}(X_i,Y_i) = \rho = \beta\sigma_y/\sigma_x \,,$$
$$\sigma_x^2 = \delta^2 + \beta^2\sigma_y^2. \tag{4.1.2}$$

The information matrix of $\underset{\sim}{\theta} = (\mu_x, \delta, \beta, \underset{\sim}{\lambda})$ and the large sample covariance matrix of $\underset{\sim}{\theta}$ will be calculated. Joint MLE or modified MLE of $\underset{\sim}{\theta}$ will be derived with the MLE of δ, β being independent of $g_{\underset{\sim}{\lambda}}$. These estimators will be used to estimate ρ and σ_x. The likelihood ratio test of H_0: $\beta = 0$ will be displayed. The test statistic does not depend on $g_{\underset{\sim}{\lambda}}$. The properties of the test will be studied for Y_i having an exponential or extreme value distribution.

4.2. Inference when Order Statistics are from any Parent Distribution but Concomitant Variables are Conditionally Normally Distributed

4.2.1. Cramèr-Rao Lower Bounds

The joint density of $Y_{n,1}, \ldots, Y_{n,k}$ is

$$n^{[k]} \prod_{i=1}^{k} g_{\underset{\sim}{\lambda}}(y_i) \; [1 - G_{\underset{\sim}{\lambda}}(y_k)]^{n-k} , \qquad (4.2.1.1)$$

$$y_1 \leq y_2 \leq \cdots \leq y_k$$

where $G_{\underset{\sim}{\lambda}}$ is the d.f. corresponding to the density $g_{\underset{\sim}{\lambda}}$. The conditional density of $X_{n[1]}, \ldots, X_{n[k]}$ given $Y_{n,1}, \ldots, Y_{n,k}$ is, by Lemma 1 of Bhattacharya [1974],

$$\prod_{i=1}^{k} \{\phi_1((x_i - \mu_x - \beta(Y_{n,i} - \mu_y))/\delta)/\delta\} \qquad (4.2.1.2)$$

and $X_{n[1]},\ldots, X_{n[k]}$ are conditionally independent. As before, ϕ_1 denotes the univariate normal density. Denote the log likelihood function of $Y_{n,1},\ldots,Y_{n,k}$ by

$$L^y_{n,k} = \log^{n[k]} + \sum_{i=1}^{k} \log g_{\underset{\sim}{\lambda}}(Y_{n,i}) + (n-k)\log[1-G_{\underset{\sim}{\lambda}}(Y_{n,k})]. \tag{4.2.1.3}$$

The total log likelihood function of $(X_{n[i]}, Y_{n,i}\ , 1 \le i \le k)$ is then, aside from constant terms,

$$L_{n,k} = -k \log\delta - (2\delta^2)^{-1} \sum_{i=1}^{k} [X_{n[i]} - \mu_x - \beta(Y_{n,i} - \mu_y)]^2 + L^y_{n,k}. \tag{4.2.1.4}$$

Let $\underset{\sim}{\alpha}$ denote the $1 \times (\ell - 1)$ parameter vector in $\underset{\sim}{\lambda}$ aside from μ_y. Then letting $\underset{\sim}{T} = \underset{\sim}{X} - \mu_x \underset{\sim}{1}$, $\underset{\sim}{Z} = (\underset{\sim}{Y} - \mu_y \underset{\sim}{1})$,

$$(\partial/\partial\mu_y)L_{n,k} = \underset{\sim}{1}'(\underset{\sim}{T} - \beta\underset{\sim}{Z})/\delta^2 \tag{4.2.1.5}$$

$$(\partial/\partial\delta)L_{n,k} = -k\delta^{-1} + \delta^{-3}(\underset{\sim}{T} - \beta\underset{\sim}{Z})'(\underset{\sim}{T} - \beta\underset{\sim}{Z}) \tag{4.2.1.6}$$

$$(\partial/\partial\beta)L_{n,k} = (\underset{\sim}{T} - \beta\underset{\sim}{Z})'\underset{\sim}{T}/\delta^2 \tag{4.2.1.7}$$

$$(\partial/\partial\mu_y)L_{n,k} = -\beta\, \underset{\sim}{1}'(\underset{\sim}{T} - \beta\underset{\sim}{Z})/\delta^2 + (\partial/\partial\mu_y)L^y_{n,k} \tag{4.2.1.8}$$

$$(\partial/\partial\underset{\sim}{\alpha})L_{n,k} = (\partial/\partial\underset{\sim}{\alpha})L^y_{n,k}\ . \tag{4.2.1.9}$$

Now let the information matrix derived from $L^{y}_{n,k}$ be

$$\underset{\sim}{I}^{y} = \begin{bmatrix} \underset{\sim}{I}_{\mu_y} & \underset{\sim}{I}_{\mu_y \underset{\sim}{\alpha}} \\ \underset{\sim}{I}'_{\mu_y \underset{\sim}{\alpha}} & \underset{\sim}{I}_{\underset{\sim}{\alpha}} \end{bmatrix} \qquad (4.2.1.10)$$

and let $u_{n,i} = E(Y_{n,i} - \mu_y)/\sigma_y$, $w_i = E(Y_{n,i} - \mu_y)^2/\sigma_y^2$ (4.2.1.11)

and $\underset{\sim}{u} = (u_{n,1},\ldots,u_{n,k})'$, $\underset{\sim}{w} = (w_1,\ldots,w_k)'$. Using the fact that $E(\underset{\sim}{T}) = \beta E(\underset{\sim}{Z})$, the information matrix $\underset{\sim}{I}$ for $\underset{\sim}{\theta} = (\mu_x,\delta,\beta,\mu_y,\underset{\sim}{\alpha})$ is:

$$\begin{array}{c|ccccc|} & \mu_x & \delta & \beta & \mu_y & \underset{\sim}{\alpha} \\ \hline \mu_x & k/\delta^2 & 0 & \sigma_y \underset{\sim}{1}'u/\delta^2 & -k\beta/\delta^2 & \underset{\sim}{0} \\ \delta & & 2k/\delta^2 & 0 & 0 & \underset{\sim}{0} \\ \beta & & & \sigma_y^2 \underset{\sim}{1}'\underset{\sim}{w}/\delta^2 & -\beta\sigma_y \underset{\sim}{1}'\underset{\sim}{u}/\delta^2 & \underset{\sim}{0} \\ \mu_y & & & & I_{\mu_y} + k\beta^2/\delta^2 & \underset{\sim}{I}_{\mu_y \underset{\sim}{\alpha}} \\ \underset{\sim}{\alpha} & & & & & \underset{\sim}{I}_{\underset{\sim}{\alpha}} \end{array} \qquad (4.2.1.12)$$

The inverse of the upper left 3×3 submatrix of $\underset{\sim}{I}$ is

$$\begin{bmatrix} \delta^2\, \underset{\sim}{1}'\, \underset{\sim}{w}/kd & 0 & -\delta^2 \underset{\sim}{1}'\underset{\sim}{u}/kd\sigma_y \\ & \delta^2/2k & 0 \\ & & \delta^2/d\sigma_y^2 \end{bmatrix}, \quad d = (\underset{\sim}{1}'\underset{\sim}{w} - (\underset{\sim}{1}'\underset{\sim}{u})^2/k). \qquad (4.2.1.13)$$

This yields $\underset{\sim}{I}^{-1} =$

$$
\begin{array}{ccccc}
\mu_x & \delta & \beta & \mu_y & \underset{\sim}{\alpha}
\end{array}
$$

$$
\begin{bmatrix}
\delta^2 \underset{\sim}{1}'\underset{\sim}{w}/kd+\beta^2 f & 0 & -\delta^2 \underset{\sim}{1}'\underset{\sim}{u}/kd\sigma_y & \beta f & -\beta f \underset{\sim}{I}_{\mu_y \underset{\sim}{\alpha}} \underset{\sim}{I}_{\alpha}^{-1} \\
 & \delta^2/2k & 0 & 0 & 0 \\
 & & \delta^2/d\sigma_y^2 & 0 & \underset{\sim}{0} \\
 & & & & \underset{\sim}{I}^{y^{-1}}
\end{bmatrix}
\qquad (4.2.1.14)
$$

where $f = (I_{\mu_y} - \underset{\sim}{I}_{\mu_y \underset{\sim}{\alpha}} \underset{\sim}{I}_{\alpha}^{-1} \underset{\sim}{I}'_{\mu_y \underset{\sim}{\alpha}})^{-1}$, $\underset{\sim}{I}^y$ is $\ell \times \ell$, and $\underset{\sim}{I}_{\alpha}$ is $(\ell-1) \times (\ell-1)$.

