SOME PROPERTIES OF T CONFIDENCE INTERVALS FOR SURVEY DATA

Cathy Campbell and Michael Meyer, University of Minnesota

1. Introduction and Background

How to make reasonable inferences about nonlinear parameters of finite populations is a problem that often troubles survey statisticians. Exact properties of nonlinear statistics are not known even when simple random sampling procedures are used. When complex sample designs are employed, the problems become even more difficult.

Suppose we wish to form a 95% confidence interval for some nonlinear parameter θ, e.g., a variance. The usual approach is to

(1) calculate $\hat{\theta}$, a point estimate of θ;
(2) calculate $\hat{\sigma}(\hat{\theta})$, an estimated standard error;
(3) assume that $T = (\hat{\theta}-\theta)/\hat{\sigma}(\hat{\theta})$ has an approximate Student's-t distribution;
(4) let $\hat{\theta} \pm t_{.975}\hat{\sigma}(\hat{\theta})$ be the confidence interval.

Attempts to validate the above procedure have continued over many years. Most emphasis has focused on variance estimation -- how to calculate $\hat{\sigma}^2(\hat{\theta})$ so that it is a "good" estimator of $Var(\hat{\theta})$. Less attention has been given to the distributional properties of T.

Two important methods of variance estimation have emerged. One is the δ-method based on the first order Taylor approximation of θ. The other, a general class of sample reuse methods, has been developed for which replicate parameter estimates are calculated after systematically omitting portions of the sample data. Details of the methods for simple random and stratified sampling are given by Mellor (1973), McCarthy (1966, 1969), Tepping (1968), and Frankel (1971), among others.

Another large sampling study was done by Bean (1975) using data from the Health Interview Survey. Her sample sizes were substantially larger than Frankel's, and she concluded that both Taylor and BRR methods yield good variance estimates and similar T's.

There have been limited empirical and minimal theoretical evaluations of the confidence intervals produced with these variance estimates. Frankel (1971) used data from the Current Population Study for a large sampling study to compare confidence intervals using Taylor, balanced repeated replication (BRR), and jackknife replication (JKR) methods of variance estimation. He found that a version of BRR was superior to the others and gave adequate confidence intervals for several nonlinear statistics but not for squared correlations. Strangely enough the deviation between the t distribution and the distribution of T appeared to <u>increase</u> for larger sample sizes.

Mellor (1973) began to investigate the joint distribution of $(\hat{\theta},\hat{\sigma}(\hat{\theta}))$ in an attempt to discover why $T \sim t$ in some cases and $T \nsim t$ in others. Using computer generated simple random samples from infinite populations, Mellor performed two simulation studies. First he established that the standard "drop out 1" jackknife procedure produced better confidence intervals than the other "drop out m" replication methods or the Taylor method. Then he did a detailed study of the jackknife T for seven different parameters. Besides assessing the validity of the confidence intervals, he evaluated the normality of $\hat{\theta}$, compared the moments of $\hat{\sigma}^2(\hat{\theta})$ with those of a constant $\cdot \chi^2$ distribution, and calculated the correlation between $\hat{\theta}$ and $\hat{\sigma}(\hat{\theta})$. He concluded that normality of $\hat{\theta}$ is the most important factor affecting the "t-ness" of T, but the correlation between $\hat{\theta}$ and $\hat{\sigma}(\hat{\theta})$ seems to be important also.

Numerical studies such as those of Frankel, Bean, and Mellor have established the legitimacy of Taylor, BRR, or JKR T's in many situations. Theoretical justification for the use of these methods is largely lacking. Asymptotically all of the methods are valid, but small sample differences are already evident. Small sample distribution results have been obtained only for special cases involving linear estimators.

This leads to the question: when does a variable of the form

$$T = \frac{\hat{\theta} - \theta}{\hat{\sigma}(\hat{\theta})}$$

have an approximate t-distribution? We hope to approach this problem by studying T's for estimating different parameters for different populations using different variance estimates. Our first attempt will be to investigate how the characteristics of the joint distribution of $\hat{\theta}$ and $\hat{\sigma}(\hat{\theta})$ affect the "t-ness" of T.

