# BAYESIAN PREDICTIVE DENSITIES AND THEIR APPLICATION TO SAMPLING THEORY

By

MELVIN VALONE FOSTER

Bachelor of Science
Southern State College
Magnolia, Arkansas
1956

Master of Science
University of Arkansas
Fayetteville, Arkansas
1963

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1973

BAYESIAN PREDICTIVE DENSITIES AND THEIR

APPLICATION TO SAMPLING THEORY

Thesis Approved:

Lyle D. Broemeling
Thesis Adviser

David L. Weeks

P. L. Claypool

G. T. Stevens, Jr.

D. D. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my gratitude to my adviser, Professor Lyle D. Broemeling, for his guidance, encouragement, and patience in the preparation of this thesis.

Also, I wish to express my appreciation to Professors P. Larry Claypool, David L. Weeks, and Gladstone T. Stevens, Jr. for serving on my advisory committee.

A special note of thanks is extended to my family, the statistics faculty and many fellow graduate students at Oklahoma State University for their interest and encouragement during my entire course of graduate study.

TABLE OF CONTENTS

# CHAPTER I

# INTRODUCTION

The objective of this thesis is to develop a system of statistical inference for finite populations. This inference system is based on a variation of the predictive approach utilized by Kalbfleisch and Sprott [1]. They used the Fisher fiducial approach to derive predictive densities whereas we will use the Bayesian approach.

## Predictive Distribution

Geisser [2] designates a predictive distribution as the distribution of an observable random variable whose distribution is completely specified as to form and constants. He also states that distributions are not rendered predictive by substituting estimates for the parameters, nor shall the posterior distribution of a vector parameter $\theta$ attain predictive status unless $\theta$ is an observable variable.

Since our interest in predictive distributions is within the Bayesian frame work, we will formulate a definition of predictive distributions in this sense. We should note, however, that predictive distributions exist within the classical framework. For example, if $X \sim N(\mu,1)$ and $Y \sim N(\mu,1)$, then $X - Y \sim N(0,2)$ would be a predictive density although not a particularly informative one. Also, as previously indicated predictive densities may be derived by the Fisher-fiducial approach.

Suppose a set of $N$ independent observations, summarized by $D$, are available on $F_x(\cdot|\theta)$ where $\theta$ is assumed to have the prior distribution $\pi(\theta)$. Also, assume the posterior distribution $P(\theta|D)$ exists and $G_y(\cdot|\theta)$ is the distribution of a function, $y = f(x_1, \ldots, x_k)$, of $k$ future observations, $x_1, x_2, \ldots, x_k$. The predictive distribution of $y$ is defined by:

$$F_y(\cdot|D) = E_\theta[G_y(\cdot|\theta)] = \int G_y(\cdot|\theta) dP(\theta|D).$$

Hence, the predictive distribution is the average of all conditional distributions of $y$ given $\theta$ with respect to the posterior distribution of $\theta$.

## The Super Population Concept

Let $U_i$ be the variate value attached to the $i^{th}$ unit in the finite population of $N$ numbered and distinguishable units. We will consider the finite population $(U_1, \ldots, U_N)$ as a vector in $R^N$, where $R$ is the real numbers. This population may be sampled by picking $n$ random integers from $1, \ldots, N$ and examining the variate values, $U_{i_1}, U_{i_2}, \ldots, U_{i_n}$. For simplicity, we will call these values $x_1, x_2, \ldots, x_n$. From these values we wish to make an inference about some function of the finite population values. In particular, we are interested in making an inference about the mean, $\bar{U} = \sum_1^N U_i/N$, of the finite population.

Cochran [3], [4] suggests that in many instances the value attached to the $i^{th}$ population unit is the realization from a super distribution. That is, the finite population can be considered as the result of a random physical process described by a probability distribution. For example, the finite populations of heights, weights, or intelligence can be considered as large random samples from a normal distribution defined by genetical mechanisms giving rise to the finite population.

Under this concept, a simple random sample from the finite population is also a simple random sample from the super population; hence an inductive inference may first be made to the super population and then a deductive inference may be made about the finite population from which the sample was obtained. We propose the following procedure to accomplish this.

The finite population is considered to be a simple random sample of size $N$ from a super population whose density is given by $f(x|\theta)$. The prior density of $\theta$ is given by $\pi(\theta)$. A simple random sample of size $n$ is drawn from the finite population and the corresponding values attached to the sample units are summarized by $D_n$. Let $P(\theta|D_n)$ be the posterior density of $\theta$ and let $y$ be some function of the values attached to the remaining finite population units. If $h(y|\theta)$ is the density of $y$, the predictive density of $y$ is defined by

$$f(y|D_n) = \int_{\theta} h(y|\theta)P(\theta|D_n)d\theta$$

which is to be utilized in making inferences about the finite population parameters.

## Organization of Thesis

In Chapter II the predictive procedure will be used to make inferences about the mean of the finite population assuming that the super population density is (1) Bernoulli, (2) exponential and (3) normal. In particular, we will use the mean of the predictive density of the finite population mean as a point estimator for the mean of the finite population. Also, if the super population density is normal, we will use the mean of the predictive density of the finite population variance as a point estimator for the variance of the finite population.

Chapter III is an extension of Chapter II. In this chapter, we assume that the super population density is normal with mean $\alpha x^g + \beta x^{g+1}$ and variance $(\sigma x^g)^2$, where $g = 0$ or $1$ and the x's are nonstochastic variables. In this case our finite population is a vector in $(R^2)^N$, where $R$ is the real numbers. The value attached to each unit then is an ordered pair $(U_i, V_i)$, $i = 1, \ldots, N$. On the basis of a simple random sample $(y_1, x_1), \ldots, (y_n, x_n)$, we will make inferences concerning $\sum_1^N U_i/N$ if $g = 0$ and concerning $\sum_1^N U_i \big/ \sum_1^N V_i$ if $g = 1$. In either case we will assume that

$$\sum_1^N V_i \quad \text{and} \quad \sum_1^N V_i^2$$

are known.

We will apply the results of Chapters II and III to stratified simple random sampling in Chapter IV. We assume that each strata is a random sample from a super population and make inference for the overall mean, $u = \frac{1}{N}\sum_{1}^{k} N_i\overline{U}_i$, where $\overline{U}_i$ is the mean of the $i^{th}$ stratum, $N_i$ is the number of units in the $i^{th}$ stratum, and $\sum_{1}^{k} N_i = N$.

Chapter V is a study of optimum allocation of sampling units among $k$ strata. Our criterion for optimality will be minimization of the variance of the predictive density subject to a fixed cost function. If there is no prior information concerning the within strata variance, we will use a two-phase sampling procedure as utilized by Draper and Guttman [5] for a Bayesian approach to allocation. We also consider allocation for estimating $r \leq k$ ($k$ number of strata) parametric linear functions of the strata means. (Des Raj [6])

# CHAPTER II

## SIMPLE RANDOM SAMPLING

This chapter is devoted to a study of simple random sampling from a finite population utilizing predictive densities. We assume the finite population $(U_1, U_2, \ldots, U_N)$ is a simple random sample from (i) a Bernoulli, (ii) an exponential, or (iii) a normal super population distribution. In all three cases, we will let $x_i$, $i = 1,2, \ldots, n$, denote the observed value attached to unit $i$ in a simple random sample without replacement of size $n$ from the finite population and will let $y_j$, $j = 1,2, \ldots, N-n$, designate the unknown value attached to the $j^{th}$ unsampled unit in the finite population. Also, we assume the prior information can be expressed either by a Jeffrey's vague prior [7] or by a conjugate prior distribution [8].

### Bernoulli Super Population

We assume the finite population $(U_1, U_2, \ldots, U_N)$ is a simple random sample from a Bernoulli distribution. We will let $p$ be the probability that $U_i = 1$ (or that $U_i$ is a success). The probability that $U_i = 0$ (or that $U_i$ is a failure) is $1 - p = q$. Our interest will be in estimating the number of successes in the finite population, or equivalently,

$$U = \sum_{1}^{N} U_i \tag{2.1}$$

on the basis of a simple random sample of size $n < N$ from the finite population.

If we define

$$X = \sum_{1}^{n} x_i$$

and

$$Y = \sum_{1}^{N-n} y_i$$

then $U$ can be expressed as

$$U = Y + X. \tag{2.2}$$

For a point estimate of $U$, we will use

$$E(U) = E(Y) + X$$

where $E(Y)$ is determined from the predictive density of $Y$. As a measure of the precision of our prediction, we use

$$V(U) = V(Y).$$

Theorem 2.1. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is a Bernoulli with unknown parameter $p$. Also, suppose the prior density on $p$ is a Jeffrey vague prior of the form

$$\pi(p) \propto \frac{1}{p(1-p)}, \quad 0 < p < 1.$$

Then the predictive density of $Y$ is

$$f(Y|X,N,n) = \frac{\binom{Y+X-1}{Y}\binom{N-X-Y-1}{N-n-Y}}{\binom{N-1}{N-n}}, \quad Y = 0,1, \ldots, N-n.$$

Proof: Zellner [7] derives the posterior of $p$ as

$$P(p|X,n) = \frac{\Gamma(n)}{\Gamma(X)\Gamma(n-X)} p^{X-1}(1-p)^{n-X-1}, \quad 0 < p < 1$$

and $X = 0,1, \ldots, n$. The distribution of $Y$ given $p$ and $N-n$ is

$$g(Y|p,N-n) = \binom{N-n}{Y} p^{Y}(1-p)^{N-n-Y}, \quad Y = 0,1, \ldots, N-n. \tag{2.3}$$

Hence, the predictive density of $Y$, $E[G(Y|p,N-n)]$, is

$$f(Y|X,N,n) = \int_0^1 \frac{\Gamma(n)}{\Gamma(X)\Gamma(n-X)} \binom{N-n}{Y} p^{X+Y-1}(1-p)^{N-X-Y-1} dp.$$

Integrating and simplifying, we obtain

$$f(Y|X,N,n) = \frac{\binom{Y+X-1}{Y}\binom{N-X-Y-1}{N-n-Y}}{\binom{N-1}{N-n}}, \quad Y = 0,1, \ldots, N-n. \tag{2.4}$$

Corollary 2.1. If the assumptions of Theorem 2.1 hold, then:

(a) $E(U) = N \frac{X}{n}$

(b) $V(U) = \frac{(N-n)N}{n+1} \frac{X}{n} \left( 1 - \frac{X}{n} \right)$.

Proof:

(a) $$\sum_{Y=0}^{N-n} Y \binom{Y+X-1}{Y} \binom{N-X-Y-1}{N-n-Y} =$$

$$= X \sum_{Y=0}^{N-n-1} \binom{n-X+N-n-1-Y-1}{N-n-1-Y} \binom{X+1+Y-1}{Y}.$$

From the equality,

$$\sum_{j=0}^{k} \binom{a+k-j-1}{k-j} \binom{b+j-1}{j} = \binom{a+b+k-1}{k}, \tag{2.5}$$

given by Feller [9] on page 65, it follows that

$$\sum_{Y=0}^{N-n} Y \binom{Y+X-1}{Y} \binom{N-X-Y-1}{N-n-Y} = X \binom{N-1}{N-n-1}.$$

Therefore, from (2.4)

$$E(Y) = \frac{N-n}{n} X$$

and from (2.2)

$$E(U) = N \frac{X}{n}. \tag{2.6}$$

(b) $$\sum_{Y=0}^{N-n} Y(Y-1) \binom{Y+X-1}{Y} \binom{N-X-Y-1}{N-n-Y} =$$

$$= X(X+1) \sum_{Y=0}^{N-n-2} \binom{n-X+N-n-2-Y-1}{N-n-2-Y} \binom{X+2+Y-1}{Y}$$

$$= X(X+1) \binom{N-1}{N-n-2}.$$

The last equality follows by equation (2.5). Therefore, from (2.4)

$$E[Y(Y-1)] = \frac{X(X+1)(N-n)(N-n-1)}{n(n+1)} . \qquad (2.7)$$

Now

$$V(Y) = E[Y(Y-1)] + E(Y) - E^2(Y).$$

Hence, substituting equations (2.6) and (2.7) in the above and simplifying, yields

$$V(U) = V(Y) = \frac{(N-n)N}{n+1} \frac{X}{n} \left( 1 - \frac{X}{n} \right).$$

<u>Theorem</u> 2.2. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is a Bernoulli with unknown parameter $p$. Also, suppose the prior density on $p$ is a Beta conjugate prior of the form

$$\pi(p|\alpha,\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad 0 < p < 1, \quad \alpha,\beta > 0.$$

Then the predictive density of $Y$ is

$$f(Y|X,\alpha,\beta,N,n) = \frac{\binom{Y+X+\alpha-1}{Y}\binom{N+\beta-X-Y-1}{N-n-Y}}{\binom{N+\alpha+\beta-1}{N-n}}, \quad Y = 0,1, \ldots, N-n.$$

Proof: The posterior distribution of $p$ as given by LaValle [10] on page 340 is

$$P(p|X,\alpha,\beta,n) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(X+\alpha)\Gamma(n-X+\beta)} p^{X+\alpha-1}(1-p)^{n-X+\beta-1}, \quad 0 < p < 1.$$

From above and equation (2.3), we can express the predictive density of Y as

$$f(Y|X,\alpha,\beta,N,n) = \int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(X+\alpha)\Gamma(n-X+\beta)} \binom{N-n}{Y} p^{X+Y+\alpha-1} (1-p)^{N-X-Y+\beta-1}\, dp$$

which reduces to

$$f(Y|X,\alpha,\beta,N,n) = \frac{\binom{Y+X+\alpha-1}{Y}\binom{N+\beta-X-Y-1}{N-n-Y}}{\binom{N+\alpha+\beta-1}{N-n}}, \quad Y = 0,1,\ \ldots,\ N-n. \tag{2.8}$$

<u>Corollary</u> 2.2. If the assumptions of Theorem 2.2 hold, then:

(a) $E(U) = \dfrac{(N+\alpha+\beta)X + (N-n)\alpha}{n+\alpha+\beta}$

(b) $V(U) = \dfrac{(N-n)(N+\alpha+\beta)(X+\alpha)(n+\beta-X)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}$ .