When $g_{\underset{\sim}{\lambda}}$ is a one parameter distribution such as the exponential, $\underset{\sim}{\alpha}$ is not defined. For this case we obtain the large sample covariance matrix

$$
\begin{array}{cccc}
\mu_x & \delta & \beta & \mu_y
\end{array}
$$

$$
\begin{bmatrix}
\delta^2 \underset{\sim}{1}'\underset{\sim}{w}/kd+\beta^2/I_{\mu_y} & 0 & -\delta^2 \underset{\sim}{1}'\underset{\sim}{u}/kd\sigma_y & \beta/I_{\mu_y} \\
 & \delta^2/2k & 0 & 0 \\
 & & \delta^2/d\sigma_y^2 & 0 \\
 & & & I_{\mu_y}^{-1}
\end{bmatrix}
\qquad (4.2.1.15)
$$

and σ_y is a function of μ_y. To place this within the framework of section 2.3, Cramèr-Rao lower bounds on the parameters ρ and σ_x which are derived from β,δ, and σ_y , can be calculated from (4.2.1.14) by standard large-sample methods.

4.2.2. Estimation.

As in previous chapters, we will place * by a parameter to denote an estimator which is derived from maximum likelihood estimators of $\underset{\sim}{\lambda}$ or by another efficient method such as linearization of $L^y_{n,k}$ (Chan [1967]). A ^ over a parameter will denote a true MLE. To estimate μ_x ,δ,β, we set partial derivatives of $L_{n,k}$ in (4.2.1.5) through (4.2.1.7) to zero. Define $\overline{X},\overline{Y}$, S_x^2, S_y^2 and S_{xy} as in (2.4.1). Dropping * and ^ for the moment we find

$$\overline{X} - \mu_x = \beta(\overline{Y} - \mu_y) \quad \text{from (4.2.1.5)}, \tag{4.2.2.1}$$

$$\delta^2 = S_x^2 + \beta^2 S_y^2 - 2\beta S_{xy} \quad \text{from (4.2.1.6)}, \tag{4.2.2.2}$$

$$\text{and} \quad \beta = S_{xy}/S_y^2 , \qquad \text{from (4.2.1.7)}. \tag{4.2.2.3}$$

Note that β is the usual regression estimator. Since it is MLE not depending on $L^y_{n,k}$ we write

$$\hat{\beta} = S_{xy}/S_y^2 . \tag{4.2.2.4}$$

Substituting $\hat{\beta}$ into (4.2.2.2) we find

$$\hat{\delta}^2 = S_x^2 - S_{xy}^2/S_y^2 \ , \tag{4.2.2.5}$$

the usual residual squared error estimator from regression. Using (4.2.2.1), (4.2.1.8) becomes

$$(\partial/\partial\mu_y)L_{n,k} = (\partial/\partial\mu_y)L_{n,k}^y \tag{4.2.2.6}$$

which along with (4.2.1.9) implies that $\underset{\sim}{\lambda}^* = (\mu_y^* \ , \ \underset{\sim}{\alpha}^*)$ are MLE or other efficient estimators based on $L_{n,k}^y$ alone. From (4.2.2.1) we complete the estimation by

$$\mu_x^* = \overline{X} + (S_{xy}/S_y^2)(\mu_y^* - \overline{Y}) \tag{4.2.2.7}$$

which, of course, is the same as (2.4.10). For the derived parameters we obtain, using (4.1.2),

$$\sigma_x^{*2} = S_x^2 + (S_{xy}^2/S_y^4)(\sigma_y^{*2} - S_y^2) \tag{4.2.2.8}$$

$$\rho^* = \sigma_y^* S_{xy}/\sigma_x^* S_y^2$$

as in (2.4.9) and (2.4.7). If μ_y^* , σ_y^* are MLE it follows that σ_x^* and ρ^* are also. This statement can be proved by writing

$\underset{\sim}{\varepsilon} = (\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4,\varepsilon_5) = (\mu_x,\mu_y,\delta^2,\sigma_y^2,\beta)$, $\underset{\sim}{\gamma} = (\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5) =$ $(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$. Thus $\underset{\sim}{\gamma} = (\varepsilon_1,\varepsilon_2,\varepsilon_3 + \varepsilon_4\ \varepsilon_5^2,\varepsilon_4,\varepsilon_4^{1/2}\ \varepsilon_5/(\varepsilon_3+\varepsilon_4\ \varepsilon_5^2)^{1/2})$ and $\underset{\sim}{\varepsilon} = (\gamma_1,\gamma_2,\gamma_3(1-\gamma_5^2),\ \gamma_4,\gamma_5(\gamma_3/\gamma_4)^{1/2})$ by (4.1.2) so the transformation is one-to-one.

Formulas for the expected values of estimators in section 2.4 are also valid here. In particular (2.4.11)and (2.4.13) imply

$$E(\hat{\beta}) = \beta\ ,\quad V(\hat{\beta}|\underset{\sim}{Y}) = \delta^2/k\ S_y^2 \qquad (4.2.2.9)$$

so $$V(\hat{\beta}) = (\delta^2/k)\ E(S_y^{-2}).$$

This exact variance should be compared with the large sample variance from (4.2.1.14),

$$\delta^2/(\underset{\sim}{1}'\underset{\sim}{w} - (\underset{\sim}{1}'\underset{\sim}{u})^2/k)\sigma_y^2 \qquad (4.2.2.10)$$

which for a complete sample equals

$$\delta^2/k\ \sigma_y^2\ . \qquad (4.2.2.11)$$

It can also be shown that

$$E(\hat{\delta}^2) = (k-2)k^{-1}\delta^2. \qquad (4.2.2.12)$$

4.2.3. Likelihood Ratio Test of Independence

By (4.2.2.6), estimation of $\underset{\sim}{\lambda}$ is unaffected by restrictions on β . Hence the log likelihood ratio depends only on the conditional normality of $\underset{\sim}{X}$. Thus the argument in section 2.5.1 applies and the likelihood ratio statistic is a function of the product moment correlation r^* and we use the test statistic T^* in (2.5.1.8). Theorem 2.5.2.1 used only the premise of conditional normality so conditional on S_y^2, T^* has an F d.f. with DF(1,k-2) and non-centrality parameter $\{\rho^2/(1-\rho^2)\}k\ S_y^2/\sigma_y^2$. The unconditional distribution of T^* then depends only on $g_{\underset{\sim}{\lambda}}(y)$ and under H_0: $\beta = 0$, $\rho = 0$ and T^* has an unconditional F d.f.

4.3. Example: Order Statistics from Extreme Value Parent Distribution

Let $Y_1,\ldots,Y_n$ denote random variables from the extreme value type I distribution with d.f.

$$G_{\underset{\sim}{\lambda}}(y) = \exp[\,-e^{-(y-\xi)/\theta}], \quad -\infty < y < \infty \ , \ \theta > 0 . \tag{4.3.1}$$

This distribution is a useful model for life-testing. An interpretation of its parameters is given by the moments,

$$\begin{aligned} E(Y_i) &= \xi + \gamma\,\theta \\ V(Y_i) &= \theta^2\pi^2/6 \end{aligned} \tag{4.3.2}$$

where γ is Euler's constant (.57721...). The location parameter ξ can be efficiently estimated by linear combinations of order statistics while the scale parameter θ cannot; see Johnson and Kotz [1970, p. 284] for a summary of efficiencies of various estimators. Harter and Moore [1968] showed that while the MLE of ξ and θ have considerable bias, the MLE of ξ based on a censored sample is still comparable to the best linear unbiased estimator in terms of mean squared error and the MLE of θ is superior to the best linear unbiased estimator.

Let

$$\Gamma_w(a) = \int_0^w u^{a-1} e^{-u} \, du \tag{4.3.3}$$

denote the incomplete gamma function and $\Gamma'_w(a) = (\partial\Gamma_w(t)/\partial t)\Big|_{t=a}$, etc. Let

$$\gamma'(a) = \Gamma'_\infty(a) - \Gamma'_{-\log p}(a), \quad \gamma''(a) = \Gamma''_\infty(a) - \Gamma''_{-\log p}(a), \tag{4.3.4}$$

where $p = k/n$ is the proportion of order statistics observed. $\gamma'(a)$, $\gamma''(a)$, etc. are related to incomplete polygamma functions. Specializing the results of Harter and Moore [1968] to the case of censoring on the right, we find as elements of the information matrix,

$$\lim_{n\to\infty} (-\theta^2/n)\, E(\partial^2 L^y_{n,k}/\partial\xi^2) = p - p \log p = a$$

$$\lim (-\theta^2/n) E(\partial^2 L^y_{n,k}/\partial\xi\partial\theta) =$$

$$-[\gamma'(2) + p\, q \log p + (1-p)^{-1} p\, q \log^2 p] = b \tag{4.3.5}$$

$$\lim_{n\to\infty} (-\theta^2/n) E(\partial^2 L^y_{n,k}/\partial\theta^2) = \gamma''(2) + 2[\gamma'(2) - \gamma'(1)] - p$$