In the next sections, we describe a computer program for generating finite populations of stratified, clustered bivariate data and present some preliminary results we have obtained with this program.

2. A Program for Generating Bivariate Populations

A key component of this study of t-confidence intervals has been the development of a computer program for generating finite populations of bivariate data. The populations created are stratified with clusters of elements as the primary sampling units. The program is very flexible and allows for forming populations with the following important features:

(1) variable-sized primary units both within (to be added) and between strata;
(2) general specification of the intraclass structure to allow for varying the degree of "clustering";
(3) specification of the correlation between the two variables;
(4) generation of normal, non-normal, or discrete (to be added) data with linear or nonlinear bivariate relationships;
(5) independent generation of each stratum of data allowing different stratum specifications.

A population consists of L strata each of which contains $A_h (h=1,\ldots, L)$ independent primary units (clusters) of M_h elements. As each stratum of data is formed independently, we shall describe the procedure for obtaining a single stratum. Each stratum contains $N_h = M_h A_h$ elements and $2N_h$ numbers since two variables (X and Y) are associated with each element.

Three primary steps are involved in the data generation:

(1) obtaining $2N_h$ iid N(0,1) variables;
(2) a linear transformation to create clusters with a desired level of correlation between X and Y;
(3) a transformation that sets the scale, location, marginal distributions, and shape of the regression line for X and Y.

To describe the data structure for a stratum the following notation will be convenient.

Let

$$\underline{U}_{hij} = \begin{pmatrix} X_{hij} \\ Y_{hij} \end{pmatrix} \qquad \begin{matrix} h=1,\ldots, L \\ i=1,\ldots, A_h \\ j=1,\ldots, M_h \end{matrix}$$

be the data for the j^{th} element in the i^{th} cluster of stratum h, and let

$$\underline{U}_{hi} = \begin{bmatrix} \underline{U}_{hi1} \\ \vdots \\ \underline{U}_{hiM_h} \end{bmatrix}$$

be the $2M_h$ vector of data for cluster i. We want the covariance matrix of $\underline{U}_{hi}$ to have the intraclass variance structure given by

$$\mathrm{Var}(\underline{U}_{hi}) = S_h = (\Sigma_h - P_h) \otimes I_{M_h} + P_h \otimes J_{M_h}$$

where J_{M_h} is an $M_h \times M_h$ matrix of ones,

$\Sigma_h(2\times 2)$ is the variance matrix of U_{hij}.

$P_h = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}$ is the covariance between $\underline{U}_{hij}$ and $\underline{U}_{hik}$ $(j \neq k)$.

The matrices Σ_h and P_h must be chosen so that Σ_h, $\Sigma_h - P_h$, and $\Sigma_h + (M_h - 1)P_h$ are all positive definite. Let $\underline{U}_h$ be the $2N_h$ vector of data for stratum h. Since the clusters are formed independently within each stratum, $\mathrm{Var}(\underline{U}_h) = S_h \otimes I_{A_h}$.

To achieve this structure, we begin by generating $\underline{t}_{hi} \sim N(0, I_{2M_h})$ as the initial vector of data for a cluster. The vector $\underline{t}_{hi}$ then undergoes a linear transformation to create the desired correlation structure. The linear transformation is not unique, but a useful computational form has been found via the spectral decomposition of S_h. After the clusters have been created another, possibly nonlinear, transformation is then performed which sets (a) the location, scale, and marginal distributions of X and Y, and (b) the shape of the bivariate distribution of X and Y.

Presently we are choosing this final transformation from the Johnson (1949a,b) translation system of bivariate distributions. This large and versatile system of bivariate distributions was chosen because

(1) the transformation is defined on bivariate normal data;
(2) the available marginal distributions for X and Y allow a four-moment approximation of any continuous distribution;
(3) the bivariate relationships between X and Y include linear and nonlinear relationships of varying shapes;
(4) the transformations are easy to compute.