Proof:

(a) $$\sum_{Y=0}^{N-n} Y\binom{Y+X+\alpha-1}{Y}\binom{N+\beta-X-Y-1}{N-n-Y} =$$

$$= (X+\alpha)\sum_{Y=0}^{N-n-1}\binom{n+\beta-X+N-n-1-Y-1}{N-n-1-Y}\binom{X+\alpha+1+Y-1}{Y}$$

$$= (X+\alpha)\binom{N+\alpha+\beta-1}{N-n-1}.$$

The last equality follows by equation (2.5). From the above result and (2.8), we have

$$E(Y) = \frac{(X+\alpha)(N-n)}{n+\alpha+\beta}$$

and

$$E(U) = \frac{(N+\alpha+\beta)X + (N-n)\alpha}{n+\alpha+\beta} .$$

(b) In a manner similar to that used in (b) of Corollary 2.1, we obtain

$$\sum_{Y=0}^{N-n} Y(Y-1)\binom{X+\alpha+Y-1}{Y}\binom{N+\beta-X-Y-1}{N-n-Y} = (X+\alpha+1)(X+\alpha)\binom{N+\alpha+\beta-1}{N-n-2}.$$

From (2.8) and the relation

$$V(Y) = E[Y(Y-1)] + E(Y) - E^2(Y),$$

we obtain

$$V(U) = V(Y) = \frac{(N-n)(N+\alpha+\beta)(X+\alpha)(n+\beta-X)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)} .$$

It should be noted that (a) and (b) of Corollary 2.2 reduce to (a) and (b) of Corollary 2.1 in the limit as $\alpha \to 0$ and $\beta \to 0$.

## Exponential Super Population

In this section we assume the finite population $(U_1, U_2, \ldots, U_N)$ is a simple random sample from an exponential distribution with unknown parameter $\beta$. Our interest will be in estimating the finite population mean,

$$\overline{U} = \frac{1}{N}\sum_{1}^{N} U_i, \tag{2.9}$$

on the basis of a simple random sample of size $n < N$ from the finite population.

If we define

$$\bar{x} = \frac{1}{n} \sum_{1}^{n} x_i$$

and

$$\bar{y} = \frac{1}{N-n} \sum_{1}^{N-n} y_i,$$

then $\bar{U}$ can be expressed as

$$\bar{U} = \frac{1}{N} [(N-n)\bar{y} + n\bar{x}]. \tag{2.10}$$

For a point estimate of $\bar{U}$, we will use

$$E(\bar{U}) = \frac{1}{N} [(N-n)E(\bar{y}) + n\bar{x}]$$

where $E(\bar{y})$ is determined from the predictive density of $\bar{y}$.

<u>Theorem</u> 2.3. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is an exponential with unknown parameter $\beta$. Also, suppose the prior density on $\beta$ is a Jeffrey vague prior of the form

$$\pi(\beta) \propto \frac{1}{\beta}, \quad \beta > 0.$$

Then the predictive density of

$$T = \frac{N-n}{n\bar{x}} \bar{y} \tag{2.11}$$

is a Beta distribution of the second type with parameters $(N-n)$ and $n$.

Proof: Since $x_i$, $i = 1,2, \ldots, n$, has an exponential distribution, then

$$g(\bar{x}|n,n\beta) = \frac{n\beta}{\Gamma(n)} (n\beta\bar{x})^{n-1} e^{-n\beta\bar{x}}, \quad \bar{x} > 0. \tag{2.12}$$

Also, $\bar{x}$ is sufficient for $\beta$ so the posterior for $\beta$ is obtained from

$$P(\beta|n,n\bar{x}) \propto g(\bar{x}|n,n\beta)\pi(\beta)$$

as

$$P(\beta|n,n\bar{x}) = \frac{n\bar{x}}{\Gamma(n)} (n\bar{x}\beta)^{n-1} e^{-n\bar{x}\beta}, \quad \beta > 0.$$

Since $\bar{y}$ given $(N-n)$ and $\beta$ has a distribution of the form in (2.12), then the predictive density of $\bar{y}$ is

$$f(\bar{y}|N,n,\bar{x}) = \int_0^\infty \frac{(N-n)[(N-n)\bar{y}]^{N-n-1} (n\bar{x})^n \beta^{N-1}}{\Gamma(N-n)\Gamma(n)} \cdot \exp\{-[(N-n)\bar{y} + n\bar{x}]\beta\}d\beta.$$

The above gamma integral reduces to

$$f(\bar{y}|N,n,\bar{x}) =$$

$$= \frac{\Gamma(N)}{\Gamma(N-n)\Gamma(n)} \left(\frac{N-n}{n\bar{x}}\right) \left[\frac{(N-n)\bar{y}}{n\bar{x}}\right]^{N-n-1} \left\{1 + \frac{N-n}{n\bar{x}}\bar{y}\right\}^{-N}, \quad \bar{y} > 0.$$

Now, if we let

$$T = \frac{N-n}{n\bar{x}}\bar{y},$$

then the desired result,

$$f(T|N-n,n) = \frac{\Gamma(N)}{\Gamma(N-n)\Gamma(n)} T^{N-n-1}(1+T)^{-N}, \quad T > 0 \tag{2.13}$$

is obtained.

Corollary 2.3. If the assumptions of Theorem 2.3 hold, then:

(a) $E(\bar{U}) = \frac{N-1}{N} \frac{n\bar{x}}{n-1}$

(b) $V(\bar{U}) = \frac{N-n}{N} \frac{N-1}{N(n-2)} \left( \frac{n\bar{x}}{n-1} \right)^2.$

Proof:

(a) From equation (2.10)

$$E(\bar{U}) = \frac{1}{N} [(N-n)E(\bar{y}) + n\bar{x}],$$

but

$$E(\bar{y}) = \frac{n\bar{x}}{N-n} E(T) = \frac{n\bar{x}}{n-1}$$

follows by equations (2.11) and (2.13).

Hence,

$$E(\bar{U}) = \frac{N-1}{N} \frac{n\bar{x}}{n-1} .$$

(b) Also, from equation (2.10)

$$V(\bar{U}) = \left( \frac{N-n}{N} \right)^2 V(y).$$

From equations (2.11) and (2.13), we have

$$V(\bar{y}) = \left( \frac{n\bar{x}}{N-n} \right)^2 V(T) = \frac{(n\bar{x})^2}{N-n} \frac{N-1}{(n-1)^2(n-2)} .$$

Hence,

$$V(\bar{U}) = \frac{N-n}{N} \frac{N-1}{N(n-2)} \left( \frac{n\bar{x}}{n-1} \right)^2 .$$

Theorem 2.4. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is an exponential with unknown parameter $\beta$. Also, suppose the prior density on $\beta$ is a conjugate prior of the form

$$\pi(\beta|\alpha,\lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \beta^{\lambda-1} e^{-\alpha\beta}, \quad \beta > 0, \quad \alpha,\lambda > 0.$$

Then the predictive density of

$$T = \frac{N-n}{\alpha+nx} \bar{y} \tag{2.14}$$

is a Beta of the second type with parameters $(N-n)$ and $n+\lambda$.

Proof: The posterior density of $\beta$ is

$$P(\beta|n,\bar{x},\alpha,\lambda) = \frac{(\alpha+n\bar{x})^{n+\lambda}}{\Gamma(n+\lambda)} \beta^{n+\lambda-1} \exp\{-(\alpha+n\bar{x})\beta\}, \quad \beta > 0$$

which follows in a manner similar to that in proof of Theorem 2.3. Also, the distribution of $\bar{y}$ given $(N-n)$ and $\beta$ is of the form given in (2.12). Hence, the predictive density of $\bar{y}$ may be expressed as

$$f(\bar{y}|n,\bar{x},\alpha,\lambda) = \int_0^{\infty} \frac{(N-n)[(N-n)\bar{y}]^{N-n-1}[\alpha+n\bar{x}]^{n+\lambda}\beta^{N+\lambda-1}}{\Gamma(N-n)\Gamma(n+\lambda)} \cdot$$

$$\cdot \exp\{-[(N-n)\bar{y} + (n\bar{x}+\alpha)]\beta\}d\beta.$$

This integral reduces to

$$f(\bar{y}|n,\bar{x},\alpha,\lambda) =$$

$$= \frac{N-n}{\alpha+n\bar{x}} \left( \frac{(N-n)\bar{y}}{\alpha+n\bar{x}} \right)^{N-n-1} \frac{\Gamma(N+\lambda)}{\Gamma(N-n)\Gamma(n+\lambda)} \left\{ 1 + \frac{N-n}{\alpha+n\bar{x}} \bar{y} \right\}^{-(N+\lambda)}.$$

If we define $T$ as in (2.14), we obtain

$$f(T|n,\bar{x},\alpha,\lambda) = \frac{\Gamma(N+\lambda)}{\Gamma(N-n)\Gamma(n+\lambda)} T^{N-n-1} \{1+T\}^{-(N+\lambda)}, \quad T > 0, \tag{2.15}$$

the desired result.

<u>Corollary</u> 2.4. If the assumptions of Theorem 2.4 hold, then:

(a) $E(\bar{U}) = \frac{N-n}{N} \frac{\alpha}{n-1+\lambda} + \frac{N-1+\lambda}{N} \frac{n\bar{x}}{n-1+\lambda}$

(b) $V(\bar{U}) = \left( \frac{\alpha+n\bar{x}}{n-1+\lambda} \right)^2 \left( \frac{N-n}{N} \right) \left( \frac{N-1+\lambda}{N} \right) \left( \frac{1}{n-2+\lambda} \right).$

Proof: The proof follows from (2.10), (2.14), and (2.15) utilizing the same general procedure as in the proof of Corollary 2.3.

Note that as $\alpha \to 0$ and $\beta \to 0$, (a) and (b) of Corollary 2.4 becomes identical to (a) and (b) of Corollary 2.3.

## Normal Super Population

In this section we assume the finite population $(U_1, U_2, \ldots, U_N)$ is a simple random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Our interest will be in estimating

$$\bar{U} = \frac{1}{N} \sum_1^N U_i$$

and

$$S^2 = \frac{1}{N} \sum_1^N (U_i - \bar{U})^2$$

on the basis of a simple random sample of size $n < N$ from the finite population.

We define

$$\bar{x} = \sum_1^n x_i/n, \quad s_x^2 = \sum_1^n (x_i - \bar{x})^2/(n-1)$$

and

$$\bar{y} = \sum_1^{N-n} y_i/(N-n), \quad s_y^2 = \sum_1^{N-n-1} (y_i - \bar{y})^2/(N-n-1).$$

With this notation, $\bar{U}$ and $S^2$ can be expressed as

$$\bar{U} = \frac{1}{N} [(N-n)\bar{y} + n\bar{x}] \tag{2.16}$$

and

$$S^2 = \frac{1}{N} \left[ (N-n-1)s_y^2 + \frac{(N-n)n}{N} (\bar{y}-\bar{x})^2 + (n-1)s_x^2 \right]. \tag{2.17}$$

To determine point estimators for $\bar{U}$ and $S^2$, it will suffice to derive the predictive densities of $\bar{y}$ and $s_y^2$. We then utilize equations (2.16) and (2.17) to obtain $E(\bar{U})$ and $E(S^2)$ with respect to their predictive densities. Also, we will use equations (2.16) and

(2.17) to obtain $V(\bar{U})$ and $V(S^2)$ which are used as a measure of the precision of our predictions of $\bar{U}$ and $S^2$, respectively.

<u>Theorem</u> 2.5. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is normal with unknown mean $\mu$ and known variance $\sigma^2$. Also, suppose the prior density on $\mu$ is a Jeffrey vague prior of the form

$$\pi(\mu)d\mu \propto d\mu, \quad -\infty < \mu < \infty.$$

Then:

(a) The predictive density of $\bar{y}$ is normal with mean $\bar{x}$ and variance $N\sigma^2/(N-n)n$.

(b) The predictive density of $s_y^2$ is a gamma with parameters $(N-n-1)/2$ and $(N-n-1)/2\sigma^2$.

Proof:

(a) As is well-known, the posterior density of $\mu$ is

$$P(\mu|\bar{x},\sigma^2/n) = \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left\{-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right\}$$

and the density of $\bar{y}$ given $\mu$ and $\sigma^2$ is

$$g_1\left(\bar{y}|\mu, \frac{\sigma^2}{N-n}\right) = \sqrt{\frac{N-n}{2\pi\sigma^2}} \exp\left\{-\frac{N-n}{2\sigma^2}(\bar{y}-\mu)^2\right\}. \qquad (2.18)$$

Hence, the predictive density of $\bar{y}$ is

$$f_1(\bar{y}|\bar{x},\sigma^2) = \int_{-\infty}^{\infty} \sqrt{\frac{(N-n)n}{(2\pi\sigma^2)^2}} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}[n(\mu-\bar{x})^2 + (N-n)(\bar{y}-\mu)^2]\right\} d\mu .$$

Completing the square with respect to $\mu$ and integrating, we obtain

$$f_1(\bar{y}|\bar{x},\sigma^2) = \sqrt{\frac{(N-n)n}{2\pi\sigma^2}} \exp\left\{-\frac{(N-n)n}{2N\sigma^2}(\bar{y}-\bar{x})^2\right\}, \quad -\infty < \bar{y} < \infty, \tag{2.19}$$

the desired result.

(b) The distribution of $s_y^2$ given $\sigma^2$ and N-n is

$$g_1\left(s_y^2 \middle| \frac{N-n-1}{2}, \frac{N-n-1}{2\sigma^2}\right) = \frac{\left(\frac{N-n-1}{2\sigma^2}\right)^{\frac{N-n-1}{2}} (s_y^2)^{\frac{N-n-1}{2}-1}}{\Gamma\left(\frac{N-n-1}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{(N-n-1)s_y^2}{2\sigma^2}\right\}. \tag{2.20}$$

Now $\sigma^2$ and (N-n) are known; hence, (2.20) is the predictive density of $s_y^2$.