$$+ p\, q \log p\, \{2 + q + (1-p)^{-1} q \log p\} = c$$

where $q = \log(-\log p)$ and $L^y_{n,k}$ is the log likelihood function based on the order statistics as defined in (4.2.1.3). So the "asymptotic information matrix" of ξ, θ is

$$n/\theta^2 \begin{bmatrix} a & b \\ b & c \end{bmatrix}. \tag{4.3.6}$$

To draw inferences from the induced order statistics $\underset{\sim}{X}$ we need to identify the transformed parameters

$$\mu_y = E(Y_i) = \xi + \gamma\theta$$

$$\sigma_y = [V(Y_i)]^{1/2} = \theta\pi/\sqrt{6} \tag{4.3.7}$$

since the parameters associated with $\underset{\sim}{X}$ are defined by conditional moments of X. The large sample covariance matrix of the MLE of μ_y and σ_y ($\hat{\xi} + \gamma\hat{\theta}$ and $\hat{\theta}\pi/\sqrt{6}$ respectively) is

$$(\sigma_y^2/n) \begin{bmatrix} (6/\pi^2)(\gamma^2 a + c - 2\gamma b) & (\sqrt{6}/\pi)(\gamma a - b) \\ (\sqrt{6}/\pi)(\gamma a - b) & a \end{bmatrix} \tag{4.3.8}$$

with corresponding information matrix

$$\underset{\sim}{I}^y = (n/\sigma_y^2) \begin{bmatrix} (\pi^2/6)a & (\pi/\sqrt{6})(b-\gamma a) \\ (\pi/\sqrt{6})(b-\gamma a) & \gamma^2 a + c - 2\gamma b \end{bmatrix} \tag{4.3.9}$$

Using this, the asymptotic joint lower bounds for variances of estimators of $(\mu_x, \delta, \beta, \mu_y, \sigma_y)$ can be readily obtained using (4.2.1.14). For example we find that the lower bound for the variance of μ_x^* from (4.2.1.14) is

$$\delta^2 \underset{\sim}{1}'\underset{\sim}{w}/kd + (\sigma_y^2/n)(6/\pi^2)\beta^2(\gamma^2 a + c - 2\gamma b)/(ac - b^2 - 4\gamma ab) \tag{4.3.10}$$

where $\underset{\sim}{w}$ and d in (4.2.1.13) can be computed using the first two moments of extreme value order statistics (see Johnson and Kotz [1970, p. 279]).

To study the test of independence of X_i and Y_i, or H_0: $\beta = 0$ we need to study the distribution of

$$U^* = \sigma_y^{-2} \sum_{i=1}^{k} (Y_{n,i} - \overline{Y})^2 = (6/\pi^2)\theta^{-2} \sum_{i=1}^{k} (Y_{n,i} - \overline{Y})^2$$

$$= (6/\pi^2) \sum_{i=1}^{k} (Z_{n,i} - \overline{Z})^2 \tag{4.3.11}$$

where $Z_{n,i}$, $1 \leq i \leq k$ are order statistics from the standard extreme value distribution ($\xi = 0$, $\theta = 1$). $E(U^*)$ is most easily approximated by simulation. For this purpose, 2000 $\underset{\sim}{Z}$ vectors were generated for each n,k combination. The average values of U^* are as follows:

(4.3.12)

$\underline{n}$	$\underline{k}$	$\underline{E(U^*)}$
20	10	1.518
100	20	1.250
1000	50	1.419

To estimate the unconditional power of the test, the conditional power was computed for each simulated value of U^* and each value of ρ by using the noncentral F distribution. These 2000 powers were then averaged to estimate the unconditional distribution of U^*. For the one moment power approximations, the corresponding noncentral F probabilities were evaluated using $E(U^*)$ in (4.3.12). The results are in Table 4.1.

TABLE 4.1

ESTIMATED POWER OF 5% TEST FOR CENSORED DATA FROM

THE EXTREME VALUE DISTRIBUTION

n	k	ρ	Power	Power from One Moment Approximation
20	10	.1	.0514	.0514
		.3	.0636	.0636
		.5	.0967	.0965
		.9	.5686	.6079
100	20	.1	.0513	.0513
		.3	.0628	.0628
		.5	.0940	.0939
		.9	.5677	.5889
1000	50	.1	.0516	.0516
		.3	.0656	.0656
		.5	.1035	.1035
		.9	.6581	.6738

The power approximation is apparently satisfactory.

4.4. Example: Order Statistics from Exponential Parent Distribution

Let Y_i have density function

$$g_{\underset{\sim}{\lambda}}(y) = \sigma_y^{-1} \exp(-y/\sigma_y) \tag{4.4.1}$$

so that $\mu_y \equiv \sigma_y$. This implies

$$L^y_{n,k} = \log n^{[k]} - k \log \sigma_y - \sigma_y^{-1} \sum_{i=1}^{k} Y_{n,i} - (n-k)\sigma_y^{-1} Y_{n,k} \quad . \tag{4.4.2}$$

$$\hat{\sigma}_y = k^{-1}\left(\sum_{i=1}^{k} Y_{n,i} + (n-k)Y_{n,k}\right) = \overline{Y} + k^{-1}(n-k)Y_{n,k} \; . \tag{4.4.3}$$

The MLE $\hat{\sigma}_y$ has several desirable properties, among them being that it is the best linear unbiased estimator for σ_y (Sarhan and Greenberg [1962, p. 354]). Its variance is σ_y^2/k, which is also the reciprocal of

$$I_{\sigma_y} = I_{\mu_y} = k\,\sigma_y^{-2}. \tag{4.4.4}$$

Using $\hat{\sigma}_y$ we can estimate μ_x , σ_x , and ρ with (4.2.2.7) and (4.2.2.8) and using (2.4.15), (2.4.16), and (2.4.17), we find that

$$\begin{aligned} \mu_x^* &= \overline{X} - (n-k)S_{xy}\, Y_{n,k}/k\, S_y^2 , \quad E(\mu_x^*) = \mu_x \\ V(\mu_x^*) &= \sigma_x^2/k + k^{-3}(n-k)^2\delta^2\, E(Y_{n,k}^2/S_y^2) \end{aligned} \tag{4.4.5}$$

(note that the last term may be large for large amount of censoring) and

$$E(\sigma_x^{*2}) = \sigma_x^2(1 + \rho^2/k) + k^{-1}\delta^2[E(\hat{\sigma}_y^2/S_y^2) - 2]$$

since

$$E(\hat{\sigma}_y^2) = k^{-1}(k+1)\sigma_y^2 \ . \tag{4.4.6}$$

From Sarhan and Greenberg [1962, p. 343] , the order statistics have the following moments, with $Z_{n,i} = Y_{n,i}/\sigma_y$:

$$\begin{aligned} a_{n,i} &= E(Z_{n,i}) = \sum_{\ell=1}^{i} (n-\ell+1)^{-1} \\ b_{n,i} &= V(Z_{n,i}) = \mathrm{Cov}(Z_{n,i}, Z_{n,j}) = \sum_{\ell=1}^{i} (n-\ell+1)^{-2}, \ i \leq j \ . \end{aligned} \tag{4.4.7}$$

So

$$\underset{\sim}{1}'\underset{\sim}{u} = \sum_{i=1}^{k} E(Z_{n,i}-1) = -(n-k)a_{n,k} \quad \text{using} \quad (4.4.3) \tag{4.4.8}$$

$$\underset{\sim}{1}'\underset{\sim}{w} = \sum_{i=1}^{k} E(Z_{n,i}-1)^2 = \sum_{i=1}^{k} (b_{n,i} + a_{n,i}^2 - 2\, a_{n,i} + 1)$$

and

$$\sum_{i=1}^{k} b_{n,i} = \sum_{i=1}^{k} \sum_{\ell=1}^{i} (n-\ell+1)^{-1} = \sum_{i=1}^{k} (k-i+1)/(n-i+1)^2$$

$$= \sum_{i=1}^{k} \{(n-i+1) + [(k-i+1)-(n-i+1)]\}/(n-i+1)^2$$

$$= a_{n,k} - (n-k)b_{n,k} \tag{4.4.9}$$

$$\sum_{i=1}^{k} a_{n,i}^2 = \sum_{i=1}^{k} (k-i+1)(n-i+1)^{-2} + 2\sum_{i>\ell}\sum (k-i+1)(n-i+1)^{-1}(n-\ell+1)^{-1}$$

$$= a_{n,k} - (n-k)b_{n,k} \tag{4.4.10}$$

$$+ 2\sum_{i>\ell}\sum \{(n-i+1) + [(k-i+1)-(n-i+1)]\}(n-i+1)^{-1}(n-\ell+1)^{-1}.$$

The rightmost term is

$$2\sum_{i>\ell}\sum (n-\ell+1)^{-1} - 2(n-k)\sum_{i>\ell}\sum (n-i+1)^{-1}(n-\ell+1)^{-1}$$

whose first term is

$$2\sum_{\ell=1}^{k} (k-\ell)(n-\ell+1)^{-1} = 2\sum_{\ell=1}^{k} \{(n-\ell+1) + [(k-\ell)-(n-\ell+1)](n-\ell+1)^{-1}$$

and second term is

$$(n-k)(b_{n,k} - a_{n,k}^2)$$

so (4.4.10) equals

$$2k - [1 + 2(n-k)]a_{n,k} - (n-k)a_{n,k}^2 \tag{4.4.11}$$

implying

$$\underset{\sim}{1}'\underset{\sim}{w} = k-(n-k)E(Z_{n,k}^2) \tag{4.4.12}$$

$$d = \underset{\sim}{1}'\underset{\sim}{w} - (\underset{\sim}{1}'\underset{\sim}{u})^2/k = k-(n-k)b_{n,k}-nk^{-1}(n-k)a_{n,k}^2. \tag{4.4.13}$$

With these quantities, the large sample covariance matrix in (4.2.1.15) can readily be evaluated.