After this transformation is complete, either variable could be categorized to define dichotomous or other discrete distributions.

Any nonlinear transformations are performed after the clusters are created; thus the actual values of the correlation and intraclass correlations may not be exactly equal to those specified by Σ_h and P_h. Since the transformations are all monotonic we do not expect the difference to be important.

By generating each stratum of data independently, differences between strata can be created in a number of ways:

(1) using different cluster sizes within each stratum;
(2) choosing different Σ_h and P_h matrices;
(3) using different transformations from the Johnson system or using the same basic transformation with different parameters.

By using all the different options of this data generation system, we will be able to create stratified clustered populations with varying and interesting properties.

All computer programs described in this paper were written in Fortran and run on a CDC 6400 computer. Uniform random numbers are generated using the multiplicative congruential method as described by Barnett (1962). Normal random numbers are generated from uniform random numbers by choosing randomly from one of five transformations. The method is described by Marsaglia and Gray (1964).

3. Sampling and Estimation Procedures

This study is based on paired selection of primary units from each stratum. A sample is drawn by randomly choosing a pair (without replacement) of clusters from each stratum. To investigate the sampling distribution of various statistics, sets of K independent samples are chosen from the same finite population. In this section we describe the statistics calculated from each sample and the summary statistics computed for each set of K samples.

The values reported in the summary output for each set of K samples were chosen to help answer the following questions:

(1) When is the distribution of the sample T statistics approximately t with appropriate degrees of freedom?
(2) When do two-sided 95% and 90% confidence intervals have the correct error rate?
(3) How do the properties of the joint distribution of the parameter point estimate and variance estimate influence the distribution of the T statistics?

In order to cover a variety of situations, we decided to estimate 12 different parameters and to calculate 9 different T's for each parameter.

The parameters to be estimated, with identification, are given in Table 3.1.

Table 3.1

Parameters Estimated for Bivariate Population

i	θ_i	i	θ_i
1	$\overline{X}$	7	$B_{y\cdot x}$*
2	$\overline{Y}$	8	$B_{x\cdot y}$
3	$S_x^{\ 2}$	9	R_{xy}**
4	$S_y^{\ 2}$	10	$\ell n(S_x^{\ 2})$
5	$\overline{X}/\overline{Y}$	11	$\ell n(S_y^{\ 2})$
6	$\overline{Y}/\overline{X}$	12	$\frac{1}{2}\ell n\left(\frac{1+R_{xy}}{1-R_{xy}}\right)$

*regression coefficient of Y on X.
**Pearson-product correlation between X and Y.

The population parameters are evaluated by taking the appropriate functions of the finite population data. Parameters 10, 11, and 12 were included to study the effect of "normalizing" transformations.

For each sample and for each θ_i, 9 different T values of the form $T = (\hat{\theta}_i-\theta_i)/\hat{\sigma}(\hat{\theta}_i)$ were calculated by using different methods to find $\hat{\theta}_i$ and $\hat{\sigma}(\hat{\theta}_i)$. The Taylor expansion method, jackknife replication, and balanced repeated replication are the basic forms used. Methods employed here are analogous to those used by Frankel (1971). Exact formulas are given here only if they differ from his.