Corollary 2.5. If the assumptions of Theorem 2.5 hold, then:

(a) $E(\bar{U}) = \bar{x}$

(b) $V(\bar{U}) = \frac{N-n}{N}\frac{\sigma^2}{n}$

(c) $E(S^2) = \frac{N-n}{N}\sigma^2 + \frac{n-1}{N} s_x^2$

(d) $V(S^2) = 2(N-n)\left(\frac{\sigma^2}{N}\right)^2.$

Proof:

(a) By equations (2.16) and (2.19), we have

$$E(\overline{U}) = \frac{1}{N}\left[(N-n)\overline{x} + n\overline{x}\right],$$

which reduces to

$$E(\overline{U}) = \overline{x}.$$

(b) From equation (2.16),

$$V(\overline{U}) = \frac{N-n}{N}^2 V(\overline{y}),$$

and by (2.19),

$$V(\overline{U}) = \frac{N-n}{N}\frac{\sigma^2}{n}.$$

(c) From equation (2.17),

$$E(S^2) = \frac{1}{N}\left[(N-n-1)E(s_y^2) + \frac{(N-n)n}{N}E(\overline{y}-\overline{x})^2 + (n-1)s_x^2\right].$$

But

$$E(s_y^2) = \sigma^2$$

and

$$E(\overline{y}-\overline{x})^2 = \frac{N\sigma^2}{(N-n)n}$$

follows from (2.20) and (2.19). Hence,

$$E(S^2) = \frac{N-n}{N}\sigma^2 + \frac{n-1}{N}s_x^2.$$

(d) From equation (2.17), we have

$$V(S^2) = \left(\frac{N-n-1}{N}\right)^2 V(s_y^2) + \left\{\frac{(N-n)n}{N^2}\right\}^2 V(\overline{y}-\overline{x})^2.$$

But,

$$\frac{(\overline{y}-\overline{x})^2}{V(\overline{y})} \sim \chi^2(1),$$

hence

$$V(\overline{y}-\overline{x})^2 = 2V^2(\overline{y}).$$

From (2.20), we have

$$V(s_y^2) = \frac{2\sigma^4}{N-n-1}.$$

So, substituting and simplifying, we have

$$V(S^2) = 2(N-n)\left(\frac{\sigma^2}{N}\right)^2.$$

Theorem 2.6. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is normal with unknown mean $\mu$ and known variance $\sigma^2$. Also, suppose the prior density on $\mu$ is a conjugate prior of the form

$$\pi\left(\mu \mid \mu_0, \frac{\sigma^2}{n_0}\right) = \sqrt{\frac{n_0}{2\pi\sigma^2}}\exp\left\{-\frac{n_0}{2\sigma^2}(\mu-\mu_0)^2\right\}, \quad -\infty < \mu < \infty.$$

Then:

(a) The predictive density of $\bar{y}$ is normal with mean $n_0\mu_0 + n\bar{x}$ and variance $(N+n_0)\sigma^2/(N-n)(n+n_0)$.

(b) The predictive density of $s_y^2$ is a gamma with parameters $(N-n-1)/2$ and $(N-n-1)/2\sigma^2$.

Proof:

(a) LaValle [10], page 347, gives the posterior distribution of $\mu$ by

$$P\left(\mu|\mu_1, \frac{\sigma^2}{n_1}\right) = \sqrt{\frac{n_1}{2\pi\sigma^2}} \exp\left\{-\frac{n_1}{2\sigma^2}(\mu-\mu_1)^2\right\}, \quad -\infty < \mu < \infty \tag{2.21}$$

where

$$\mu_1 = (n_0+n)^{-1}(n\bar{x} + n_0\mu_0)$$

and

$$n_1 = n_0 + n.$$

From (2.18) and (2.21), we obtain the predictive density of $\bar{y}$ as

$$f_1(\bar{y}|\bar{x},\sigma^2) = E\left[g_1\left(\bar{y}|\mu, \frac{\sigma^2}{N-n}\right)\right]$$

$$= \int_{-\infty}^{\infty} \sqrt{\frac{(N-n)n_1}{(2\pi\sigma^2)^2}} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}[n_1(\mu-\mu_1)^2 + (N-n)(\bar{y}-\mu)^2]\right\} d\mu.$$

Completing the square with respect to $\mu$ and integrating, we obtain

$$f_1(\bar{y}|\bar{x},\sigma^2) = \sqrt{\frac{(N-n)(n+n_0)}{2(N+n_0)\pi\sigma^2}} \cdot$$

$$\cdot \exp\left\{ -\frac{(N-n)(n+n_0)}{2(N+n_0)\sigma^2} \left( \bar{y} - \frac{n_0\mu_0 + n\bar{x}}{n_0+n} \right)^2 \right\}, \quad -\infty < \bar{y} < \infty, \qquad (2.22)$$

the desired result.

(b) Follows exactly as in (b) of Theorem 2.5.

Corollary 2.6. If the assumptions of Theorem 2.6 hold, then:

(a) $E(\bar{U}) = \frac{n\bar{x} + n_0\mu_0}{n+n_0} + \frac{nn_0}{N(n+n_0)} (\bar{x}-\mu_0)$

(b) $V(\bar{U}) = \frac{N-n}{N} \frac{N+n_0}{N} \frac{\sigma^2}{n+n_0}$

(c) $E(S^2) = \left( \frac{N-n-1}{N} + \frac{n}{n+n_0} \frac{N+n_0}{N} \right) \sigma^2 + \frac{n-1}{N} s_x^2$

(d) $V(S^2) = 2\left( \frac{\sigma^2}{N} \right)^2 \left\{ (N-n-1) + \left[ \frac{N+n_0}{N} \frac{n}{n+n_0} \right]^2 \right\}$

Proof:

(a) From equation (2.16),

$$E(\bar{U}) = \frac{N-n}{N} E(\bar{y}) + \frac{n}{N} \bar{x}$$

and by (2.22), we obtain

$$E(\bar{U}) = \frac{N-n}{N} \frac{n\bar{x} + n_0\mu_0}{n+n_0} + \frac{n}{N} \bar{x}.$$

Simplifying, yields

$$E(\bar{U}) = \frac{n\bar{x} + n_0\mu_0}{n+n_0} + \frac{nn_0}{N(n+n_0)} (\bar{x}-\mu_0).$$

(b) The result follows from (2.16) and (2.22) as

$$V(\bar{U}) = \left( \frac{N-n}{N} \right)^2 V(\bar{y}) = \frac{N-n}{N} \frac{N+n_0}{N} \frac{\sigma^2}{n+n_0} .$$

(c) From equation (2.17), we have

$$E(S^2) = \frac{1}{N} \left[ (N-n-1)E(s_y^2) + \frac{(N-n)n}{N} V(\bar{y}) + (n-1)s_x^2 \right]$$

which simplifies to

$$E(S^2) = \left( \frac{N-n-1}{N} + \frac{n}{n+n_0} \frac{N+n_0}{N} \right) \sigma^2 + \frac{n-1}{N} s_x^2 ,$$

using (2.21) and (2.22).

(d) Again, from equation (2.17), we have

$$V(S^2) = \frac{1}{N^2} \left[ (N-n-1)^2 V(s_y^2) + \left[ \frac{(N-n)n}{N} \right]^2 V(\bar{y}-\bar{x})^2 \right].$$

From (d) of Corollary 2.5,

$$V(\bar{y}-\bar{x})^2 = 2V^2(\bar{y}) = 2\left[ \frac{N+n_0}{N-n} \frac{\sigma^2}{n+n_0} \right]^2,$$

the last equality follows from (2.22).

By (2.20), we obtain

$$V(s_y^2) = \frac{2\sigma^4}{(N-n-1)} .$$

Substituting and simplifying, yields

$$V(S^2) = 2\left( \frac{\sigma^2}{N} \right)^2 \left\{ (N-n-1) + \left[ \frac{N+n_0}{N} \frac{n}{n+n_0} \right]^2 \right\}.$$

Theorem 2.7. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is normal with unknown mean $\mu$ and unknown variance $\sigma^2$. Also, suppose the joint prior density on $\mu$ and $\sigma^2$ is a Jeffrey vague prior of the form

$$\pi(\mu,\sigma^2) \propto \frac{1}{\sigma^2} , \quad -\infty < \mu < \infty, \quad \sigma^2 > 0.$$

Then:

(a) The predictive density of $\bar{y}$ is a t-distribution with $(n-1)$ degrees of freedom, location parameter $\bar{x}$, and precision $n(N-n)/Ns_x^2$.

(b) The predictive density of $u = s_y^2/s_x^2$ is an F-distribution with $(N-n-1)$ and $(n-1)$ degrees of freedom.

Proof:

(a) The joint distribution of $\bar{x}$ and $s_x^2$ given $\mu$ and $\sigma^2$ is

$$g(\overline{x},s_x^2|\mu,\sigma^2) = \frac{\sqrt{\frac{n}{2\pi\sigma^2}}\left(\frac{(n-1)s_x^2}{2\sigma^2}\right)^{\frac{n-1}{2}}}{s_x^2\ \Gamma\left(\frac{n-1}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}\left[\ n(\overline{x}-\mu)^2 + (n-1)s_x^2\ \right]\right\}.$$

Since $(\overline{x},s_x^2)$ is sufficient for $(\mu,\sigma^2)$, the joint posterior of $\mu$ and $\sigma^2$ can be expressed as

$$P(\mu,\sigma^2|\overline{x},s_x^2) \propto g(\overline{x},s_x^2|\mu,\sigma^2)\pi(\mu,\sigma^2).$$

Integration of $g(\overline{x},s_x^2|\mu,\sigma^2)\pi(\mu,\sigma^2)d\mu d\sigma^2$ yields $s_x^2$; hence,

$$P(\mu,\sigma^2|\overline{x},s_x^2) = \frac{\sqrt{\frac{n}{2\pi\sigma^2}}\left(\frac{(n-1)s_x^2}{2\sigma^2}\right)^{\frac{n+1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\left(\frac{(n-1)s_x^2}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}\left[\ n(\mu-\overline{x})^2 + (n-1)s_x^2\ \right]\right\}. \qquad (2.23)$$

The predictive density of $\overline{y}$ is obtained from (2.18) and (2.23) as

$$f_1(\bar{y}|\bar{x},s_x^2) = \int_{-\infty}^{\infty}\int_0^{\infty} \frac{\sqrt{\frac{N-n}{2\pi\sigma^2}}\sqrt{\frac{n}{2\pi\sigma^2}}\left(\frac{(n-1)s_x^2}{2\sigma^2}\right)^{\frac{n+1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\left(\frac{(n-1)s_x^2}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}\left[\, n(\bar{x}-\mu)^2 + (N-n)(\bar{y}-\mu)^2 + (n-1)s_x^2 \,\right]\right\} d\mu d\sigma^2 .$$

Noting that

$$n(\bar{x}-\mu)^2 + (N-n)(\bar{y}-\mu)^2 = N\left(\mu - \frac{n\bar{x} + (N-n)\bar{y}}{N}\right)^2 + \frac{(N-n)n}{N}(\bar{y}-\bar{x})^2$$

and integrating over $\mu$, yields

$$f_1(\bar{y}|\bar{x},s_x^2) = \int_0^{\infty} \frac{\sqrt{\frac{N-n}{2\pi N}}\left(\frac{(n-1)s_x^2}{2\sigma^2}\right)^{\frac{n+2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\left(\frac{(n-1)s_x^2}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s_x^2 + \frac{(N-n)n}{N}(\bar{y}-\bar{x})^2\right]\right\} d\sigma^2 .$$

The integral is in the form

$$\int_0^{\infty}\left(\frac{1}{\sigma^2}\right)^k \exp\left\{-\frac{z}{\sigma^2}\right\} d\sigma^2 = (z)^{-(k-1)}\Gamma(k-1). \qquad (2.24)$$

Hence, we obtain

$$f_1(\bar{y}|\bar{x},s_x^2) = \frac{\sqrt{\frac{N-n}{2\pi N}}\left(\frac{(n-1)s_x^2}{2}\right)^{\frac{n-1}{2}}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left\{\frac{n-1}{2}\,s_x^2 + \frac{(N-n)n}{2N}\,(\bar{y}-\bar{x})^2\right\}^{-\frac{n}{2}}.$$

After some algebraic simplifications, we obtain

$$f_1(\bar{y}|\bar{x},s_x^2) = \sqrt{\frac{n(N-n)}{s_x^2 N(n-1)\pi}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left\{1 + \frac{(N-n)n}{Ns_x^2}\,\frac{(\bar{y}-\bar{x})^2}{n-1}\right\}^{-\frac{n}{2}}, \qquad (2.25)$$

the desired result.

(b) The predictive density of $s_y^2$ is obtained from (2.20) and (2.23) as

$$f_2(s_y^2|\bar{x},s_x^2) = \int_{-\infty}^{\infty}\int_0^{\infty} g_2(s_y^2|\mu,\sigma^2)P(\mu,\sigma^2|\bar{x},s_x^2)d\mu d\sigma^2.$$

Integrating over $\mu$, we obtain

$$f_2(s_y^2|\bar{x},s_x^2) = \int_0^{\infty} \frac{\left(\frac{(n-1)s_x^2}{2}\right)^{\frac{n-1}{2}}\left(\frac{(N-n-1)s_y^2}{2}\right)\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{N-n-1}{2}\right)s_y^2}\,\cdot$$

$$\cdot\ \exp\left\{-\frac{1}{2\sigma^2}\left[\ (N-n-1)s_y^2 + (n-1)s_x^2\ \right]\right\}d\sigma^2.$$

By equation (2.24), this reduces to

$$f_2(s_y^2|\bar{x},s_x^2) = \frac{\Gamma\left(\frac{N-2}{2}\right)\left(\frac{(n-1)s_x^2}{2}\right)^{\frac{n-1}{2}}\left(\frac{(N-n-1)s_y^2}{2}\right)^{\frac{N-n-1}{2}}}{s_y^2\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{N-n-1}{2}\right)} \cdot \left\{\frac{N-n-1}{2}s_y^2 + \frac{n-1}{2}s_x^2\right\}^{-\frac{N-2}{2}}$$

and after some algebraic simplification, we have

$$f_2(s_y^2|\bar{x},s_x^2) = \frac{\Gamma\left(\frac{N-2}{2}\right)\left(\frac{N-n-1}{n-1}\frac{s_y^2}{s_x^2}\right)^{\frac{N-n-1}{2}}}{s_y^2\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{N-n-1}{2}\right)}\left\{1 + \frac{N-n-1}{n-1}\frac{s_y^2}{s_x^2}\right\}^{-\frac{N-2}{2}} . \qquad (2.26)$$

Let $u = s_y^2/s_x^2$, then $u \sim F(N-n-1,n-1)$.