An indication of the power of the r^* correlation test is given by the expected value of

$$U^* = \sum_{i=1}^{k} Z_{n,i}^2 - k^{-1} \left(\sum_{i=1}^{k} Z_{n,i}\right)^2. \tag{4.4.14}$$

By combining (4.4.9) and (4.4.11), the expected value of the first term is

$$E\left(\sum_{i=1}^{k} Z_{n,i}^2\right) = 2k - 2(n-k)a_{n,k}-(n-k)(b_{n,k}+a_{n,k}^2). \tag{4.4.15}$$

Using the fact that $V[\Sigma Z_{n,i} + (n-k)Z_{n,k}] = k$ we obtain

$$E\left\{\left(\sum_{i=1}^{k} Z_{n,i}\right)^2\right\} = k + (E\,\Sigma\, Z_{n,i})^2 - (n-k)^2 b_{n,k}$$

$$-2(n-k)\ \mathrm{Cov}(\Sigma\, Z_{n,i}\ ,\ Z_{n,k}). \tag{4.4.16}$$

The second term is obtained from (4.3.15). For the covariance term,

$$\mathrm{Cov}(\Sigma Z_{n,i}, Z_{n,k}) = E(\sum_{i=1}^{k} Z_{n,i}\, Z_{n,k}) - E(Z_{n,k})\,\Sigma\, E(Z_{n,i})$$

$$= \sum_{i=1}^{k} \mathrm{Cov}(Z_{n,i}\,,\, Z_{n,k}) = \sum_{i=1}^{k} b_{n,i} \quad (\text{since } i \leq k)$$

$$= a_{n,k} - (n-k)b_{n,k} \text{ by (4.4.9).} \qquad (4.4.17)$$

Thus we obtain

$$E\{(\Sigma Z_{n,i})^2\} = k^2 + k - 2(n-k)(k+1)a_{n,k} + (n-k)^2(b_{n,k} + a_{n,k}^2) \qquad (4.4.18)$$

and using (4.4.15),

$$E(U^*) = k-1 + k^{-1}(n-k)[2\, a_{n,k} - n(b_{n,k} + a_{n,k}^2)]. \qquad (4.4.19)$$

For large amounts of censoring, $E(U^*)$ is small in comparison with the normal distribution (for $n = 1000$, $k = 50$, $E(U^*) = 6.79$ for normal distribution, 0.011 for exponential). Now since $a_{n,k} \doteq - \log(1-p)$, $b_{n,k} \doteq \frac{1}{n-k} - \frac{1}{n} = p(1-p)/n$ where $p = k/n$,

$$(k-1)^{-1}E(U^*) \doteq 1 - p^{-2}(1-p)\log^2(1-p) \qquad (4.4.20)$$

for large n and fixed p. This quantity does not exceed .5 until $p > .95$. Since U^* relates to the non-centrality parameter of the conditional F distribution, power characteristics of the test of

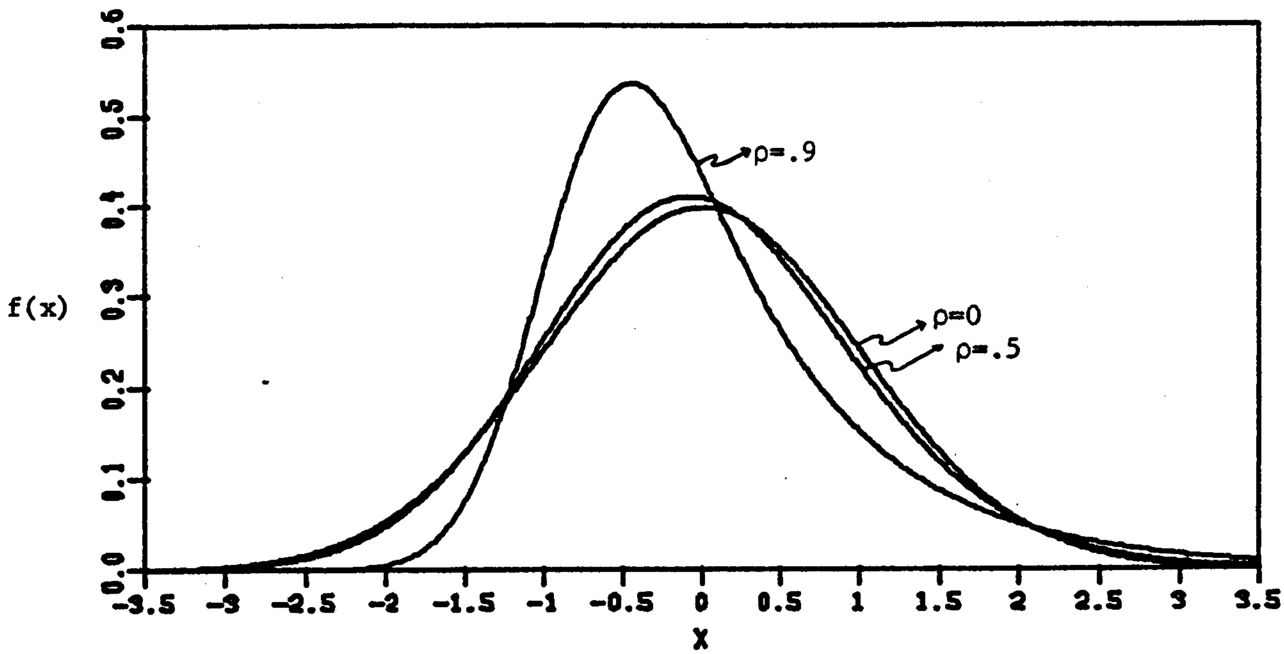
H_0: $\beta = 0$ will be inferior with this exponential model with any substantial amount of censoring. Where the exponential distribution is most dense (in its left tail), there is little variation in the smallest order statistics. Either the linear regression model or the assumption of homoscedastic variances for this setup is unreasonable. Thus while the extreme value distribution model performed quite well, this theory may be of limited utility in some non-normal models.

It is interesting to consider the marginal density of X in this model where Y has the exponential distribution. Taking $\sigma_x = \sigma_y = 1$ and $\mu_x = 0$ for simplicity, X has the standard normal density when $\rho = 0$, and when $\rho \neq 0$ it has density

$$f(x) = \int_0^\infty [2\pi(1-\rho^2)]^{-1/2} \exp\{-[x-\rho(y-1)]^2/2(1-\rho^2)\}\exp(-y)dy$$
$$= |\rho|^{-1}\exp\{(1-3\rho^2-2\rho x)/2\rho^2\}\Phi((x+\rho-(1-\rho^2)/\rho)/\sqrt{1-\rho^2}) \tag{4.4.21}$$

where $\Phi(\cdot)$ is the standard normal d.f. This density closely resembles the normal density for moderate ρ and becomes more skewed as ρ becomes large in absolute value; however its first two moments remain unchanged. A graph of $f(x)$ is shown in Figure 4.1.

Figure 4.1. Graph of Marginal Density of X_i when Y_i has an Exponential Distribution



A model in which the marginal distribution of X should depend on ρ is one in which both X and Y are reflections of another variable with a strength of association ρ.

The problems with this exponential model can be remedied if instead of linear regression holding the regression on X_i is log-linear, i.e.,

$$E(X_i|Y_i) = \mu_x + \beta(T_i - \mu_y) \tag{4.4.22}$$

where $T_i = \log Y_i$. As far as estimation of parameters associated with X_i is concerned, we could use the extreme value distribution for Y_i as discussed earlier. This is a generalization of the log linear model since $-T_i$ has the extreme value distribution with parameters $\xi = -\log \sigma_y$, $\theta = 1$. However, the present discussion will be for the log linear model with Y_i having an exponential distribution.