We introduce the following notation:

$\hat{\theta}_{iT}$ = "whole sample" or first order Taylor estimate of θ_i

$\{\hat{\theta}_{ij}^{\ J};j=1,\ldots,L\}$ = individual jackknife replicate estimates of θ_i

$\hat{\theta}_{iJ}$ = mean of $\{\hat{\theta}_{ij}^{\ J}\}$

$\{\hat{\theta}_{ij}^{\ B};j=1,\ldots,k\}$ = individual balanced repeated replication estimates of θ_i

$\hat{\theta}_{iB}$ = mean of $\{\hat{\theta}_{ij}^{\ B}\}$

$\sigma_{iT}^{\ 2}$ = first order Taylor expansion estimate of $\mathrm{Var}(\hat{\theta}_{iT})$

$$\hat{\sigma}_{iJT}^{\ 2} = (1-f)\sum_{j=1}^{L}(\hat{\theta}_{ij}^{\ J} - \hat{\theta}_{iT})^2$$

$$\hat{\sigma}_{iJR}^{\ 2} = (1-f)\sum_{j=1}^{L}(\hat{\theta}_{ij}^{\ J} - \hat{\theta}_{iJ})^2$$

$$\hat{\sigma}_{iBT}^{\ 2} = \frac{1-f}{k}\sum_{j=1}^{k}(\hat{\theta}_{ij}^{\ B} - \hat{\theta}_{iT})^2$$

$$\hat{\sigma}_{iBR}^{\ 2} = \frac{1-f}{k}\sum_{j=1}^{k}(\hat{\theta}_{ij}^{\ B} - \hat{\theta}_{iB})^2 .$$

The 9 different T's are defined in Table 3.2.

Table 3.2

Definition of T Statistics			
T	Mnemonic Identification	Point Estimate	Variance Estimate
1	Taylor	$\hat{\theta}_{iT}$	$\hat{\sigma}_{iT}^{\ 2}$
2	JKRT	$\hat{\theta}_{iJ}$	$\hat{\sigma}_{iJT}^{\ 2}$
3	JKRR	$\hat{\theta}_{iJ}$	$\hat{\sigma}_{iJR}^{\ 2}$
4	JKTT	$\hat{\theta}_{iT}$	$\hat{\sigma}_{iJT}^{\ 2}$
5	JKTR	$\hat{\theta}_{iT}$	$\hat{\sigma}_{iJR}^{\ 2}$
6	BHRT	$\hat{\theta}_{iB}$	$\hat{\sigma}_{iBT}^{\ 2}$
7	BHRR	$\hat{\theta}_{iB}$	$\hat{\sigma}_{iBR}^{\ 2}$
8	BHTT	$\hat{\theta}_{iT}$	$\hat{\sigma}_{iBT}^{\ 2}$
9	BHTR	$\hat{\theta}_{iT}$	$\hat{\sigma}_{iBR}^{\ 2}$

Because of computational limitations the corresponding estimates for the complement samples were not calculated in either the jackknife replication or balanced repeated replication methods. Combining the complement and main samples seems quite desirable especially with jackknife replication where there is an inherent imbalance in the samples because only one observation from each stratum is ever dropped to form a replicate.

For each sample drawn from a population, the following quantities are calculated for i=1,...,12 and saved for later summary:

(1) individual replicate estimates $\hat{\theta}_{ij}^{\ J}$ (j=1,...,L); $\hat{\theta}_{ij}^{\ B}$(j=1,...,k)

(2) parameter point estimates $\hat{\theta}_{iT}$, $\hat{\theta}_{iJ}$, $\hat{\theta}_{iB}$

(3) variance point estimates $\hat{\sigma}_{iT}^{\ 2}$, $\hat{\sigma}_{iJT}^{\ 2}$, $\hat{\sigma}_{iJR}^{\ 2}$, $\hat{\sigma}_{iBT}^{\ 2}$, $\hat{\sigma}_{iBR}^{\ 2}$

(4) the 9 T-statistics defined in Table 3.2.

After choosing a set of K samples from the same finite population, the following summary statistics are calculated and included in the computer output:

(1) for each point estimate, variance estimate, and T in (2), (3), and (4) above, the following sample moments and functions of them:

m_1', m_2, m_3, m_4; v_1 = mean absolute deviation

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}};\ b_2 = \frac{m_4}{m_2^2};\ G = \frac{v_1}{m_2^{1/2}};$$

(2) the between-sample and within-sample variance of the $\hat{\theta}_{ij}^{\ B}$ and $\hat{\theta}_{ij}^{\ J}$, and the intrasample correlations;

(3) for each of the T's
   (a) $r(\hat{\theta}_i,\hat{\sigma}_i^2)$ = sample correlation between point estimate of θ_i and variance estimate;
   (b) $r(|\hat{\theta}_i-\theta_i|,\hat{\sigma}_i^2)$;

(4) for each of the T's, the number of values falling in specified intervals of the t-distribution.