Corollary 2.7. If the assumptions of Theorem 2.7 hold, then:

(a) $E(\bar{U}) = \bar{x}$

(b) $V(\bar{U}) = \frac{N-n}{N}\frac{n-1}{n-3}\frac{s_x^2}{n}$, $n > 3$

(c) $E(S^2) = \frac{n-1}{n-3}\frac{N-3}{N}s_x^2$, $n > 3$

(d) $V(S^2) = \frac{2}{n-5}\left\{\frac{n-1}{n-3}\frac{s_x^2}{N}\right\}^2\left\{(N-n-1)(N-4) + (n-2)\right\}$, $n > 5$.

Proof: Let

$$t = \sqrt{\frac{(N-n)n}{Ns_x^2}}\,(\bar{y}-\bar{x})$$

in (2.25), then $t$ has a Student's t-distribution with $(n-1)$ degrees of freedom. Since

$$E(t) = 0 \quad \text{and} \quad V(t) = \frac{n-1}{n-3} ,$$

then

$$E(\bar{y}) = \bar{x} \tag{2.27}$$

and

$$V(\bar{y}) = \frac{n-1}{n-3} \frac{N s_x^2}{(N-n)N} . \tag{2.28}$$

Also, if we let $u = s_y^2/s_x^2$ in (2.26), then

$$E(u) = \frac{n-1}{n-3} \quad \text{and} \quad V(u) = \frac{2(n-1)^2(N-4)}{(N-n-1)(n-5)(n-3)^2} .$$

Therefore,

$$E(s_y^2) = \frac{n-1}{n-3} s_x^2 \tag{2.29}$$

and

$$V(s_y^2) = \frac{2(n-1)^2(N-4)}{(N-n-1)(n-5)(n-3)^2} s_x^4 . \tag{2.30}$$

(a) The result follows immediately from (2.16) and (2.27).

(b) From equation (2.16), we obtain

$$V(\bar{U}) = \left(\frac{N-n}{N}\right)^2 V(\bar{y}),$$

and from (2.28) it follows that

$$V(\bar{U}) = \frac{N-n}{N}\,\frac{n-1}{n-3}\,\frac{s_x^2}{n} .$$

(c) Equation (2.17) yields

$$E(S^2) = \frac{1}{N}\left[(N-n-1)E(s_y^2) + \frac{(N-n)n}{N}\,V(\bar{y}) + (n-1)s_x^2\right].$$

Substituting results from (2.29), (2.28), and simplifying, yields

$$E(S^2) = \frac{n-1}{n-3}\,\frac{N-3}{N}\,s_x^2 .$$

(d) We obtain

$$V(S^2) = \frac{1}{N^2}\left\{(N-n-1)^2 V(s_y^2) + \left[\frac{(N-n)n}{N}\right]^2 V(\bar{y}-\bar{x})^2\right\} \qquad (2.31)$$

from (2.17). Now

$$\begin{aligned} V(\bar{y}-\bar{x})^2 &= E(\bar{y}-\bar{x})^4 - V^2(\bar{y}) \\ &= \left[\frac{N\,s_x^2}{(N-n)n}\right]E(t^4) - \left[\frac{n-1}{n-3}\,\frac{N\,s_x^2}{(N-n)n}\right]^2 , \end{aligned}$$

where

$$t = \sqrt{\frac{(N-n)n}{N\,s_x^2}}\,(\bar{y}-\bar{x}) \sim t(n-1).$$

Since

$$E(t^4) = \frac{3(n-1)^2}{(n-3)(n-5)} ,$$

we then have

$$V(\bar{y}-\bar{x})^2 = 2\left\{\frac{N\ s_x^2}{(N-n)n}\ \frac{n-1}{n-3}\right\}^2 \left(\frac{n-2}{n-5}\right).$$

Substituting the above and (2.30) into (2.31), we obtain

$$V(S^2) = \frac{2}{n-5}\left\{\frac{n-1}{n-3}\ \frac{s_x^2}{N}\right\}^2 \left\{(N-n-1)(N-4) + (n-2)\right\}.$$

Theorem 2.8. Let $x_i$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from a finite population and suppose the super population distribution is normal with unknown mean $\mu$ and unknown variance $\sigma^2$. Also, suppose the joint prior density of $\mu$ and $\sigma^2$ is the Normal-inverted gamma density defined by

$$\pi(\mu,\sigma^2|\mu_0,\psi_0,n_0,\nu_0) = \sqrt{\frac{n_0}{2\pi\sigma^2}}\ \exp\left\{-\frac{n_0}{2\sigma^2}(\mu-\mu_0)^2\right\} \cdot \frac{\left(\frac{\nu_0\psi_0}{2\sigma^2}\right)^{\frac{\nu_0}{2}+1}}{\left(\frac{\nu_0\psi_0}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)} \exp\left\{-\frac{\nu_0\psi_0}{2\sigma^2}\right\}$$

for $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$, and zero otherwise, where

$$-\infty < \mu_0 < \infty, \quad n_0,\psi_0,\nu_0 > 0.$$

Then:

(a) The predictive density of $\bar{y}$ is a t-distribution with $(n+\nu_0)$ degrees of freedom, location parameter $(n\bar{x} + n_0\mu_0)/(n_0 + n)$, and precision $(N-n)(n+n_0)(n+\nu_0)/(N+\nu_0)\nu_1\psi_1$.

(b) The predictive density of $u = s_y^2/\psi_1$ is an F-distribution with $(N-n-1)$ and $\nu_1$ degrees of freedom, where

$$\psi_1 = \nu_1^{-1}\left\{ \frac{n\ n_0}{n+n_0}(\bar{x}-\mu_0)^2 + (n-1)s_x^2 + \nu_0\psi_0 \right\}$$

and

$$\nu_1 = n + \nu_0.$$

Proof: The proof follows the same general procedure as the proof of Theorem 2.7.

Corollary 2.8. If the assumptions of Theorem 2.8 hold, then:

(a) $$E(\bar{U}) = \frac{n\bar{x} + n_0\mu_0}{n + n_0} + \frac{n\ n_0}{N(n + n_0)}(\bar{x} - \mu_0)$$

(b) $$V(\bar{U}) = \frac{N - n}{N}\,\frac{N + n_0}{N}\,\frac{\nu_1\psi_1}{n_1(\nu_1 - 2)}$$

(c) $$E(S^2) = \frac{\nu_1\psi_1}{\nu_1 - 2}\left[(N-n-1) + \frac{n}{n + n_0}\,\frac{N + n_0}{N}\right] + \frac{n - 1}{N}s_x^2$$

(d) $$V(S^2) = \left(\frac{\nu_1\psi_1}{\nu_1 - 2}\right)^2\left(\frac{1}{\nu_1 - 4}\right)\left(\frac{2}{N^2}\right)\cdot \left\{(N+n_0-3) + (\nu_1-1)\left(\frac{n}{n+n_0}\right)^2\left(\frac{N+n_0}{N}\right)^2\right\}.$$

Proof: The results follow utilizing a procedure similar to the proof of Corollary 2.7.

## CHAPTER III

## REGRESSION AND RATIO ESTIMATORS

If there is an ordered pair, $(Y_i, X_i)$, of observable values attached to the $i^{th}$ unit in the finite population of $N$ units, we will consider the finite population as a vector in $(R^2)^N$, where $R$ is the real numbers. Also, we will represent the finite population by $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$.

In addition, we assume the finite population is a simple random sample from a joint probability distribution such that the conditional distribution of $Y$ given $X$ is

$$f(Y|X,\alpha',\beta,\sigma^2) = \tag{3.1}$$

$$= \sqrt{\frac{1}{2\pi\sigma^2X^{2g}}} \exp\left\{- \frac{1}{2\sigma^2X^{2g}} (Y - \alpha'X^g - \beta X^{g+1})^2\right\}, \quad -\infty < Y < \infty.$$

$-\infty < \alpha'$, $\beta < \infty$, $\sigma^2 > 0$, and $g = 0$ or $g = 1$. Hence, a simple random sample without replacement of size $n < N$ from the finite population is a simple random sample from the distribution (3.1). Based on this sample, we derive predictive estimators of

$$\overline{Y} = \frac{1}{N} \sum_{1}^{N} Y_i$$

for particular assumptions on the parameters in (3.1). For all of our derivations, we assume

$$\bar{X} = \frac{1}{N}\sum_{1}^{N} X_i \quad \text{and} \quad \sum_{1}^{N} X_i^2$$

are known. This knowledge does not imply that each individual $X_i$ is known.

The notation used in this chapter will be slightly different from that used in the preceeding chapter. We will let $(y_i, x_i)$, $i = 1,2, \ldots, n$, denote the observed values attached to the $i^{th}$ sampled unit in the simple random sample of size $n$ from the finite population. Also, $(u_i, v_i)$, $i = 1,2, \ldots, N-n$, will denote the unknown values attached to the $i^{th}$ unsampled unit remaining in the finite population.

The following terminology will be used in this chapter:

$$\bar{y} = \frac{1}{n}\sum_{1}^{n} y_i \qquad \bar{x} = \frac{1}{n}\sum_{1}^{n} x_i$$

$$s_y^2 = \sum_{1}^{n} (y_i - \bar{y})^2 \qquad s_x^2 = \sum_{1}^{n} (x_i - \bar{x})^2$$

$$\bar{u} = \frac{1}{N-n}\sum_{1}^{N-n} u_i \qquad \bar{v} = \frac{1}{N-n}\sum_{1}^{N-n} v_i$$

and

$$s_{xy} = \sum_{1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

Since we can express $\bar{Y}$ by

$$\bar{Y} = \frac{1}{N}\,[(N-n)\bar{u} + n\bar{y}], \tag{3.2}$$

it will suffice to derive predictive estimators of $\bar{u}$, or equivalently, of $T = (N-n)\bar{u}$ in order to predict $\bar{Y}$.

## Regression Estimators

If $g = 0$ in (3.1), then

$$E(Y|X) = \alpha' + \beta X$$

and

$$V(Y|X) = \sigma^2.$$

Hence, there is a linear regression of $Y$ on $X$ which we will utilize in predicting $\bar{Y}$.

Theorem 3.1. Let $(y_i, x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $g = 0$. Also, suppose that $\sigma^2$ is known and that the joint prior density on $\alpha$ and $\beta$ $(\alpha = \alpha' + \beta\bar{x})$ is a Jeffrey vague prior of the form

$$\pi(\alpha,\beta)d\alpha d\beta \propto d\alpha d\beta, \quad -\infty < \alpha, \quad \beta < \infty.$$

Then the predictive density of $\bar{u}$ is normal with mean

$$\bar{y} + \frac{N}{N-n}(\bar{X} - \bar{x})\frac{s_{xy}}{s_x^2}$$

and variance

$$\frac{N}{N-n}\frac{\sigma^2}{n}\left[1 + \frac{nN}{N-n}\frac{(\bar{X}-\bar{x})^2}{s_x^2}\right].$$

Proof: Lindley [11] on page 207 writes the joint posterior distribution of $\alpha$ and $\beta$ as

$$P(\alpha,\beta|\hat{\alpha},\hat{\beta}) = \frac{\sqrt{ns_x^2}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2]\right\}$$

where

$$\hat{\alpha} = \bar{y} \quad \text{and} \quad \hat{\beta} = \frac{s_{xy}}{s_x^2} .$$

The distribution of $\bar{u}$ given $\alpha$ and $\beta$ is normal with mean $\alpha + \beta(\bar{v}-\bar{x})$ and variance $\sigma^2/(N-n)$. Hence, the predictive density of $\bar{u}$ is

$$f(\bar{u}|\hat{\alpha},\hat{\beta}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{\frac{N-n}{2\pi\sigma^2}}\sqrt{\frac{ns_x^2}{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}[n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2 + (N-n)(\bar{u}-\alpha-\beta(\bar{v}-\bar{x}))^2]\right\} d\alpha d\beta. \tag{3.3}$$

The equality

$$n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2 + (N-n)(\bar{u}-\alpha-\beta(\bar{v}-\bar{x}))^2 =$$

$$= N(\alpha-\mu_\alpha)^2 + \left(s_x^2 + \frac{n(N-n)}{N}(\bar{v}-\bar{x})^2\right)(\beta-\mu_\beta)^2 + \frac{\frac{n(N-n)}{N}s_x^2}{s_x^2 + \frac{n(N-n)}{N}(\bar{v}-\bar{x})^2}(\bar{u}-\hat{\alpha}-\hat{\beta}(\bar{v}-\bar{x}))^2,$$

where

$$\mu_\alpha = \frac{n\hat{\alpha} + (N-n)\bar{u} - (N-n)(\bar{v}-\bar{x})\beta}{N}$$

and

$$\mu_\beta = \frac{\hat{\beta}s_x^2 + \frac{n(N-n)}{N}(\bar{v}-\bar{x})(\bar{u}-\hat{\alpha})}{s_x^2 + \frac{n(N-n)}{N}(\bar{v}-\bar{x})^2}$$

is obtained by expanding the left hand side and completing the square with respect to $\alpha$ and $\beta$. Substituting into (3.3) and integrating with respect to $\alpha$ and $\beta$, we obtain

$$f(\bar{u}|\hat{\alpha},\hat{\beta}) = \sqrt{\frac{1}{2\pi K\sigma^2}} \exp\left\{-\frac{1}{2K\sigma^2}(\bar{u}-\hat{\alpha}-\hat{\beta}(\bar{v}-\bar{x}))^2\right\}$$

where

$$K = \frac{N}{n(N-n)} + \frac{(\bar{v}-\bar{x})^2}{s_x^2}.$$

Since

$$\bar{v} = \frac{1}{N-n}(N\bar{X} - n\bar{x}),$$

then

$$\bar{v} - \bar{x} = \frac{N}{N-n}(\bar{X} - \bar{x}).$$

Therefore,

$$E(\bar{u}) = \hat{\alpha} + \hat{\beta}(\bar{v}-\bar{x}) = \bar{y} + \frac{N}{N-n}(\bar{X}-\bar{x})\hat{\beta} \tag{3.4}$$

and

$$V(\bar{u}) = \sigma^2\left\{\frac{N}{n(N-n)} + \frac{(\bar{v}-\bar{x})^2}{s_x^2}\right\} = \frac{N}{N-n}\frac{\sigma^2}{n}\left\{1 + \frac{nN}{N-n}\frac{(\bar{X}-\bar{x})^2}{s_x^2}\right\}. \tag{3.5}$$

<u>Corollary</u> 3.1. If the assumptions of Theorem 3.1 hold, then:

(a) $E(\bar{Y}) = \bar{y} + \hat{\beta}(\bar{X}-\bar{x})$

(b) $V(\bar{Y}) = \frac{(N-n)}{N}\frac{\sigma^2}{n}\left\{1 + \frac{n(N-n)}{N}\frac{(\bar{X}-\bar{x})^2}{s_x^2}\right\}$.