Using properties of the extreme value distribution,

$$\begin{aligned} E(T_i) &= \log \sigma_y - \gamma \\ V(T_i) &= \pi^2/6 \end{aligned} \tag{4.4.23}$$

and since $2k\,\hat{\sigma}_y/\sigma_y \sim \chi^2_{2k}$ (Epstein and Sobel [1953] in Sarhan and Greenberg [1962, p. 362]) it follows that the moment generating function of $\log \hat{\sigma}_y$ is

$$m(t) = (\sigma_y/k)^t \, \Gamma(t+k)/\Gamma(k). \tag{4.4.24}$$

Therefore,

$$\begin{aligned} E(\log \hat{\sigma}_y) &= m'(0) = \log \sigma_y + \psi(k) - \log k \\ V(\log \hat{\sigma}_y) &= m''(0) - m'(0)^2 = \psi'(k) \end{aligned} \tag{4.4.25}$$

where

$$\psi(k) = \Gamma'(k)/\Gamma(k) = -\gamma + \sum_{i=1}^{k-1} i^{-1}$$

and (4.4.26)

$$\psi'(k) = \pi^2/6 - \sum_{i=1}^{k-1} i^{-2} \doteq k^{-1}.$$

Note that $\lim_{k\to\infty} \psi(k) - \log k = \lim_{k\to\infty} \psi'(k) = 0$. An unbiased estimator of $\mu_T = E(T_i)$ is

$$\mu_T^* = \log \hat{\sigma}_y + \log k - \sum_{i=1}^{k-1} i^{-1} \tag{4.4.27}$$

with variance $\psi'(k)$. From Johnson and Kotz [1970, p. 285], if $k = n$, $\log \hat{\sigma}_y$ is the MLE of $E(T_i)$ and for any k,n, $V(\mu_T^*)$ is equal to the variance of the MLE of $E(T_i)$ for a complete sample of size k.

From μ_T^* an unbiased estimator of μ_x can be produced as well as simple estimators of ρ and σ_x since $V(T_i)$ is a known constant. All parameters are estimated in the manner described previously with Y replaced by $T = \log Y$, e.g.

$$\begin{aligned}
\hat{\beta} &= S_{XT}/S_T^2 \\
\mu_x^* &= \bar{X} + (S_{XT}/S_T^2)(\bar{T} - \mu_T^*) \\
\sigma_x^{*2} &= S_x^2 + (S_{XT}^2/S_T^4)(\pi^2/6 - S_T^2) \\
\rho^* &= (\pi/\sqrt{6})S_{XT}/\sigma_x^* S_T^2 .
\end{aligned} \qquad (4.4.28)$$

The power function of the test of $\beta = 0$ for this model is different from that of section 4.3 since here we are sampling from the opposite tail of the distribution of Y_i. The average values of U^* here are:

n	k	$E(U^*)$
20	10	5.653
100	20	12.244
1000	50	30.545

The power estimates calculated as in table 4.1 are shown in table 4.2.

TABLE 4.2

ESTIMATED POWER OF 5% TEST FOR CENSORED DATA

UNDER THE LOG-LINEAR MODEL

n	k	ρ	Power	Power from One Moment Approximation
20	10	.1	.0560	.0560
		.3	.1102	.1100
		.5	.2486	.2569
		.9	.9079	.9957
100	20	.1	.0629	.0628
		.3	.1804	.1808
		.5	.4516	.4810
		.9	.9953	.9999
1000	50	.1	.0847	.0846
		.3	.3910	.3988
		.5	.8289	.8783
		.9	1.0000	1.0000

The one moment approximation is again satisfactory. By comparison with table 2.3 the power characteristics of this model, when it holds, are greatly superior to those of the bivariate normal model under censoring.

CHAPTER V

DISTRIBUTION FREE TESTS OF INDEPENDENCE BASED ON INDUCED ORDER STATISTICS

5.1. Introduction.

In this chapter we will derive locally most powerful rank tests for independence of X_i and Y_i when (X_i, Y_i) are from a general bivariate distribution. The tests will be explored in more detail for the case when X_i is conditionally normally distributed or when linear regression-type alternatives are appropriate. Empirical power studies will be made for Y_i having a normal or extreme value distribution. An asymptotic study will then be made of relative efficiencies of rank tests of independence with regards to local regression alternatives. Finally, an omnibus nonparametric test of independence due to Hoeffding [1948] will be studied for the case of censored bivariate samples.

5.2. Locally Most Powerful Rank Test of Independence

Suppose that Y_i has density $g(y)$ with d.f. $G(y)$ and $f_\theta(x|y)$ is the conditional density of X_i given $Y_i = y$. We assume that θ is a parameter such that X_i and Y_i are independent if $\theta = 0$ and that θ is not a parameter of $g(y)$. When $\theta = 0$ we write $f_0(x|y) = f_0(x)$. With $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ defined as in section 1.4, the joint density of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ is

$$n^{[k]}\{\prod_{i=1}^{k} f_\theta(x_i|y_i)g(y_i)\}\ [1-G(y_k)]^{n-k},\ y_1 \leq y_2 \leq \cdots \leq y_k\ . \qquad (5.2.1)$$

The density when X_i and Y_i are independent is

$$n^{[k]}\{\prod_{i=1}^{k} f_0(x_i)g(y_i)\}\ [1-G(y_k)]^{n-k}\ ,\quad y_1 \leq y_2 \leq \cdots \leq y_k\ . \qquad (5.2.2)$$

The likelihood ratio then is

$$L_{n,k} = \prod_{i=1}^{k}\{f_\theta(X_{n[i]}|Y_{n,i})/(f_0(X_{n[i]})\}. \qquad (5.2.3)$$

It can easily be seen that

$$Q_{n,k} = (\partial L_{n,k}/\partial\theta)\Big|_{\theta=0} = \sum_{i=1}^{k}[\partial \log\ f_\theta(X_{n[i]}|Y_{n,i})/\partial\theta]\Big|_{\theta=0}. \qquad (5.2.4)$$

Let $S_{n,i}$ be the rank of $Y_{n,i}$ among $Y_1,\ldots,Y_n$ $(S_{n,i} = i)$, and $\underset{\sim}{S}_n = (S_{n,1},\ldots,S_{n,k})$. For continuous distributions, the probability of a tie is zero and ranks are well defined with probability 1.

For the induced order statistics, the rank of $X_{n[i]}$ among $X_1,\ldots,X_n$ is unknown. Define $R_{k,i}$ to be the rank of $X_{n[i]}$ among $X_{n[1]},\ldots,X_{n[k]}$ and $R_k = (R_{k,1},\ldots,R_{k,k})$. Then following, for example, Cox and Hinkley [1974, p. 188], the locally most powerful rank test (ℓ.m.p.r.t.) for testing H_0: $\theta = 0$ using only $\underset{\sim}{R}_k$ and $\underset{\sim}{S}_n$ is based on the statistic

$$T_{n,k} = E(Q_{n,k} | \underset{\sim}{R}_k , \underset{\sim}{S}_n , \theta = 0)$$

$$= \sum_{i=1}^{k} E\{[\partial \log f_\theta(X_{n[i]} | Y_{n,i})/\partial\theta |_{\theta=0}] | R_{k,i} , S_{n,i} , \theta = 0\}$$

$$= \sum_{i=1}^{k} c_{n,k}(i, R_{k,i}). \tag{5.2.5}$$

Computation of $c_{n,k}(i,R_{k,i})$ and therefore $T_{n,k}$ is facilitated by using the fact that under H_0 , $\underset{\sim}{X}$ represents a complete random sample of size k from $f_0(x)$. In that case $X_{n[R_1]},\ldots,X_{n[R_k]}$ are the k order statistics of the k i.i.d.r.v. $X_{n[i]}$, $1 \leq i \leq k$ where we have abbreviated $R_{k,i}$ by R_i. Now under H_0: X and Y independent, $\underset{\sim}{R}_k$ can take on any of the k! permutations of the integers 1,...,k. This fact gives rise to an exact permutation test for H_0:θ=0.

Hoeffding [1951] showed that, under H_0 ,

$$E(T_{n,k}) = k^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} c_{nk} (i,j), \text{ and}$$

$$V(T_{n,k}) = (k-1)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} d^2_{nk} (i,j), \tag{5.2.6}$$

where

$$d_{n,k}(i,j) = c_{n,k}(i,j) - k^{-1} \sum_{g=1}^{k} c_{n,k}(g,j) - k^{-1} \sum_{h=1}^{k} c_{n,k}(i,h) +$$

$$k^{-2} \sum_{g=1}^{k} \sum_{h=1}^{k} c_{n,k}(g,h). \qquad (5.2.7)$$

He also showed that under suitable conditions on the $c_{n,k}(i,j)$, $V(T_{n,k})^{-1/2}\ (T_{n,k} - E\ T_{n,k})$ converges in distribution to the standard normal distribution under H_0 as $k \to \infty$. [Note that we have introduced the dependence of $c_{n,k}(i,j)$ on the additional parameter n which was not the case in Hoeffding [1951] , however, the same theory still applies.] Motoo (1957) weakened the conditions on the $c_{n,k}(i,j)$ to the following:

$$\lim_{k\to\infty} \sum_{|d_{n,k}(i,j)/d_{n,k}| > \varepsilon} (d_{n,k}(i,j)^2/d_{n,k}^2) = 0 \qquad (5.2.8)$$

where $d_{n,k}^2 = k^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} d_{n,k}(i,j)^2$.