Other calculations and tabulations done as part of the analysis will be described in later sections.

4. Data Base for Present Analysis

Results reported here are based on K=200 samples chosen from each of 9 finite populations. Each population consists of either 6 or 12 strata, and each stratum contains A_h=40 clusters. Selecting two clusters without replacement from each stratum gives a finite population correction of (1-f)=0.95 in all cases. Unless stated otherwise all populations have clusters which contain M_h=6 elements, and all strata were identically generated. Descriptions of the populations are given in Table 4.1. Study of the table reveals that we have hardly begun to explore the variety of population structures that can be created with this program. The richness of the Johnson system and methods for specifying stratification remain largely untouched.

Table 4.1

Description of Finite Population

Pop'n ID	# of Strata	Parameters of Generating Distribution ρ_{xy}	P	X	Y
1	12	.7	$\begin{bmatrix} .5 & .3 \\ .3 & .5 \end{bmatrix}$	N(10,4)	N(30,25)
2-6	6	0	∅	N(10,4)	N(30,25)
2-12	12	"	"	"	"
3	6	.7	∅	N(10,4)	N(30,25)
4-6	6	0	∅	Johnson f_u: $\mu=24, \sigma^2=36$, $\sqrt{\beta_1}=0, \beta_2=36$	f_L=log-normal $\mu=1, \sigma^2=.65$ $\sqrt{\beta_1}=1.75, \beta_2=8.9$
4-12	12	"	"	"	"
5-6	6	.7	$\begin{bmatrix} .2 & .05 \\ .05 & .1 \end{bmatrix}$	N(10,4)	N(30,25)
5-12	12	"	"	"	"
6*	12	.7	$\begin{bmatrix} .2 & .05 \\ .05 & .1 \end{bmatrix}$	N(10,4)	N(30,25)

*Stratification: 4 strata with M_h = 3
4 strata with M_h = 6
4 strata with M_h = 12

5. Preliminary Results

In this section we present the results of some simple analyses done to determine when the T statistics yield 95% and 90% confidence intervals for θ_i which have the correct error rate. The critical values of t used in this analysis were obtained from the t-distribution with L-1 degrees of freedom. Although others have used L degrees of freedom, Frankel (1971) notes that there is no theoretical basis for doing so. We chose L-1 because the rank of the quadratic forms involved in calculating $\hat{\sigma}^2_{iJR}$ and $\hat{\sigma}^2_{iBR}$ for sample means is L-1. A survey of the intervals does not reveal an excessive number of undercoverages; however, the issue of appropriate degrees of freedom needs to be examined more thoroughly in the future.

We wish to assess the effect of the following variables on the distribution of the T's:

(1) characteristics of population structure, such as cluster size, choice of Σ_h and P_h, stratification effects, marginal distributions of X and Y, etc.;
(2) functional form of the paramter estimate;
(3) type of t-statistic being used.

The effect of population structure can be estimated only by making comparisons across different populations. Many populations will have to be generated before these effects can be accurately determined. Each parameter is estimated once for every population, and each T is calculated for every parameter. The nested arrangement of these variables is analogous to the arrangement of treatments in a split-split plot design with population characteristics as whole plot treatments, parameters as subplot treatments, and T's as sub-subplot treatments. This hierarchical structure of treatments means that T effects can be measured most accurately and population effects least accurately.

With the limited data we now have, we will focus on comparing the different T's and make only general comments about the parameter and population comparisons. The results presented here are based only on the properties of the 95% and 90% confidence intervals for θ_i.