Proof:

(a) From (3.2), we have

$$E(\bar{Y}) = \frac{N-n}{N} E(\bar{u}) + \frac{n}{N}\bar{y}$$

and by (3.4), we obtain

$$E(\bar{Y}) = \bar{y} + \hat{\beta}(\bar{X}-\bar{x}).$$

(b) The result follows immediately from (3.2) and (3.5).

<u>Theorem</u> 3.2. Let $(y_i,x_i)$, $i = 1,2, \ldots, n,$ be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $\alpha'$ and $g$ both zero. Also, suppose $\sigma^2$ is known and the prior density on $\beta$ is a Jeffrey vague prior of the form

$$\pi(\beta)d\beta \propto d\beta, \quad -\infty < \beta < \infty.$$

Then the predictive density of $\bar{u}$ is normal with mean $\hat{\beta}\bar{v}$ and variance

$$\left\{\frac{1}{N-n} + \frac{\bar{v}^2}{n\bar{x}^2 + s_x^2}\right\}\sigma^2$$

where

$$\hat{\beta} = \frac{\sum_{1}^{n} x_i y_i}{n\bar{x}^2 + s_x^2} .$$

Proof: The proof follows in same manner as the proof of Theorem 3.1.

Corollary 3.2. If the assumptions of Theorem 3.2 hold, then:

(a) $E(\bar{Y}) = \frac{1}{N}\,[n\bar{y} + \hat{\beta}(N\bar{X}-n\bar{x})]$

(b) $V(\bar{Y}) = \left\{ \frac{N-n}{N} + \frac{(N\bar{X} - n\bar{x})^2}{N(n\bar{x}^2 + s_x^2)} \right\} \frac{\sigma^2}{N} .$

Proof: The proof is similar to the proof of Corollary 3.1.

Theorem 3.3. Let $(y_i,x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $g = 0$. Also, suppose the joint prior density on $\alpha,\beta$, and $\sigma^2$ $(\alpha = \alpha' + \beta\bar{x})$ is a Jeffrey vague prior of the form

$$\pi(\alpha,\beta,\sigma^2) \propto \frac{1}{\sigma^2} , \quad -\infty < \alpha, \quad \beta < \infty, \quad \sigma^2 > 0.$$

Then the predictive density of $\bar{u}$ is a t-distribution with $(n-2)$ degrees of freedom, location parameter $\hat{\alpha} + \hat{\beta}(\bar{v}-\bar{x})$, and precision

$$\left\{ \left[ \frac{N}{n(N-n)} + \frac{(\bar{v}-\bar{x})^2}{s_x^2} \right] S^2 \right\}^{-1}$$

where

$$\hat{\alpha} = y$$

$$\hat{\beta} = \frac{s_{xy}}{s_x^2}$$

and

$$S^2 = s_y^2 - \frac{s_{xy}^2}{s_x^2} .$$

Proof: Lindley [11] on page 205 writes the joint posterior distribution of $\alpha$, $\beta$, and $\sigma^2$ as

$$P(\alpha,\beta,\sigma^2|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = \sqrt{\frac{ns_x^2}{(2\pi\sigma^2)^2}}\ \frac{\left(\frac{S^2}{2\sigma^2}\right)^{\frac{n}{2}}}{\frac{S^2}{2}\ \Gamma\left(\frac{n-2}{2}\right)} \cdot \exp\left\{-\frac{1}{2\sigma^2}\left[S^2 + n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2\right]\right\}.$$

The distribution of $\bar{u}$ given $\alpha$, $\beta$, and $\sigma^2$ is normal with mean $\alpha + \beta(\bar{v}-\bar{x})$ and variance $\sigma^2/(N-n)$. Hence, the predictive density of $\bar{u}$ is

$$f(\bar{u}|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \sqrt{\frac{(N-n)ns_x^2}{(2\pi\sigma^2)^2}}\ \frac{\left(\frac{S^2}{2\sigma^2}\right)^{\frac{n}{2}}}{\frac{S^2}{2}\ \Gamma\left(\frac{n-2}{2}\right)} \cdot \exp\left\{-\frac{1}{2\sigma^2}\left[S^2 + (N-n)(\bar{u}-\alpha-\beta(\bar{v}-\bar{x}))^2 + n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2\right]\right\} d\alpha d\beta d\sigma^2. \qquad (3.6)$$

Expanding and completing the square on $\alpha$ and then on $\beta$, it can be shown that

$$(N-n)(\bar{u}-\alpha-\beta(\bar{v}-\bar{x}))^2 + n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2 =$$

$$= N(\alpha-\mu_\alpha)^2 + \left( s_x^2 + \frac{n(N-n)}{N} (\bar{v}-\bar{x})^2 \right)(\beta-\mu_\beta)^2$$

$$+ \frac{\frac{n(N-n)}{N} s_x^2 \; (\bar{u}-\hat{\alpha}-\hat{\beta}(\bar{v}-\bar{x}))^2}{\frac{n(N-n)}{N} (\bar{v}-\bar{x})^2 + s_x^2}$$

where

$$\mu_\alpha = \frac{(N-n)\bar{u} - (N-n)\beta(\bar{v}-\bar{x}) + n\hat{\alpha}}{N}$$

and

$$\mu_\beta = \frac{\frac{n(N-n)}{N} (\bar{v}-\bar{x})(\bar{u}-\hat{\alpha}) + s_x^2\hat{\beta}}{\frac{n(N-n)}{N} (\bar{v}-\bar{x})^2 + s_x^2} .$$

Hence, if we substitute into (3.6) and integrate with respect to $\alpha$ and $\beta$, we obtain an integral in the form of (2.24) which reduces to

$$f(\bar{u}|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = \sqrt{\frac{1}{\pi K S^2}} \, \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \left\{ 1 + \frac{(\bar{u}-\hat{\alpha}-\hat{\beta}(\bar{v}-\bar{x}))^2}{(n-2)KS^2} \right\}^{-\frac{n-1}{2}}$$

where

$$K = \frac{N}{n(N-n)} + \frac{(\bar{v}-\bar{x})^2}{s_x^2} .$$

Note that the above distribution reduces to a standardized t-distribution with $(n-2)$ degrees of freedom, if we let

$$t = \sqrt{\frac{1}{KS^2}}\,(\bar{u}-\hat{\alpha}-\hat{\beta}(\bar{v}-\bar{x})). \tag{3.7}$$

Corollary 3.3. If the assumptions of Theorem 3.3 hold, then:

(a) $$E(\bar{Y}) = \bar{y} + \hat{\beta}(\bar{X}-\bar{x}) \tag{3.8}$$

(b) $$V(\bar{Y}) = \frac{S^2}{n-4}\left[\frac{N-n}{nN} + \frac{(\bar{X}-\bar{x})^2}{s_x^2}\right], \quad n > 4, \tag{3.9}$$

where $\hat{\beta}$ and $S^2$ are as defined in Theorem 3.3.

Proof:

(a) From (3.2), we have

$$E(\bar{Y}) = \frac{N-n}{N}\,E(\bar{u}) + \frac{n}{N}\,\bar{y}.$$

But, by (3.7)

$$E(\bar{u}) = \hat{\alpha} + \hat{\beta}(\bar{v}-\bar{x}).$$

Hence, combining and simplifying, yields

$$E(\bar{Y}) = \bar{y} + \hat{\beta}(\bar{X}-\bar{x}).$$

(b) From (3.2) and (3.7), we obtain

$$V(\bar{Y}) = \left(\frac{N-n}{N}\right)^2 \frac{KS^2}{n-4}$$

which reduces to

$$V(\bar{Y}) = \frac{S^2}{(n-4)}\left[\frac{N-n}{nN} + \frac{(\bar{X}-\bar{x})^2}{s_x^2}\right].$$

We note that (3.8) is the usual least squares regression estimator for $\overline{Y}$. Also, we can write (3.9) as

$$V(\overline{Y}) = \frac{s_y^2}{n-4}(1-P^2)\left[\frac{N-n}{nN} + \frac{(\overline{X}-\overline{x})^2}{s_x^2}\right],$$

where

$$P^2 = \frac{s_{xy}^2}{s_x^2\, s_y^2}.$$

This result is similar to the variance of the least squares regression estimator as given by Cochran [12] page 194.

<u>Theorem</u> 3.4. Let $(y_i,x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $\alpha'$ and $g$ both zero. Also, suppose the joint prior density on $\beta$ and $\sigma^2$ is a Jeffrey vague prior of the form

$$\pi(\beta,\sigma^2) \propto \frac{1}{\sigma^2}, \quad -\infty < \beta < \infty, \quad \sigma^2 > 0.$$

Then the predictive density of $T = (N-n)\overline{u}$ is a t-distribution with $(n-1)$ degrees of freedom, location parameter $\hat{\beta}(N\overline{X}-n\overline{x})$, and precision

$$\frac{(n-1)(n\overline{x}^2 + s_x^2)}{[(N-n)(n\overline{x}^2+s_x^2) + S_1^2]\left[S^2 + \frac{ns_x^2}{n\overline{x}^2+s_x^2}\left(\overline{y} - \frac{s_{xy}}{s_x^2}\overline{x}\right)^2\right]}$$

where

$$\hat{\beta} = \frac{\sum_{1}^{n} x_i y_i}{n\bar{x}^2 + s_x^2}$$

and

$$S_1 = \sum_{1}^{N-n} v_i.$$

Proof: The proof follows in the same manner as the proof of Theorem 3.3.

Corollary 3.4. If the assumptions of Theorem 3.4 hold, then:

(a) $E(\bar{Y}) = \hat{\beta}\bar{X} + \frac{n}{N}(\bar{y} - \hat{\beta}\bar{x})$

(b) $V(\bar{Y}) = \frac{1}{N^2}\left\{\frac{N-n}{n-3} + \frac{(N\bar{X} - n\bar{x})^2}{(n-3)(n\bar{x}^2+s_x^2)}\right\}\left\{S^2 + \frac{ns_x^2}{n\bar{x}^2+s_x^2}\left(\bar{y} - \frac{s_{xy}}{s_x^2}\bar{x}\right)^2\right\}.$

Proof: The proof is similar to the proof of Corollary 3.3.

## Ratio Estimators

If $g = 1$ in (3.1), then

$$E(Y_i|X_i) = \alpha' X_i + \beta X_i^2$$

and

$$V(Y_i|X_i) = \sigma^2 X_i^2.$$

Now if $R_i' = Y_i/X_i$, then there is a linear regression of $\frac{Y}{X}$ on $X$ which will be utilized in predicting $\bar{Y}$.

In order to simplify the notation we encounter in this section, we define the following additional terms:

$$R_i = \frac{y_i}{x_i}, \qquad \bar{R} = \frac{1}{n}\sum_1^n \frac{y_i}{x_i},$$

$$S_1 = \sum_1^{N-n} v_i, \qquad S_2 = \sum_1^{N-n} v_i^2,$$

$$s_{Rx} = \sum_1^n (R_i - \bar{R})(x_i - \bar{x}) = n(\bar{y} - \bar{R}\bar{x}),$$

and

$$s_R^2 = \sum_1^n (R_i - \bar{R})^2.$$

<u>Theorem</u> 3.5. Let $(y_i,x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $g = 1$. Also, suppose that $\sigma^2$ is known and that the joint prior density on $\alpha$ and $\beta$ $(\alpha = \alpha' + \beta\bar{x})$ is a Jeffrey vague prior of the form

$$\pi(\alpha,\beta)d\alpha d\beta \propto d\alpha d\beta, \quad -\infty < \alpha, \quad \beta < \infty.$$

Then the predictive density of $T = (N-n)\bar{u}$ is normal with mean

$$S_1\bar{R} - \frac{s_{Rx}}{s_x^2}(S_2 - \bar{x}S_1)$$

and variance

$$\frac{(nS_2 + S_1^2)s_x^2 + n(S_2 - \bar{x}S_1)^2}{ns_x^2}\,\sigma^2.$$

Proof: Since $R_i = Y_i/X_i \sim N(\alpha + \beta(x_i - \bar{x}), \sigma^2)$, the joint posterior density of $\alpha$ and $\beta$ is

$$P(\alpha,\beta|\hat{\alpha},\hat{\beta}) = \frac{\sqrt{ns_x^2}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2\right]\right\}$$

where

$$\hat{\alpha} = \bar{R}, \qquad \hat{\beta} = \frac{s_{Rx}}{s_x^2} .$$

If we let

$$T = (N-n)\bar{u},$$

then the density of $T$ given $\alpha$ and $\beta$ is normal with mean $\alpha S_1 + \beta(S_2 - \bar{x}S_1)$ and variance $S_2\sigma^2$. Hence, the predictive density of $T$ is

$$f(T|\hat{\alpha},\hat{\beta}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\sqrt{ns_x^2}}{(2\pi\sigma^2)\sqrt{2\pi\sigma^2 S_2}} \cdot \exp\left\{-\frac{1}{2\sigma^2 S_2}\left[nS_2(\alpha-\hat{\alpha})^2 + s_x^2 S_2(\beta-\hat{\beta})^2 + (T-\alpha S_1-\beta(S_2-\bar{x}S_1))^2\right]\right\} d\alpha d\beta. \qquad (3.10)$$