For many densities f_θ , the quantity $c_{n,k}(i,j)$ will factor into a product $a_{n,k}(i)\ b_{n,k}(j)$, so that the resulting test statistic is a linear rank statistic, and the resulting test is easy to apply. There are distributions, however, for which such factorization is not possible. An example is the Gumbel bivariate exponential distribution.

Assume now the more restricted model $f_\theta(x|y) = [2\pi(1-\theta^2)]^{-1/2} \cdot \exp\{-(x-\theta y)^2/2(1-\theta^2)\}$ which does yield linear rank statistics. Location and scale parameters can be ignored because the rank tests to be considered are location and scale invariant. For this model,

$$Q_{n,k} = \sum_{i=1}^{k} X_{n[i]} Y_{n,i} . \tag{5.2.9}$$

Define $u_{k,i} = E(X_{k,i})$ where $X_{k,i}$ is the i^{th} order statistic out of k from $f_0(x) = (2\pi)^{-1/2} \exp(-x^2/2)$. Also define $t_{n,i} = E(Y_{n,i})$ where as before $Y_{n,i}$ is the i^{th} order statistic out of n from the parent distribution $g(y)$. Then

$$\begin{aligned} E(X_{n[i]}|R_i\ , \ \theta = 0) &= u_{k,R_i} , \\ E(Y_{n,i}|S_{n,i}) &= t_{n,i} , \end{aligned} \tag{5.2.10}$$

and since $X_{n[i]}$ and $Y_{n,i}$ are independent when $\theta = 0$,

$$T_{n,k} = E(Q_{n,k}|\underset{\sim}{R}_k\ \underset{\sim}{S}_n\ , \ \theta = 0) = \sum_{i=1}^{k} u_{k,R_i}\ t_{n,i} . \tag{5.2.11}$$

This statistic is useful not only for the conditional normal model but also for general monotonic association alternatives. As in Ghosh and Sen [1971] , one can also define "mixed rank statistics" which are asymptotically efficient. These are

$$M^x_{n,k} = \sum_{i=1}^{k} X_{n[i]}\, t_{n,i}$$

or (5.2.12)

$$M^y_{n,k} = \sum_{i=1}^{k} u_{k,R_i}\, Y_{n,i} .$$

The statistic $M^y_{n,k}$ might be preferable since the scores $u_{n,i}$ are more dependent on $g(y)$ for large amounts of censoring.

Since under H_0 all permutations of R_i , $1 \leq i \leq k$ among the integers $1,\ldots,k$ are equally likely, $T_{n,k}$ provides a non-parametric test of independence. By Theorem II.3.1.c of Hájek and Šidák [1967, p. 61] , under H_0 ,

$$E(T_{n,k}) = k\ \bar{u}\ \bar{t}$$

and (5.2.13)

$$V(T_{n,k}) = \sigma_u^2 \sum_{i=1}^{k} (t_{n,i} - \bar{t})^2 ,$$

where

$$\bar{u} = k^{-1} \sum_{i=1}^{k} u_{k,i} , \quad \bar{t} = k^{-1} \sum_{i=1}^{k} t_{n,i} , \text{ and}$$

(5.2.14)

$$\sigma_u^2 = (k-1)^{-1} \sum_{i=1}^{k} (u_{k,i} - \bar{u})^2 .$$

These results hold true for all $t_{n,i}$ and $u_{k,i}$ whether or not these scores are defined as in (5.2.7). Assume that $t_{n,1},\ldots,t_{n,k}$ are any real constants such that $\sum_{i=1}^{k}(t_{n,i}-\bar{t})^2>0$ and that there exists a square integrable function $\tau(u)$ such that

$$\lim_{k\to\infty}\int_0^1 [u_{k,1+[v\,k]} - \tau(v)]^2 dv = 0. \tag{5.2.15}$$

Hájek and Šidák [1967] have shown (Theorem V.1.6.a) that under H_0, $T_{n,k}$ is asymptotically normal for $k\to\infty$ with mean and variance given in (5.2.13). The asymptotic variance of $T_{n,k}$ can also be used:

$$\lim_{k\to\infty} V(T_{n,k}) = [\sum_{i=1}^{k}(t_{n,i}-\bar{t})^2]\int_0^1 [\tau(u) - \bar{\tau}]^2\,du\ , \tag{5.2.16}$$

where $\bar{\tau} = \int_0^1 \tau(u)du$. So for large k, critical values from the standard normal distribution can be compared to

$$Z_{n,k} = (k-1)^{1/2}\, r_{n,k} \tag{5.2.17}$$

where $r_{n,k} = (\sum_{i=1}^{k} u_{k,R_i}\, t_{n,i} - k\bar{u}\,\bar{t})/[\sum_{i=1}^{k}(u_{k,i}-\bar{u})^2 \sum_{i=1}^{k}(t_{n,i}-\bar{t})^2]^{1/2}$

(5.2.18)

can be used as a rank measure of correlation. Another approximation is to treat

$$(k-2)^{1/2}\, r_{n,k}/(1-r_{n,k}^{2})^{1/2} \tag{5.2.19}$$

as an approximate t random variable with $k-2$ D.F. This approximation is accurate for small sample sizes in many situations (e.g. logistic or normal generating distributions).

The paper by Shirahata [1975] also deals with the ℓ.m.p.r.t. under censoring but in the case where X and Y are both time-censored, i.e. X_i is observed if it is one of the n_1 smallest X_i and Y_i is observed if it is one of the n_2 smallest Y_i.

5.3. Empirical Power Studies

In this section we study power properties of the rank tests of independence discussed in section 5.2, assuming the conditional normality model for X given Y and for g(y) being the normal or extreme value distribution. We write ρ instead of θ to be consistent with discusssions of parametric tests. The power is estimated here by generating 1000 values of $r_{n,k}$ for each n,k,ρ, estimating the 95^{th} percentile of $|r_{n,k}|$ for $\rho = 0$ and computing the proportion of $|r_{n,k}|$ greater than this critical value when $\rho \neq 0$. Blom scores defined in (1.4.9) are employed for $u_{k,i}$ and for $t_{n,i}$ when g(y) is the normal density. The empirical critical values are:

n	k	Empirical Critical Value	Critical Value Estimated from (5.2.19)
20	10	.620	.632
100	20	.457	.444
1000	50	.274	.279

When X and Y are truly from a bivariate normal distribution and normal scores are used for $u_{k,i}$ and $t_{n,i}$ the following power estimates result:

TABLE 5.1

ESTIMATED POWER OF 5% NORMAL SCORES TEST FOR

CENSORED DATA FROM THE BIVARIATE NORMAL DISTRIBUTION

n	k	ρ	Estimated Power
20	10	.1	.061
		.3	.086
		.5	.142
		.9	.743
100	20	.1	.045
		.3	.082
		.5	.139
		.9	.895
1000	50	.1	.064
		.3	.148
		.5	.319
		.9	.993

Except for $n = 100$ $k = 20$ $\rho = .9$ these results compare quite favorably to the power of the best parametric test for the bivariate normal case as found in table 2.3. The normal scores test statistic has a null distribution that is independent of $f_\theta(x|y)$ and $g(y)$ in the sense that it has a permutation distribution depending only on the assumption that all permutations of $\underset{\sim}{R}_k$ are equally likely. Therefore, the normal scores test may be preferred over the parametric test which assumes that either $f_0(x)$ or $g(y)$ is normal before its null distribution is known. The normal scores test may also be appropriate in the presence of non-linear regression or heteroscedasticity.

In order to examine power properties of the normal scores test when the scores are not optimal we now assume the extreme value/normal model of section 4.3 but continue to use normal scores for $t_{n,i}$ and $u_{k,i}$. The power estimates are found in table 5.2.

TABLE 5.2

ESTIMATED POWER OF 5% NORMAL SCORES TEST FOR CENSORED DATA FROM THE EXTREME VALUE/NORMAL DISTRIBUTION

n	k	ρ	Estimated Power
20	10	.1	.062
		.3	.071
		.5	.102
		.9	.499
100	20	.1	.043
		.3	.056
		.5	.064
		.9	.513
1000	50	.1	.054
		.3	.089
		.5	.106
		.9	.652

By comparing those results with table 4.1, we see that very little power is lost over the parametric test even though only the scores for $\underset{\sim}{X}$ are optimal. The normal scores test of independence may perform well for many bivariate distributions.