With K=200 samples chosen from each population we expect E_{95}=10 T's to fall outside the 95% central region of the t-distribution and E_{90}=20 T's to all outside the 90% region. For a given sample T distribution, we observe the frequencies O_{95} and O_{90} of T's outside the specified regions. If the distribution of T is in fact t, then $\sigma(O_{95})$=3.08 and $\sigma(O_{90})$=4.24. A set of 200 T's is considered to yield "adequate" 95% confidence intervals if $|O-E| \leq \sigma(O)$. If all the T-distributions are true t's, about two-thirds of the confidence intervals should be adequate.

In Table 5.1 we examine the adequacy of the confidence intervals by T and by population. The numbers in the body of the table are the number of parameters (θ_i) for which the confidence intervals are adequate with that particular T.

Looking at the row marginals of Table 5.1, we quickly see that the balanced repeated replication T's yield the best confidence intervals. The two T's, BHTT and BHTR, which use the whole sample point estimate $\hat{\theta}_{iT}$ have a slight edge. The worst confidence intervals come from the JKRR and JKTR T's which use $\hat{\sigma}^2_{iJR}$, the sum of squares around the replicate mean, as the variance estimator. Falling between these two groups are the JKTT, JKRT, and Taylor T's.

Our numerical results indicate that the poor performance of JKRR and JKTR may be partially explained by a negative bias in the variance estimator $\hat{\sigma}^2_{iJR}$. A simple calculation shows that $\hat{\sigma}^2_{iJR}$ does in fact have a small negative bias when θ_i is a population mean or total and that the bias approaches zero when L is large.

Other comparisons of the T's based on the values of χ^2 goodness-of-fit statistics and other functions of the coverage rates lend firm support to the general superiority of the balanced

Table 5.1

Adequacy of Confidence Intervals by T and by Population*

T		Population 1	2-6	2-12	3	4-6	4-12	5-6	5-12	6	% adequate
Taylor	95%	3	9	10	7	4	5	10	9	9	61%
	90%	3	9	11	11	3	4	7	11	4	58
JKRT	95%	3	10	8	8	3	5	10	10	9	61
	90%	4	10	9	9	3	4	7	12	7	60
JKRR	95%	0	5	8	2	3	5	3	6	4	33
	90%	1	2	7	1	3	3	1	11	4	31
JKTT	95%	3	9	9	10	4	5	11	9	11	66
	90%	4	9	10	10	3	5	7	11	6	60
JKTR	95%	0	5	8	3	1	5	3	6	5	33
	90%	2	2	8	2	4	4	2	9	5	35
BHRT	95%	4	8	9	9	6	8	9	11	10	69
	90%	7	9	10	12	6	7	8	11	5	69
BHRR	95%	4	8	9	10	6	8	9	11	9	69
	90%	6	9	10	12	6	6	9	12	5	69
BHTT	95%	7	8	10	9	6	8	7	10	11	70
	90%	7	10	10	11	6	6	10	11	6	71
BHTR	95%	6	8	10	10	6	9	8	10	11	72
	90%	7	10	10	11	6	6	9	11	6	70
% adequate	95%	28%	65	75	63	36	54	65	76	73	59
	90%	38%	65	79	73	37	42	56	92	44	58

*Entries are the number out of 12 parameters for which confidence intervals are adequate.