The portion of the exponent of $e$ in brackets may be rewritten by

$$nS_2(\alpha-\hat{\alpha})^2 + s_x^2 S_2(\beta-\hat{\beta})^2 + (T-\alpha S_1-\beta(S_2-\bar{x}S_1))^2 =$$

$$= w(\alpha-\mu_\alpha)^2 + \frac{S_2K}{w}(\beta-\mu_\beta)^2 + \frac{nS_2s_x^2}{K}(T-S_1\hat{\alpha}-\hat{\beta}(S_2-\bar{x}S_1))^2$$

where

$$w = nS_2 + S_1^2 ,$$

$$K = ws_x^2 + n(S_2 - \bar{x}S_1)^2,$$

$$\mu_\alpha = \frac{nS_2\hat{\alpha} + TS_1 - \beta S_1(S_2 - \bar{x}S_1)}{nS_2 + S_1^2} ,$$

and

$$\mu_\beta = \frac{wS_2s_x^2\hat{\beta} + nS_2(S_2 - \bar{x}S_1)(T - S_1\hat{\alpha})}{wS_2s_x^2 + nS_2(S_2 - \bar{x}S_1)^2} .$$

Now integrating (3.10) with respect to $\alpha$ and $\beta$, yields

$$f(T|\hat{\alpha},\hat{\beta}) = \sqrt{\frac{ns_x^2}{2\pi K\sigma^2}} \exp\left\{-\frac{ns_x^2}{2K\sigma^2}(T - S_1\hat{\alpha} - \hat{\beta}(S_2 - \bar{x}S_1))^2\right\}.$$

Hence,

$$E(T) = S_1\hat{\alpha} + \hat{\beta}(S_2 - \bar{x}S_1) \tag{3.11}$$

and

$$V(T) = \frac{(nS_2 + S_1^2)s_x^2 + n(S_2 - \bar{x}S_1)^2}{ns_x^2}\sigma^2. \tag{3.12}$$

Corollary 3.5. If the assumptions of Theorem 3.5 hold, then:

(a) $$E(\bar{Y}) = \bar{X}\bar{R} + \frac{n}{N}(\bar{y} - \bar{x}\bar{R}) \left[ \frac{\sum_1^N X_i^2 - N\bar{x}\bar{X}}{s_x^2} \right]$$

(b) $$V(\bar{Y}) = \frac{\sigma^2}{N^2} \left[ \frac{\left( \sum_1^N X_i^2 - N\bar{x}\bar{X} \right)^2}{s_x^2} - s_x^2 + \frac{N(N-n)}{n} \bar{x}^2 \right],$$

where

$$s_x^2 = \sum_1^N (X_i - \bar{X})^2.$$

Proof:

(a) Since

$$\bar{Y} = \frac{1}{N} [T + n\bar{y}], \tag{3.13}$$

then by (3.11), we have

$$E(\bar{Y}) = \frac{1}{N} [S_1\hat{\alpha} + \hat{\beta}(S_2 - \bar{x}S_1) + n\bar{y}]. \tag{3.14}$$

Note that

$$\hat{\alpha} = \bar{R},$$

$$\hat{\beta} = \frac{n}{s_x^2} (\bar{y} - \bar{x}\bar{R}),$$

$$S_1 = N\bar{X} - n\bar{x},$$

and

$$S_2 = \sum_1^N X_i^2 - \sum_1^n x_i^2.$$

Substituting these into (3.14) and simplifying, yields

$$E(\overline{Y}) = \overline{R}\overline{X} + \frac{n}{N}(\overline{y} - \overline{R}\overline{x})\left[\frac{\sum_1^N X_i^2 - N\overline{x}\overline{X}}{s_x^2}\right]. \tag{3.15}$$

(b) By (3.13) and (3.12), we have

$$V(\overline{Y}) = \frac{\sigma^2}{N^2}\left\{S_2 + \frac{1}{n}S_1^2 + (S_2 - \overline{x}S_1)^2/s_x^2\right\}.$$

Using the identities

$$\left(\sum_{i=1}^N X_i^2 - N\overline{X}\overline{x}\right)^2 = (S_2 - \overline{x}S_1)^2 + 2s_x^2(S_2 - \overline{x}S_1) + s_x^4$$

and

$$\frac{1}{n}\left(\sum_{i=1}^N X_i\right)^2 = \frac{1}{n}S_1^2 + 2\overline{x}S_1 + n\overline{x}^2,$$

the equality reduces to

$$V(\overline{Y}) = \frac{\sigma^2}{N^2}\left\{\frac{\left(\sum_1^N X_i^2 - N\overline{x}\overline{X}\right)^2}{s_x^2} - s_x^2 + \frac{N(N-n)}{n}\overline{X}^2\right\}. \tag{3.16}$$

We remark that to use these formulas, it is necessary to know $\sum_1^N X_i^2$ which we did not require for the regression estimators.

Note that if

$$\bar{x} \doteq \bar{X}$$

and

$$\frac{S_X^2}{N-1} \doteq \frac{s_x^2}{n-1}$$

then (3.15) reduces to

$$E(\bar{Y}) = \bar{X}\bar{R} + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{x}\bar{R}),$$

the classical Hartley-Ross ratio estimator. Also, under these conditions (3.16) reduces to

$$V(\bar{Y}) = \frac{\sigma^2}{N^2}\left\{\frac{N-n}{n-1} S_X^2 + \frac{N(N-n)}{n} \bar{X}^2\right\}.$$

We now state two theorems and two corollaries whose proofs are similar to the proofs of Theorem 3.5 and Corollary 3.5, respectively.

<u>Theorem</u> 3.6. Let $(y_i, x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $\alpha' = 0$ and $g = 1$. Also, suppose $\sigma^2$ is known and the prior density on $\beta$ is a Jeffrey vague prior of the form

$$\pi(\beta)d\beta \propto d\beta, \quad -\infty < \beta < \infty.$$

Then the predictive density of $T = (N-n)\bar{u}$ is normal with mean $\hat{\beta}S_2$ and variance

$$\frac{S_2 \sum_1^N X_i^2}{n\bar{x}^2 + s_x^2} \sigma^2$$

where

$$\hat{\beta} = \frac{n\bar{y}}{n\bar{x}^2 + s_x^2} .$$

Corollary 3.6. If the assumptions of Theorem 3.6 hold, then:

(a) $E(\bar{Y}) = \frac{1}{N} \hat{\beta} \sum_1^N X_i^2$

(b) $V(\bar{Y}) = \dfrac{\left( \sum_1^N X_i^2 \right)^2 - \sum_1^n x_i^2 \sum_1^N X_i^2}{N^2(n\bar{x}^2 + s_x^2)} \sigma^2 .$

Theorem 3.7. Let $(y_i, x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $\beta = 0$ and $g = 1$. Also, suppose that $\sigma^2$ is known and that the prior density on $\alpha'$ is a Jeffrey vague prior of the form

$$\pi(\alpha')d\alpha' \propto d\alpha', \quad -\infty < \alpha' < \infty.$$

Then the predictive density of $T = (N-n)\bar{u}$ is normal with mean

$$\bar{R}(N\bar{X} - n\bar{x})$$

and variance

$$\left\{ n(N-n) + (N\bar{X} - n\bar{x})^2 \right\} \frac{\sigma^2}{n} .$$

Corollary 3.7. If the assumptions of Theorem 3.7 hold, then:

(a) $E(\bar{Y}) = \bar{R}\bar{X} + \frac{n}{N}(\bar{y} - \bar{R}\bar{x})$

(b) $V(\bar{Y}) = \{n(N-n) + (N\bar{X} - n\bar{x})^2\} \frac{\sigma^2}{nN^2}$ .

We now consider the case when the variance of the super population distribution is unknown.

Theorem 3.8. Let $(y_i, x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $g = 1$. Also, suppose that the joint prior density on $\alpha$, $\beta$, and $\sigma^2$ $(\alpha = \alpha' + \beta\bar{x})$ is a Jeffrey vague prior of the form

$$\pi(\alpha,\beta,\sigma^2) \propto \frac{1}{\sigma^2}, \quad -\infty < \alpha, \quad \beta < \infty, \quad \sigma^2 > 0.$$

Then the predictive density of $T = (N-n)\bar{u}$ is a t-distribution with $(n-2)$ degrees of freedom, location parameter

$$S_1\bar{R} + \hat{\beta}(S_2 - \bar{x}S_1)$$

and precision

$$\frac{n(n-2)S_2 s_x^2}{s^2[(nS_2 + S_1^2)s_x^2 + nS_2(S_2 - \bar{x}S_1)^2]}$$

where

$$\hat{\beta} = \frac{s_{Rx}}{s_x^2}$$

and

$$S^2 = s_R^2 - \frac{s_{Rx}^2}{s_x^2} .$$

Proof: Note that

$$R_i = \frac{y_i}{x_i} \sim N(\alpha + \beta(x_i - \bar{x}), \sigma^2)$$

and proceeding in a manner similar to that in the proof of Theorem 3.3, we obtain the joint posterior of $\alpha$, $\beta$, and $\sigma^2$ as

$$P(\alpha,\beta,\sigma^2|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = \sqrt{\frac{ns_x^2}{(2\pi\sigma^2)^2}} \frac{\left(\frac{S^2}{2\sigma^2}\right)^{\frac{n}{2}}}{\left(\frac{S^2}{2}\right) \Gamma\left(\frac{n-2}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2} [S^2 + n(\alpha-\hat{\alpha})^2 + s_x^2(\beta-\hat{\beta})^2]\right\}.$$

If we let

$$T = (N-n)\bar{u},$$

then

$$g(T|\alpha,\beta,\sigma^2) = \sqrt{\frac{1}{2\pi\sigma^2 S_2}} \exp\left\{-\frac{1}{2\sigma^2 S_2} (T-\alpha S_1-\beta(S_2-\bar{x}S_1))^2\right\}.$$

Hence, the predictive density of $T$,

$$f(T|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = E[g(T|\alpha,\beta,\sigma^2)],$$

can be expressed as

$$f(T|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \sqrt{\frac{ns_x^2}{(2\pi\sigma^2)^3 S_2}} \; \frac{\left(\frac{S^2}{2\sigma^2}\right)^{\frac{n}{2}}}{\left(\frac{S^2}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2 S_2}\left[S_2 S^2 + nS_2(\alpha-\hat{\alpha})^2 + S_2 s_x^2(\beta-\hat{\beta})^2 + (T-\alpha S_1-\beta(S_2-\bar{x}S_1))^2\right]\right\} d\alpha d\beta d\sigma^2 .$$

By completing the square on $\alpha$ and $\beta$, the portion in brackets of the exponent of $e$ can be written as

$$S_2S^2 + nS_2(\alpha-\hat{\alpha})^2 + S_2s_x^2(\beta-\hat{\beta})^2 + (T-\alpha S_1-\beta(S_2-\bar{x}S_1))^2 =$$

$$= S_2S^2 + W(\alpha-\mu_\alpha)^2 + \frac{K}{W}(\beta-\mu_\beta)^2 + \frac{nS_2^2 s_x^2}{K}(T-\hat{\alpha}S_1-\hat{\beta}(S_2-\bar{x}S_1))^2$$

where

$$W = nS_2 + S_1^2 ,$$

$$K = WS_2s_x^2 + nS_2(S_2 - \bar{x}S_1)^2,$$

$$\mu_\alpha = \frac{nS_2\hat{\alpha} + S_1T - \beta S_1(S_2 - \bar{x}S_1)}{nS_2 + S_1^2} ,$$

and

$$\mu_\beta = \frac{WS_2 s_x^2 \hat{\beta} + nS_2(S_2 - \bar{x}S_1)(T - S_1\hat{\alpha})}{K} .$$

Now integration with respect to $\alpha$, $\beta$, and $\sigma^2$ yields

$$f(T|\hat{\alpha},\hat{\beta},\hat{\sigma}^2) = \sqrt{\frac{nS_2 s_x^2}{\pi S^2 K}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \cdot \tag{3.17}$$

$$\cdot \left\{ 1 + \frac{n(n-2)S_2 s_x^2}{S^2 K} \frac{(T-\hat{\alpha}S_1-\hat{\beta}(S_2-\bar{x}S_1))^2}{n-2} \right\}^{-\frac{n-1}{2}} .$$

Corollary 3.8. If the assumptions of Theorem 3.8 hold, then:

(a) $E(\bar{Y}) = \bar{R}\bar{X} + \frac{n}{N}(\bar{y} - \bar{x}\bar{R}) \frac{\sum_1^N X_i^2 - N\bar{x}\bar{X}}{s_x^2}$

(b) $V(\bar{Y}) = \frac{S^2}{N^2(n-4)} \quad \frac{1}{n}\left(\sum_1^N X_i\right)^2 - \sum_1^N X_i^2 + \left(\sum_1^N X_i^2 - N\bar{x}\bar{X}\right)^2 / s_x^2$ .