5.4. Asymptotic Relative Efficiencies of Linear Rank Tests

Assume that optimum scores for $\underset{\sim}{R}_k^x$ are generated asymptotically from a square integrable function $\tau_1(u)$ as in (5.2.15) and likewise that scores for $\underset{\sim}{R}_k^y$ are generated by $\psi_1(u)$ satisfying

$$\lim_{n\to\infty} \int_0^p [t_{n,1+[un]} - \psi_1(u)]^2 du = 0 \tag{5.4.1}$$

where $k = pn$ and $p \varepsilon (0,1]$ is fixed. We assume that the linear rank statistic from $\tau_1(u)$ and $\psi_1(u)$ is locally most powerful for testing H_0: $\theta = 0$ and is asymptotically efficient relative to the best parametric test in the Pitman (local alternatives) sense. Now let $\tau_2(u)$ and $\psi_2(u)$ be other corresponding score functions which are not necessarily optimal. Define

$$\rho_x^2 = \frac{\{\int_0^1 \tau_1(u)\tau_2(u)\,du - \bar{\tau}_1\bar{\tau}_2\}^2}{\int_0^1[\tau_1(u)-\bar{\tau}_1]^2 du \int_0^1[\tau_2(u)-\bar{\tau}_2]^2 du} \tag{5.4.2}$$

and

$$\rho_y^2 = \frac{\{p^{-1}\int_0^p \psi_1(u)\psi_2(u)\,du - \bar{\psi}_1\bar{\psi}_2\}^2}{p^{-1}\int_0^p[\psi_1(u)-\bar{\psi}_1]^2 du \; p^{-1}\int_0^p[\psi_2(u)-\bar{\psi}_2]^2 du} \tag{5.4.3}$$

where $\bar{\tau}_i = \int_0^1 \tau_i(u)\,du$, $\quad i=1,2$

and (5.4.4)

$$\overline{\psi}_i = p^{-1} \int_0^p \psi_i(u)\,du \ , \quad i=1,2 \ .$$

Note that ρ_x^2 and ρ_y^2 can be interpreted as squared correlation coefficients between corresponding "correct" and "incorrect" scores. We now argue heuristically based on the results of Ghosh and Sen [1971] and Basu [1967] that the asymptotic relative efficiency of the linear rank test based on (τ_2,ψ_2) to that based on (τ_1,ψ_1) is given by

$$e_{12} = \rho_x^2 \rho_y^2 . \tag{5.4.5}$$

For example, choosing correct X-scores $(\tau_1(u) \equiv \tau_2(u))$ yields $\rho_x^2 = 1$. However, choosing Wilcoxon scores for X $(\tau_2(u) = u)$ when the optimal scores are normal scores $(\tau_1(u) = \Phi^{-1}(u))$ yields $\rho_x^2 = 3/\pi$. If this same mistake is made in choosing scores for Y $(\psi_1(u) = \Phi^{-1}(u),\ \ \psi_2(u) = u)$, the following component quantities for ρ_y^2 are:

$$\overline{\psi}_1 = p^{-1} \int_0^p \Phi^{-1}(u)\,du = p^{-1} \int_{-\infty}^a (2\pi)^{-1/2} y\ e^{-y^2/2} dy = -\phi(a)/p \tag{5.4.6}$$

where $a = \Phi^{-1}(p)$, $\phi(u) = (2\pi)^{-1/2}\ e^{-u^2/2}$,

$$\overline{\psi}_2 = p/2 \ ,$$

$$p^{-1} \int_0^p u\, \Phi^{-1}(u)du = -\phi(a) + \Phi(2^{1/2}\ a)/2p\pi^{1/2} \ , \tag{5.4.7}$$

$$p^{-1} \int_0^p [\psi_1(u) - \overline{\psi}_1]^2 du = 1 - a\ p^{-1}\phi(a) - p^{-2}\phi(a)^2 \ , \tag{5.4.8}$$

and

$$p^{-1} \int_0^p [\psi_2(u) - \overline{\psi}_2]^2 du = p^2/12. \tag{5.4.9}$$

So in this case $\rho_y^2 = 3\{\pi^{-1/2}\Phi(2^{1/2}a)p^{-1} - \phi(a)\}^2/\{p^2 - ap\phi(a) - \phi(a)^2\}$.

(5.4.10)

This squared correlation of Y - score functions has respective values .636, .902, and .955 $(3/\pi)$ when $p = .00057$, .5, and 1 respectively. Thus for example very little efficiency is lost in using the Spearman correlation coefficient (which incorporates Wilcoxon scores) instead of the normal scores correlation for censored bivariate normal data.

5.5. Hoeffding's General Test of Independence as applied to Censored Bivariate Samples

Suppose that two random variables have a d.f. H(x,y) with continuous joint and marginal density functions. Let

$$\Delta = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D^2(x,y)\,dH(x,y) \tag{5.5.1}$$

where

$$D(x,y) = H(x,y) - H(x,\infty)H(\infty,y). \tag{5.5.2}$$

The "D" test based on a random sample $\{(X_i,Y_i),\ 1 \leq i \leq n\}$ was devised by Hoeffding in 1948. It depends only on the ranks of the $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ vectors and tests H_0: $\Delta = 0$, being consistent against <u>all</u> alternatives to independence. The test can also be applied to a censored bivariate sample $\{(X_{n[i]}, Y_{n,i}),\ 1 \leq i \leq n\}$ since under the null hypothesis all permutations of the ranks $\underset{\sim}{R}_k$ are equally likely (fixing the ranks $\underset{\sim}{S}_n$) due to $\{X_{n[i]},\ 1 \leq i \leq k\}$ representing a random sample of size k under H_0.

Hoeffding derived an unbiased estimator of Δ based on a U-statistic. Alternatively Δ can be estimated with the sample bivariate empirical distribution:

$$\hat{\Delta} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D_n^2(x,y)\,d\,H_n(x,y) \tag{5.5.3}$$

where

$$\begin{aligned} H_n(x,y) &= n^{-1}\sum_{i=1}^{n} I\{X_i \leq x,\ Y_i \leq y\}, \\ D_n(x,y) &= H_n(x,y) - H_n(x,\infty)H_n(\infty,y), \end{aligned} \tag{5.5.4}$$

and $I\{A\}$ is the indicator function for the set A. Blum, Kiefer, and Rosenblatt [1961] developed another test of this type in the spirit of Kolmogorov-Smirnov tests. Their test statistic is

$$B_n = \sup_{x,y} D_n(x,y). \tag{5.5.5}$$

For a censored bivariate sample the following information is available:

$$F^0_{n,k}(x) = k^{-1} \sum_{i=1}^{n} I\{X_i \leq x\}c_i \tag{5.5.6}$$

$$= k^{-1} \sum_{i=1}^{k} I\,\{X_{n[i]} \leq x\},$$

$$G_{n,k}(y) = n^{-1} \sum_{i=1}^{k} I\,\{Y_{n,i} \leq y\}, \tag{5.5.7}$$

$$H_{n,k}(x,y) = n^{-1} \sum_{i=1}^{k} I\{X_{n[i]} \leq x\}I\{Y_{n,i} \leq y\}, \tag{5.5.8}$$

and

$$D_{n,k}(x,y) = H_{n,k}(x,y) - F^0_{n,k}(x)G_{n,k}(y) \tag{5.5.9}$$

where

$$c_i = I\{\text{rank of } Y_i \text{ among } Y_1,\ldots,Y_n \text{ is } \leq k\}. \tag{5.5.10}$$

When $Y_{n,k} \geq y$ the following simplifications hold:

$$G_{n,k}(y) = n^{-1} \sum_{i=1}^{n} I\{Y_i \leq y\} \tag{5.5.11}$$

and

$$H_{n,k}(x,y) = n^{-1} \sum_{i=1}^{n} I\{X_i \leq x\} I\{Y_i \leq y\}. \tag{5.5.12}$$

These sample distribution functions have the following properties:

$$E[G_{n,k}(y) \mid Y_{n,k} \geq y] = G(y) \ , \tag{5.5.13}$$

and

$$E[H_{n,k}(x,y) \mid Y_{n,k} \geq y] = H(x,y) \ . \tag{5.5.14}$$

and if θ denotes the dependence parameter as in section 5.2,

$$E[F^0_{n,k}(x) \mid \theta = 0] = F(x) \tag{5.5.15}$$

and

$$E[D_{n,k}(x,y)\,|\,Y_{n,k} \geq y, \;\; \theta = 0] = 0 \tag{5.5.16}$$

since $E[c_i] = k/n.$ (5.5.17)

Here $F(x)$ and $G(y)$ are respectively the marginal d.f. of X_i and Y_i. A statistic for testing independence is

$$\hat{\Delta}_{n,k} = \int_{-\infty}^{\infty} \int_{-\infty}^{Y_{n,k}} D^2_{n,k}(x,y)\, d\, H_{n,k}(x,y). \tag{5.5.18}$$

An alternative test statistic analogous to (5.5.5) is $B_{n,k} = \sup\limits_{\substack{-\infty < x < \infty \\ -\infty < y \leq Y_{n,k}}} D_{n,k}.$

These statistics should perhaps be studied in more detail. However, Hoeffding's D test can be carried out without difficulty. It would be interesting to see how such an omnibus test of independence performs with respect to a simple test directed at simple alternatives. To this end, 1000 censored bivariate normal random vectors were generated for different n,k, and ρ. The empirical upper 95th percentile when $\rho = 0$ was used as a critical value since the null distribution of D is tabulated only for very small sample sizes. The resulting critical values are:

n	k	$D_{.95}$
20	10	.00503
100	20	.00218
1000	50	.00067

Power is estimated by computing the proportion of observed D statistics greater than $D_{.95}$. The results are in table 5.3.