Table 5.2

Adequacy of 95% Confidence Intervals by Parameter Type*

Parameter Groups	Population 1	2-6	2-12	3	4-6	4-12	5-6	5-12	6	% adequate
$\bar{X}$	0	3	1	7	1	9	9	9	8	68%
$\bar{Y}$	7	8	6	0	7	9	9	8	9	
S_x	0	0	0	4	0	0	0	4	0	31
S_y	0	7	8	7	0	0	7	8	6	
ℓnS_x^2	1	7	3	7	0	0	7	7	3	51
ℓnS_y^2	2	9	9	5	0	1	8	6	8	
** $\bar{X}/\bar{Y}$		9	9		9	9				99
$\bar{Y}/\bar{X}$		9	9		8	9				
*** $\bar{X}/\bar{Y}$	0			7			7	9	8	73
$\bar{Y}/\bar{X}$	5			7			7	9	7	
** $B_{y \cdot x}$		1	9		2	4				63
$B_{x \cdot y}$		7	9		4	9				
*** $B_{y \cdot x}$	5			9			3	9	6	69
$B_{x \cdot y}$	6			5			3	7	9	
** R_{xy}		7	9		4	4				67
*** R_{xy}	0			4			6	0	8	40
** Z(R)		3	9		4	4				56
*** Z(R)	4			6			4	6	8	62

*Entries are the number (out of 9) T's that had adequate 95% confidence intervals.

** $R_{xy} \approx 0$

*** $R_{xy} \approx 0.7$

repeated replication T's and the general inferiority of JKRR and JKTR.

The column marginals of Table 5.1 allow some comparisons to be made among the different populations. The relations we expect to find do appear, but it is too early to assess their magnitude. Large populations (more strata) are better than small. Unclustered populations are better than clustered. Normal distributions are better than skew, heavy-tailed distributions. The stratified population (6) is somewhat worse than the corresponding unstratified populations (2-12 and 5-12), but this effect is masked by the larger cluster sizes in population 6.

Some slightly more curious results can be found in Tables 5.2 and 5.3 which look at the adequacy of the confidence intervals as a function of the type of parameter being estimated. For Table 5.2 we grouped the parameters into the following classes:

(1) means
(2) variances
(3) log-variances
(4) ratios with $R_{xy} \approx 0$
(5) ratios with $R_{xy} \approx 0.7$
(6) regression coefficients with $R_{xy} \approx 0$
(7) regression coefficients with $R_{xy} \approx 0.7$
(8) correlations with $R_{xy} \approx 0$
(9) correlations with $R_{xy} \approx 0.7$
(10) Fisher's z with $R_{xy} \approx 0$
(11) Fisher's z with $R_{xy} \approx 0.7$.

The data in the body of Table 5.2 are the number (out of 9) of the T's that gave adequate 95% confidence intervals for a given parameter-population combination.

Before over-interpreting the results in Table 5.2 we should note that the behavior of the 9 T's within a population-parameter cell is very highly correlated; therefore, the effective sample size for estimating the marginal percentages is not nearly so large as it seems. The effects discussed here are supported by other calculations not included in this manuscript.

(1) The easiest parameter to estimate is the ratio of two independent means. This is true even in the extremely non-normal populations 4-6 and 4-12.
(2) Confidence intervals for means are more erratic than expected. This is the parameter for which the theoretical t-distributions should be valid.
(3) Variances are practically impossible to estimate with the sample sizes and populations considered here. The logarithmic transformation helped, but not so much as we had hoped.

(4) The value of Fisher's z-transformation for estimating correlations is questionable. Other tabulations have shown the approximate t-distribution for transformed correlations to be worse than this table indicates.

Another interesting view of the confidence intervals is given in Table 5.3 which tabulates the number of adequate confidence intervals by parameter and T summing across populations. The numbers in the table are the number (out of 9) of populations for which the 95% and 90% confidence intervals are adequate.

The most important feature of Table 5.3 is the general utility of balanced repeated replication for estimating all parameters except variances and log-variances. Remember, only two-thirds of the confidence intervals should be "adequate" -- with our definition -- even under ideal circumstances. The poor behavior of the confidence intervals of the $\bar{X}$'s is an anomoly for which we have no explanation, but it seems to be related to the poor results for S_x^2. Ratios seem to be estimated well by all T's except JKRR and JKTR. The poor performance of these T's occurs for all parameters.