Proof:

(a) It follows from

$$\bar{Y} = \frac{1}{N}[T + n\bar{y}] \tag{3.18}$$

and (3.17), that

$$E(\bar{Y}) = \frac{1}{N}[\hat{\alpha}S_1 + \hat{\beta}(S_2 - \bar{x}S_1) + n\bar{y}].$$

Recall that

$$\hat{\alpha} = \bar{R},$$

$$\hat{\beta} = \frac{s_{Rx}}{s_x^2} = \frac{n}{s_x^2}\,(\bar{y} - \bar{R}\bar{x}),$$

$$S_1 = N\bar{X} - n\bar{x},$$

and

$$S_2 = \sum_1^N X_i^2 - \sum_1^n x_i^2 .$$

Hence,

$$E(\bar{Y}) = \frac{1}{N}\left\{ N\bar{R}\bar{X} + n(\bar{y}-\bar{R}\bar{x}) + \frac{n}{s_x^2}\,(\bar{y}-\bar{R}\bar{x})\left[\sum_1^N X_i^2 - N\bar{x}\bar{X} - s_x^2\right]\right\} = \tag{3.19}$$

$$= \bar{R}\bar{X} + \frac{n}{N}\,(\bar{y}-\bar{R}\bar{x})\,\frac{\sum_1^N X_i^2 - N\bar{x}\bar{X}}{s_x^2} .$$

(b) From (3.18) and (3.17), we obtain

$$V(Y) = \frac{1}{N^2}\,\frac{S^2 K}{n(n-4)\,S_2 s_x^2}$$

$$= \frac{S^2}{N^2(n-4)}\left\{ S_2 + \frac{S_1^2}{n} + \frac{(S_2 - \bar{x}S_1)^2}{s_x^2}\right\}.$$

Using substitutions similar to that in part (a), we have

$$V(\bar{Y}) = \frac{S^2}{N^2(n-4)}\left\{ \frac{1}{n}\left(\sum_1^N X_i\right)^2 - \sum_1^N X_i^2 + \frac{1}{s_x^2}\left(\sum_1^N X_i^2 - N\bar{x}\bar{X}\right)^2\right\}. \tag{3.20}$$

We note that if

$$\bar{x} \doteq \bar{X}$$

and

$$\frac{1}{N-1} \sum_{1}^{N} (X_i - \bar{X})^2 \doteq \frac{1}{n-1} \sum_{1}^{n} (x_i - \bar{x})^2$$

then (3.19) reduces to

$$E(\bar{Y}) = \bar{R}\bar{X} + \frac{n(N-1)}{N(n-1)} (\bar{y} - \bar{R}\bar{x}),$$

the well-known Hartley-Ross ratio estimator. Also, it is easy to verify that (3.20) reduces to

$$V(\bar{Y}) = \frac{S^2}{N^2(n-4)} \left\{ \frac{N-n}{n-1} \left[ \sum_{1}^{N} X_i^2 - N\bar{X}^2 \right] + \frac{N(N-n)}{n} \bar{X}^2 \right\}.$$

We will now state two theorems and two corollaries whose proofs are omitted, but note that the proofs are similar to the proof of Theorem 3.8 and the proof of Corollary 3.8, respectively.

Theorem 3.9. Let $(y_i, x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $\alpha' = 0$ and $g = 1$. Also, suppose the joint prior density on $\beta$ and $\sigma^2$ is a Jeffrey vague prior of the form

$$\pi(\beta,\sigma^2) \propto \frac{1}{\sigma^2}, \quad -\infty < \beta < \infty, \quad \sigma^2 > 0.$$

Then the predictive density of $T = (N-n)\bar{u}$ is a t-distribution with $(n-1)$ degrees of freedom, location parameter

$$\hat{\beta} \sum_{1}^{N} X_i^2 - n\bar{y},$$

and precision

$$\left\{ \frac{[(N-n)w + S_2^2] \left[ S^2 + \frac{ns_x^2}{w} (\bar{y} - \hat{\beta}\bar{x})^2 \right]}{(n-1)w} \right\}^{-1},$$

where

$$w = n\bar{x}^2 + s_x^2 ,$$

$$\hat{\beta} = \frac{n\bar{y}}{w} ,$$

and

$$S^2 = s_R^2 - \frac{s_{Rx}^2}{s_x^2} .$$

Corollary 3.9. If the assumptions of Theorem 3.9 hold, then:

(a) $E(\bar{Y}) = \hat{\beta} \sum_{1}^{N} X_i^2$

(b) $V(\bar{Y}) = \frac{(N-n)w + S_2^2}{(n-3)w} \left\{ S^2 + \frac{ns_x^2}{w} (\bar{y} - \hat{\beta}\bar{x})^2 \right\}.$

Theorem 3.10. Let $(y_i, x_i)$, $i = 1,2, \ldots, n$, be a simple random sample of size $n$ from the finite population $((Y_1,X_1),(Y_2,X_2), \ldots, (Y_N,X_N))$ and suppose the super population distribution is given by (3.1) with $\beta = 0$ and $g = 1$. Also, suppose the joint prior density on $\alpha'$ and $\sigma^2$ is a Jeffrey vague prior of the form

$$\pi(\alpha', \sigma^2) \propto \frac{1}{\sigma^2}, \quad -\infty < \alpha' < \infty, \quad \sigma^2 > 0.$$

Then the predictive density of $T = (N-n)\bar{u}$ is a t-distribution with $(n-1)$ degrees of freedom, location parameter

$$\bar{R}(N\bar{X} - n\bar{x})$$

and precision

$$\frac{n(n-1)}{(nS_2 + S_1^2)s_R^2}.$$

<u>Corollary</u> 3.10. If the assumptions of Theorem 3.10 hold, then:

(a) $E(\bar{Y}) = \bar{R}\bar{X} + \frac{n}{N}(\bar{y} - \bar{R}\bar{x})$

(b) $V(\bar{Y}) = \frac{s_R^2}{n(n-3)N^2}\{nS_2 + S_1^2\}.$

## CHAPTER IV

## STRATIFIED RANDOM SAMPLING

In this chapter we introduce a stratification concept and some notation associated with stratification which will be utilized in the next chapter. Also, we will derive predictive estimators for some of the more interesting cases of stratification.

### Concept of Stratification

Suppose the finite population of interest can be partitioned into $k$ subsets or strata. Let $N_h$ be the number of units in stratum $h$ and $U_{hi}$ be the variate value attached to the $i^{th}$ unit in stratum $h$, $i = 1,2, \ldots, N_h$; $h = 1,2, \ldots, k$. Each stratum of the finite population may be considered as a vector in $R^{N_h}$, where $R$ is the real numbers. A stratum may then be sampled by randomly obtaining $n_h$ integers from the label set, $\{1,2, \ldots, N_h\}$ and observing the values,

$$U_{hi_1}, U_{hi_2}, \ldots, U_{hi_{n_h}},$$

attached to the units. For simplicity, we will denote the observed values by $x_{h1}, x_{h2}, \ldots, x_{hn_h}$ and the unobserved values remaining in the population by $y_{h1}, y_{h2}, \ldots, y_{ht_h}$, where $t_h = N_h - n_h$. Based on a sample of size $n_h$ from each stratum, we wish to make an inference not only about the stratum mean,

$$\overline{Y}_h = \frac{1}{N_h} \sum_{1}^{N_h} U_{hi},$$

but also about the overall mean,

$$\overline{Y} = \frac{1}{N} \sum_{1}^{k} N_h \overline{Y}_h,$$

where

$$N = \sum_{1}^{k} N_h .$$

As an extension of Cochran's [3], [4] suggestion, suppose that the value, $U_{hi}$, attached to the $i^{th}$ unit in stratum $h$ is the realization from a super distribution, say $f_h(U_h | \theta_h)$, $h = 1,2, \ldots, k$. That is, the finite population, $(U_{h1}, U_{h2}, \ldots, U_{hN_h})$, in stratum $h$ is the result of a random physical process described by a probability distribution. For example, suppose the finite populations of heights, weights, or intelligence is stratified on the basis of race, then we can consider each stratum as large random samples from a normal distribution defined by genetical mechanisms peculiar to that race.

Under this concept, a simple random sample from a stratum is also a simple random sample from the super distribution giving rise to that stratum. Now the $h^{th}$ stratum mean can be predicted from the $h^{th}$ predictive density and the overall mean is predicted by

$$E(\overline{Y}) = \frac{1}{N} \sum_{1}^{k} N_h E(\overline{Y}_h)$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_{1}^{k} N_h^2 \, V(\bar{Y}_h).$$

Three features of this concept should be noted. First, it is not necessary for the super distributions to be members of the same general class of distributions. Second, the prior distribution for stratum $i$ and the prior distribution for stratum $j$, $i \neq j$, are not required to be members of the same general class of distributions. Third, it is not appropriate to group the $k$ strata into one population and obtain a simple random sample from this one population in order to estimate the overall mean. Hence, we do not consider the problem of increased precision by stratification.

## Stratified Sampling

In this section we assume the finite population can be stratified as previously discussed, and we will derive a predictive estimator for the overall mean,

$$\bar{Y} = \frac{1}{N} \sum_{1}^{k} N_h \, \bar{Y}_h,$$

for some of the more interesting super populations. Although it is not necessary, we will require the super distribution and the prior distribution for each stratum to belong to the same general classes of distributions, respectively. Also, we will omit defining notation which is an obvious extension of the notation used in Chapters II and III to stratification.

First, suppose the super distribution for each stratum belongs to the class of Bernoulli distributions. If we assume a Jeffrey vague prior distribution on the parameter, $p_h$, in each stratum, it follows by Corollary 2.1 that the predictive estimator for the proportion of successes, $\bar{Y}$, is

$$E(\bar{Y}) = \frac{1}{N} \sum_{1}^{k} N_h \frac{x_h}{n_h} \tag{4.1}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_{1}^{k} \frac{N_h(N_h - n_h)}{n_h + 1} \frac{x_h}{n_h} \left( 1 - \frac{x_h}{n_h} \right). \tag{4.2}$$

Now assume the super distribution for each stratum belongs to the class of normal distributions whose variance is known for each stratum. If we assume a Jeffrey vague prior distribution for the parameter, $\mu_h$, in each stratum, then by Corollary 2.5 the predictive estimator for the overall mean, $\bar{Y}$, is

$$E(\bar{Y}) = \frac{1}{N} \sum_{1}^{k} N_h \bar{x}_h \tag{4.3}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_{1}^{k} N_h(N_h - n_h) \frac{\sigma_h^2}{n_h} . \tag{4.4}$$

Also, if the prior distribution for the parameter, $\mu_h$, in each stratum belongs to the class of normal distributions, that is

$$\mu_h \sim N\left( \mu_{0h}, \frac{\sigma_h^2}{n_{0h}} \right),$$

then by Corollary 2.6 the predictive estimator for the overall mean, $\bar{Y}$, is

$$E(\bar{Y}) = \frac{1}{N}\sum_{1}^{k}\left[ N_h \frac{n_h\bar{x}_h + n_{0h}\mu_{0h}}{n_h + n_{0h}} + \frac{n_h \, n_{0h}}{n_h + n_{0h}} (\bar{x}_h - \mu_{0h}) \right] \tag{4.5}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2}\sum_{1}^{k} (N_h - n_h)(N_h + n_{0h}) \frac{\sigma_h^2}{n_h + n_{0h}} . \tag{4.6}$$

Again, assume the super population for each stratum belongs to the class of normal distributions but whose variance is unknown for each stratum. Now if a Jeffrey vague prior distribution is assumed for the parameters, $\mu_h$ and $\sigma_h^2$, in each stratum, then the predictive estimator for the overall mean, $\bar{Y}$, is obtained from Corollary 2.7 as

$$E(\bar{Y}) = \frac{1}{N}\sum_{1}^{k} N_h \, \bar{x}_h \tag{4.7}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2}\sum_{1}^{k} N_h(N_h - n_h) \frac{n_h-1}{n_h-3} \frac{s_{hx}^2}{n_h} . \tag{4.8}$$

Now, if it is appropriate to assume a normal-inverted gamma distribution as the prior distribution for the parameters, $\mu_h$ and $\sigma_h^2$, in each stratum, then it follows by Corollary 2.8 that the predictive estimator of the overall mean, $\bar{Y}$, is

$$E(\bar{Y}) = \frac{1}{N}\sum_{1}^{k}\left[ N_h \frac{n_h\bar{x}_h + n_{0h}\mu_{0h}}{n_h + n_{0h}} + \frac{n_h \, n_{0h}}{n_h + n_{0h}} (\bar{x}_h - \mu_{0h}) \right] \tag{4.9}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_1^k (N_h - n_h)(N_h + n_{0h}) \frac{\nu_{1h} \psi_{1h}}{n_{1h}(\nu_{1h} - 2)} . \tag{4.10}$$

Now suppose the super distribution for each stratum belongs to the class of distributions defined by (3.1) with $g_h = 0$ and $\sigma_h^2$ known for all $h$. If we assume a Jeffrey vague prior distribution for the parameters, $\alpha_h$ and $\beta_h$, in each stratum, then by Corollary 3.1 the predictive estimator of the overall mean, $\bar{Y}$, is

$$E(\bar{Y}) = \frac{1}{N} \sum_1^k N_h \{\bar{y}_h + \hat{\beta}_h(\bar{X}_h - \bar{x}_h)\} \tag{4.11}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_1^k N_h(N_h - n_h) \frac{\sigma_h^2}{n_h} \left\{ 1 + \frac{n_h(N_h - n_h)(\bar{X}_h - \bar{x}_h)^2}{N_h \, s_{hx}^2} \right\}. \tag{4.12}$$

If $\sigma_h^2$ is unknown for all $h$ and we assume a Jeffrey vague prior distribution for the parameters, $\alpha_h$, $\beta_h$, and $\sigma_h^2$, in each stratum, then by Corollary 3.3 the predictive estimator of the overall mean, $\bar{Y}$, is given by (4.11) with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_1^k \frac{N_h^2 \, S_h^2}{n_h - 4} \left\{ \frac{N_h - n_h}{n_h \, N_h} + \frac{(\bar{X}_h - \bar{x}_h)^2}{s_{hx}^2} \right\}. \tag{4.13}$$

Again, assume the super distribution for each stratum belongs to the class of distributions defined by (3.1) but with $g_h = 1$ and $\sigma_h^2$ known for all $h$. If we assume a Jeffrey vague prior distribution for the parameters, $\alpha_h$ and $\beta_h$, in each stratum, then by Corollary 3.5 the predictive estimator for the overall mean, $\bar{Y}$, is

$$E(\bar{Y}) = \frac{1}{N} \sum_1^k N_h \left\{ \bar{X}_h \bar{R}_h + \frac{n_h}{N_h} (\bar{y}_h - \bar{x}_h \bar{R}_h) \frac{\sum_1^{N_h} X_{hi}^2 - N_h \bar{x}_h \bar{X}_h}{s_{hx}^2} \right\} \tag{4.14}$$

with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_1^k \sigma_h^2 \left\{ \frac{\left( \sum_1^{N_h} X_{hi}^2 - N_h \bar{x}_h \bar{X}_h \right)^2}{s_{hx}^2} - S_{hX}^2 + \frac{N_h (N_h - n_h)}{n_h} \bar{X}_h^2 \right\}. \tag{4.15}$$

If we assume $\sigma_h^2$ is unknown for all $h$ and retain the other previous assumptions, then by Corollary 3.8 the predictive estimator of the overall mean, $\bar{Y}$, is given by (4.14), but with precision

$$V(\bar{Y}) = \frac{1}{N^2} \sum_1^k \frac{S_h^2}{n_h - 4} \left\{ \frac{1}{n_h} \left( \sum_1^{N_h} X_{hi} \right)^2 - \sum_1^{N_h} X_{hi}^2 + \left( \sum_1^{N_h} X_{hi}^2 - N_h \bar{x}_h \bar{X}_h \right)^2 \Big/ s_{hx}^2 \right\}. \tag{4.16}$$

## CHAPTER V

## OPTIMUM ALLOCATION

In this chapter we assume the finite population may be stratified into k strata as discussed in Chapter IV and that we are interested in predicting a linear combination of the stratum mean, say

$$T = \sum_{1}^{k} \ell_h Y_h,$$

where $\ell_h$, $h = 1,2, \ldots, k$, is a constant and $\overline{Y}_h$ is the mean of stratum h. Also, we assume the total resources, C, for the sample survey is fixed and that

$$C = \sum_{1}^{k} t_h^g c_h, \quad g > 0 \tag{5.1}$$

where $c_h$ is the cost associated with sampling one unit in stratum h and $t_h$ is the total number of units sampled in stratum h.