TABLE 5.3

ESTIMATED POWER OF 5% HOEFFDING D TEST FOR

CENSORED BIVARIATE NORMAL DISTRIBUTION

n	k	ρ	Estimated Power
20	10	.1	.066
		.3	.084
		.5	.117
		.9	.573
100	20	.1	.045
		.3	.075
		.5	.118
		.9	.782
1000	50	.1	.055
		.3	.113
		.5	.236
		.9	.970

By comparison with table 2.3 the power is reduced from that of the likelihood ratio test but is not devastated. The power reduction is the smallest for the case $n = 1000$ $k = 50$ in which a high proportion of censoring is present.

CHAPTER VI

SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

6.1. Summary.

Many tools of classical multivariate theory have been extended for use in analyzing censored multivariate samples consisting of order statistics of one variate and induced order statistics of other variates. Maximum likelihood and simplified estimators have been developed for means, variances, covariances, correlation coefficients, and multiple, partial and canonical correlation coefficients. Statistics for testing total or partial independence of random vectors have also been derived. Linear rank tests for independence in the bivariate case have also been proposed and Hoeffding's general test of independence in this case has been studied.

Properties of estimators have been studied by deriving Cramèr-Rao lower bounds for dispersion and by simulations. Power properties of some of the proposed test procedures have been studied by deriving conditional power functions and by making empirical power calculations. The rank tests for independence were shown to have many advantages over the parametric tests since the power characteristics of the two types of tests are similar while the rank tests make fewer assumptions.

6.2. Suggestions for Further Research

There are several open areas remaining in this research. For one, the unconditional distribution of test statistics such as (2.5.1.8) could be derived if the distribution of the sample variance calculated from a censored normal sample could be obtained. This may not turn out to be a tractable problem. Also, it would be advantageous to prove formally that the partial correlation test in section 3.4.2 is actually the likelihood ratio test [the author could find no formal proof of this even for the complete sample case].

Another question worth considering is how much information regarding tests of independence is lost when only induced order statistics are observed and not the entire vector of concomitant data. This question could be studied by comparing the power characteristics of the rank test proposed in section 5.2 with those of the test proposed by, for example, Brown *et al*. [1974].

BIBLIOGRAPHY

Anderson, T. W. 1958. An Introduction to Multivariate Statistical Analysis. New York: Wiley.

Basu, A. P. 1967. On the large sample properties of a generalized Wilcoxon-Mann-Whitney statistic. Annals of Mathematical Statistics 38: 905-15.

Bhattacharya, P. K. 1974. Convergence of sample paths of normalized sums of induced order statistics. Annals of Statistics 2: 1034-9.

Blum, J. R., Kiefer, J., and Rosenblatt, M. 1961. Distribution free tests of independence based on the sample distribution function. Annals of Mathematical Statistics 32: 485-98.

Brown, B. W., Hollander, M., and Korwar, R. M. 1974. Nonparametric tests of independence for censored data with applications to heart transplant studies. In Reliability and Biometry, Editors F. Proschan and R. J. Serfling, Philadelphia: S.I.A.M. (pp. 327-54).

Chan, L. K. 1967. Remark on the linearized maximum likelihood estimate [for type II double censoring]. Annals of Mathematical Statistics 38: 1876-81.

Cohen, A. C. 1965. Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. Technometrics 7: 579-88.

Cox, D. R. 1972. Regression models and life-tables. Journal of the Royal Statistical Society Series B 34: 187-202.

Cox, D. R., and Hinkley, D. V. 1974. Theoretical Statistics. London: Chapman and Hall Ltd.

David, H. A. 1970. Order Statistics. New York: Wiley.

David, H. A., and Galambos, J. 1974. The asymptotic theory of concomitants of order statistics. Journal of Applied Probability 11: 762-70.

David, H. A., and Moeschberger, M. L. 1978. The Theory of Competing Risk. London: Griffin.

Des Raj 1953. On estimating the parameters of bivariate normal population from doubly and singly linearly truncated samples. Sankhya 12: 277-90.

Gehan, E. A. 1965. A generalized Wilcoxon test for comparing arbitrarily singly censored samples. Biometrika 52: 203-23.

Ghosh, M., and Sen, P. K. 1971. On a class of rank order tests for regression with partially informed stochastic predictors. Annals of Mathematical Statistics 42: 650-61.

Glasser, M. 1965. Regression analysis with dependent variable censored. Biometrics 21: 300-7.

Glasser, M. 1967. Exponential survival with covariance. Journal of the Americal Statistical Association 62: 561-8.

Gupta, A. K. 1952. Estimation of the mean and standard deviation of a normal population from a censored sample. Biometrika 39: 260-73.

Hájek, J., and Šidák, Z. 1967. Theory of Rank Tests. New York: Academic Press.

Halperin, M. 1952. Maximum likelihood estimation in truncated [meaning censored] samples. Annals of Mathematical Statistics 23: 226-38.

Harter, H. L. 1961. Expected values of normal order statistics. Biometrika 48: 151-65. Correction 48: 476.

Harter, H. L., and Moore, A. H. 1968. Maximum likelihood estimation, from doubly censored samples, of the parameters of the first asymptotic distribution of extreme values. Journal of the American Statistical Association 63: 889-901.

Hoeffding, W. 1948. A non-parametric test of independence. Annals of Mathematical Statistics 19: 546-57.

Hoeffding, W. 1951. A combinatorial central limit theorem. Annals of Mathematical Statistics 22: 558-66.

Johnson, N. L., and Kotz, S. 1970. Distributions in Statistics: Continuous Univariate Distributions - 1. Boston: Houghton Mifflin Company.

Kalbfleisch, J. D. 1978. Likelihood methods and nonparametric tests. Journal of the Americal Statistical Association 73: 167-70.

Kaplan, E. L. and Meier, P. 1958. Nonparmetric estimation from incomplete observations. Journal of the Americal Statistical Association 53: 457-81.

Majumdar, H., and Sen, P. K. 1978. Nonparametric testing for simple regression under progressive censoring with staggering entry and random withdrawal. Communications in Statistics - Theory and Methods A7(4): 349-71.

Miller, R. G. 1976. Least squares regression with censored data. Biometrika 63: 449-64.

Motoo, M. 1957. On the Hoeffding's combinatorial central limit theorem. Annals of the Institute of Statistical Mathematics, Tokyo 8: 145-54.

Plackett, R. L. 1958. Linear estimation from censored data. Annals of Mathematical Statistics 29: 351-60. [Note: errors in this paper are reported in Chan, 1967].

Rao, C. R. 1965. Linear Statistical Inference and Applications. New York: Wiley.

Sarhan, A. E. and Greenberg, B. G. (Editors) 1962. Contributions to Order Statistics. New York: Wiley.

Saw, J. G. 1958. Moments of sample moments of censored samples from a normal population. Biometrika 45: 211-21. Correction 45: 587-8.

Saw, J. G. 1959. Estimation of the normal population parameters given a singly censored sample. Biometrika 46: 150-5.

Saw, J. G. 1961. The bias of the maximum likelihood estimates of the location and scale parameters given a type II censored normal sample. Biometrika 48: 448-51.

Saw, J. G. 1962. Efficient moment estimators when the variables are dependent with special reference to type II censored normal data. Biometrika 49: 155-61.

Sen, P. K. 1976. Weak convergence of progressively censored likelihood ratio statistics and its role in asymptotic theory of life testing. Annals of Statistics 4: 1247-57.

Sen, P. K. 1977. Rank analysis of covariance under progressive censoring. Institute of Statistics Mimeo Series No. 1118, University of North Carolina.

Shirahata, S. 1975. Locally most powerful rank tests for independence with censored data. Annals of Statistics 3: 241-5.

Silvey, S. D. 1961. A note on maximum likelihood in the case of dependent random variables. Journal of the Royal Statistical Society Series B 23: 444-52.

Smith, D. W. 1978. A simplified approach to the maximum likelihood estimation of the covariance matrix. American Statistician 32: 28-9.

Watterson, G. A. 1959. Linear estimation in censored samples from multivariate normal populations. Annals of Mathematical Statistics 30: 814-24.

Zippin, C., and Armitage, P. 1966. Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. Biometrics 22: 665-72.