Table 5.3

Number of Adequate Confidence Intervals for Parameter by T Combinations*

Parameter		Taylor	JKRT	JKRR	JKTT	JKTR	BHRT	BHRR	BHTT	BHTR
$\bar{X}$	95%	5	6	6	5	5	5	5	5	5
	90%	6	6	3	4	5	6	6	6	6
$\bar{Y}$	95%	7	7	4	8	3	8	8	8	8
	90%	8	7	6	8	6	8	8	8	8
S_x^2	95%	1	0	0	1	0	1	1	2	2
	90%	3	2	1	3	1	2	2	5	4
S_y^2	95%	6	6	1	6	2	6	5	5	6
	90%	4	3	2	4	2	5	5	4	4
$\bar{X}/\bar{Y}$	95%	8	8	5	9	6	7	8	8	9
	90%	8	8	4	8	4	7	7	7	7
$\bar{Y}/\bar{X}$	95%	9	9	5	9	4	8	8	9	9
	90%	6	8	2	6	5	6	6	6	6
$B_{y \cdot x}$	95%	6	4	4	5	3	6	6	7	7
	90%	5	7	2	5	3	8	9	9	9
$B_{x \cdot y}$	95%	7	7	3	9	3	8	8	7	7
	90%	7	6	3	4	4	7	7	6	7
R_{xy}	95%	4	3	1	4	2	7	7	7	7
	90%	5	5	3	5	2	7	6	7	7
$\ell n S_x^2$	95%	5	4	0	5	0	4	4	7	6
	90%	4	3	1	7	0	6	5	7	7
$\ell n S_x^2$	95%	5	6	5	5	6	5	6	4	6
	90%	5	4	2	3	3	4	5	4	3
$z(R_{xy})$	95%	2	6	2	6	2	7	8	7	7
	90%	2	5	3	7	2	9	9	8	8

*Entries are the number out of 9 populations for which the confidence intervals are adequate.

6. Summary

Our ultimate goal is to learn how the distribution of sample T's is affected by the joint distribution of $\hat{\theta}$ and $\hat{\sigma}^2(\hat{\theta})$. To begin the investigation we have developed a method for generating data to help elucidate this relationship. A first step in the analysis is to characterize the situations (populations, sample sizes, parameters, types of estimators) which produce good T's and those which produce bad T's.

We have made some progress in this direction. An important result is that the BHTT and BHTR T's are superior to the others and are quite reliable for all parameters except variances. As others have found for simple random sampling, we have discovered that the T's for estimating variances are erratic. Our future analyses will incorporate data from the complement samples. In particular, this should considerably improve the performance of the JKRT's.

Literature Cited

Barnett, V.C. (1962). "The behavior of pseudo-random sequences generated on computers by the multiplicative congruential method." Mathematics of Computation, 16, 63-69.

Bean, J.A. (1975). Distribution and Properties of Variance Estimators for Complex Multistage Probability Samples. National Center for Health Statistics, Series 2, Number 65.

Frankel, M.R. (1971). Inference from Survey Samples. Institute for Social Research, The University of Michigan, Ann Arbor, MI.

Johnson, N.L. (1949a). "Systems of frequency curves generated by methods of translation." Biometrika, 36, 149-176.

Johnson, N.L. (1949b). "Bivariate distributions based on simple translation systems." Biometrika, 36, 297-304.

McCarthy, P.J. (1966). Replication: An Approach to the Analysis of Data from Complex Surveys. National Center for Health Statistics, Series 2, Number 14.

McCarthy, P.J. (1969). Pseudoreplication: Further Evaluation and Application of the Balanced Half-Sample Technique. National Center for Health Statistics, Series 2, Number 31.

Marsaglia, G. and T.A. Gray (1964). "A convenient method for generating normal variables." SIAM Review 6, 3, 260-264.

Mellor, R.W. (1973). "Subsample replication variance estimators." Unpublished Ph.D. dissertation, Harvard University.

Tepping, B.J. (1968). "The estimation of variance in complex surveys." Proceedings of the Social Statistics Section of the American Statistical Association.