The objective in this chapter is to allocate the resources, C, among the k strata in order to achieve a minimum for the expected precision of the predictive estimator. If the prior information for the variance in each stratum is not informative, a complete solution for the allocation is not known. In this event we propose the following ad hoc procedure utilized by Draper and Guttman [13] in their Bayesian approach to allocation in stratified sampling.

The total sample will be selected in two phases. First, a sample of size $n_h$ is obtained from stratum $h$, $h = 1,2, \ldots, k$. The value attached to a unit in this sample will be represented by $x_{hi}$, $i = 1,2, \ldots, n_h$; $h = 1,2, \ldots, k$. Second, a sample of size $m_h$ is obtained from the remaining units in stratum $h$, $h = 1,2, \ldots, k$. The value attached to a unit in this second sample will be represented by $y_{hi}$, $i = 1,2, \ldots, m_h$; $h = 1,2, \ldots, k$. There are $N_h-n_h-m_h$ units remaining in stratum $h$, $h = 1,2, \ldots, k$, and we will let $z_{hi}$ represent the unobserved value attached to $i^{th}$ remaining unit in stratum $h$.

Assume that the first-phase sample has been obtained and that $n_h$, $h = 1,2, \ldots, k$, were determined so that

$$C > \sum_1^k n_h^g \, c_h.$$

Our objective now is to determine $m_h$, $h = 1,2, \ldots, k$, such that the expected precision of the predictive estimator is a minimum subject to

$$C = \sum_1^k (m_h + n_h)^g \, c_h.$$

## Bernoulli Super Distribution

In this section we will assume the following:

(1) The super distribution in each stratum is a Bernoulli distribution with parameter, $p_h$, $h = 1,2, \ldots, k$.

(2) The prior distribution for the parameter, $p_h$, $h = 1,2, \ldots, k$, is a Jeffrey vague prior.

We now state two lemmas whose proofs are omitted because the proof of each is analogous to the proof of Theorem 2.1.

Lemma 5.1. Suppose assumptions (1) and (2) hold. Let $x_{hi}$, $i = 1,2, \ldots, n_h$, be a simple random sample of size $n_h$ from stratum h. Let $y_{hi}$, $i = 1,2, \ldots, m_h$, be a future simple random sample of size $m_h$ from stratum h. Let

$$x_h = \sum_1^{n_h} x_{hi} \quad \text{and} \quad y_h = \sum_1^{m_h} y_{hi}.$$

Then the predictive density of $y_h$ is

$$f_h^{(1)}(y_h|x_h) = \frac{\binom{y_h+x_h-1}{y_h}\binom{m_h+n_h-x_h-y_h-1}{m_h-y_h}}{\binom{m_h+n_h-1}{m_h}}, \quad y_h = 0,1, \ldots, m_h. \qquad (5.2)$$

Lemma 5.2. Suppose assumptions (1) and (2) hold. Let $x_{hi}$, $i = 1,2, \ldots, n_h$, be a first-phase simple random sample of size $n_h$ from stratum h and let $y_{hi}$, $i = 1,2, \ldots, m_h$, be a second-phase simple random sample of size $m_h$ from stratum h. Let $z_{hi}$, $i = 1,2, \ldots, N_h-n_h-m_h$, represent the unobserved value attached to the $i^{th}$ remaining unit in stratum h. Also, let

$$x_h = \sum_1^{n_h} x_{hi}, \qquad y_h = \sum_1^{m_h} y_{hi},$$

$$w_h = x_h + y_h, \qquad t_h = m_h + n_h,$$

and

$$z_h = \sum_{1}^{N_h - t_h} z_{hi}.$$

Then the predictive density of $z_h$ is

$$f_h^{(2)}(z_h | w_h) = \frac{\binom{z_h + w_h - 1}{z_h} \binom{N_h - z_h - w_h - 1}{N_h - t_h - z_h}}{\binom{N_h - 1}{t_h - 1}}, \quad z_h = 0, 1, \ldots, N_h - t_h. \qquad (5.3)$$

If we utilize a two-phase sampling scheme as previously discussed, then

$$Y = \sum_{1}^{k} \sum_{i=1}^{N_h} U_{hi}$$

can be written as

$$Y = \sum_{1}^{k} [z_h + y_h + x_h]$$

where $z_h$, $y_h$, and $x_h$ are defined in Lemma 5.2. The predictive estimator of $Y$ is

$$E(Y) = \sum_{1}^{k} [E(z_h) + y_h + x_h]$$

which will reduce to

$$E(Y) = \sum_{1}^{k} \frac{(N_h - t_h) w_h}{t_h} \qquad (5.4)$$

using (5.3).

The precision of this estimator is

$$V(Y) = \sum_{1}^{k} V(z_h)$$

and obtaining $V(z_h)$ from (5.3), we have

$$V(Y) = \sum_{1}^{k} \frac{(N_h - t_h) N_h w_h}{(t_h + 1) t_h} \left( 1 - \frac{w_h}{t_h} \right) . \tag{5.5}$$

Also, the expected precision of the predictive estimator based on the results of the first-phase sample can be derived utilizing Lemma 5.1 and is

$$E[V(Y)] = \sum_{1}^{k} N_h \left( \frac{N_h}{m_h + n_h} - 1 \right) \frac{x_h}{n_h + 1} \left( 1 - \frac{x_h}{n_h} \right) . \tag{5.6}$$

Theorem 5.1. Suppose assumptions (1) and (2) hold. Let $n_h$, $h = 1,2, \ldots, k$, be an allocation for the first-phase sample such that

$$C > \sum_{1}^{k} n_h^g c_h, \quad g > 0.$$

Then (5.6) is minimized subject to (5.1) if

$$m_h = \left\{ \frac{C (q_h c_h^{-1})^a}{\sum_{1}^{k} (q_h c_h^{-1})^a \, c_h} \right\}^{\frac{1}{g}} - n_h, \quad h = 1,2, \ldots, k \tag{5.7}$$

where

$$q_h = \left( 1 - \frac{x_h}{n_h} \right) \frac{N_h^2 \ x_h}{n_h + 1} \ , \quad h = 1,2, \ \ldots, \ k.$$

and

$$a = \frac{g}{g+1} \ .$$

Proof: Apply the method of Lagrange multipliers to minimize

$$E[V(Y)] = \sum_1^k \left\{ \frac{q_h}{m_h + n_h} - N_h^{-1} \ q_h \right\}$$

subject to

$$C = \sum_1^k (m_h + n_h)^g \ c_h, \quad g > 0.$$

It is possible that formula (5.7) could produce values such that $m_h < 0$ or $m_h > N_h - n_h$. If $m_h < 0$, then stratum h has been oversampled. Hence, this stratum should be omitted from the second-phase sample. If $m_h > N_h - n_h$, then set $m_h = N_h - n_h$. In either case it is recommended that these strata be deleted in determining the second-phase allocation and that the fixed cost be adjusted correspondingly. A new allocation is then calculated for the remaining strata.

## Normal Super Distribution

In this section we assume the super distribution in each stratum is a normal distribution with mean $\mu_h$ and variance $\sigma_h^2$, $h = 1,2, \ldots, k$. Formulas are derived for allocating the total resources, C, in order to minimize the expected precision of the predictive estimator of

$$\overline{Y} = \frac{1}{N} \sum_{1}^{k} \sum_{1}^{N_h} U_{hi}.$$

Suppose $\sigma_h^2$ is known for all $h$ and assume a Jeffrey vague prior distribution for $\mu_h$, $h = 1,2, \ldots, k$. The precision of the predictive estimator of $\overline{Y}$ is given in (4.4) by

$$V(\overline{Y}) = \frac{1}{N^2} \sum_{1}^{k} N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} . \tag{5.8}$$

Since $V(\overline{Y})$ does not involve sample observations, it is not necessary to resort to the two-phase sampling technique. Hence, we determine the $n_h$, $h = 1,2, \ldots, k$, which minimizes (5.8) subject to (5.1) by the Lagrange multiplier technique, and we obtain

$$n_h = \frac{(w_h^2 \sigma_h^2 c_h^{-1})^{\frac{1}{g+1}}}{\left\{ \sum_{1}^{k} (w_h^2 \sigma_h^2 c_h^{-1})^{\frac{g}{g+1}} c_h \right\}^{\frac{1}{g}}}, \quad h = 1,2, \ldots, k$$

where $w_h = N_h/N$. This is the classical result of Neyman given in Cochran [12] page 97 with $g = 1$.

Now suppose $\sigma_h^2$ is unknown for all $h$ and assume a Jeffrey vague prior distribution for $\mu_h$ and $\sigma_h^2$, $h = 1,2, \ldots, k$. If a one-phase sample is used, then the precision of the predictive estimator as given by (4.8) involves the sample observations. Hence, we will utilize the two-phase sampling scheme.

In addition to the terminology given at the first of this chapter, we will let

$$\bar{x}_h = \frac{1}{n_h} \sum_1^{n_h} x_{hi}, \quad \bar{y}_h = \frac{1}{m_h} \sum_1^{m_h} y_{hi},$$

$$t_h = m_h + n_h, \qquad \bar{z}_h = \frac{1}{N_h - t_h} \sum_1^{N_h - t_h} (z_{hi}),$$

$$s_{hx}^2 = \sum_1^{n_h} (x_{hi} - \bar{x}_h)^2 \Big/ (n_h - 1)$$

and

$$s_{hy}^2 = \sum_1^{m_h} (y_{hi} - \bar{y}_h)^2 \Big/ (m_h - 1) .$$

for $h = 1, 2, \ldots, k$. Hence, we can express

$$\bar{Y} = \frac{1}{N} \sum_1^k \sum_1^{N_h} U_{hi}$$

in the form

$$\bar{Y} = \frac{1}{N} \sum_1^k \{(N_h - t_h)\bar{z}_h + m_h\bar{y}_h + n_h\bar{x}_h\}. \qquad (5.9)$$

We now state three lemmas without proofs, but remark that the proofs of each follow in the same manner as the proof of Theorem 2.7. In each lemma we assume the super distribution is normal with mean $\mu_h$

and variance $\sigma_h^2$. Also, we assume the prior distribution of $\mu_h$ and $\sigma_h^2$ is a Jeffrey vague prior.

Lemma 5.3. Let $x_{hi}$, $i = 1,2, \ldots, n_h$, be a simple random sample of size $n_h$ from stratum h. Let $y_{hi}$, $i = 1,2, \ldots, m_h$, be a future simple random sample of size $m_h$ from stratum h. Then the predictive density of $\bar{y}_h$ is a t-distribution with $(n_h-1)$ degrees of freedom, location parameter $\bar{x}_h$, and precision $n_h m_h/(n_h + m_h)s_{hx}^2$.

Lemma 5.4. Let $x_{hi}$, $i = 1,2, \ldots, n_h$, be a simple random sample of size $n_h$ from stratum h. Let $y_{hi}$, $i = 1,2, \ldots, m_h$, be a future simple random sample of size $m_h$ from stratum h. Then the predictive density of $u = s_{hy}^2/s_{hx}^2$ is an F-distribution with $(m_h-1)$ and $(n_h-1)$ degrees of freedom.

Lemma 5.5. Let $x_{hi}$, $i = 1,2, \ldots, n_h$, be a first-phase simple random sample of size $n_h$ from stratum h and let $y_{hi}$, $i = 1,2, \ldots, m_h$, be a second-phase simple random sample of size $m_h$ from stratum h. Then the predictive density of $\bar{z}_h$ is a t-distribution with $(m_h+n_h-1)$ degrees of freedom, location parameter

$$\frac{m_h\bar{y}_h + n_h\bar{x}_h}{t_h}$$

and precision

$$\frac{(N_h-t_h)t_h(t_h-1)}{N_h\ q_h}$$

where

$$q_h = \frac{m_h n_h}{t_h} (\bar{y}_h - \bar{x}_h)^2 + (m_h - 1)s_{hy}^2 + (n_h - 1)s_{hx}^2. \qquad (5.10)$$

From Lemma 5.5, we obtain

$$E(\bar{Y}) = \frac{1}{N} \sum_1^k N_h \frac{m_h \bar{y}_h + n_h \bar{x}_h}{t_h}$$

as the predictive estimator of (5.9). Also, we obtain the precision as

$$V(\bar{Y}) = \frac{1}{N} \sum_1^k \frac{N_h (N_h - t_h) q_h}{t_h (t_h - 3)}$$

where $q_h$ is defined in (5.10). Now $V(\bar{Y})$ is a function of unobserved sample values; hence, we consider $E[V(\bar{Y})]$ where expectation is with respect to the predictive densities of $\bar{y}_h$ and $s_{hy}^2$, $h = 1,2, \ldots, k$. From Lemma 5.3 and 5.4, we obtain

$$E[V(\bar{Y})] = \frac{1}{N^2} \sum_1^k N_h^2 \left\{ \frac{n_h - 1}{n_h - 3} s_{hx}^2 \left( \frac{1}{m_h + n_h} - \frac{1}{N_h} \right) \right\}. \qquad (5.11)$$

<u>Theorem</u> 5.2. Suppose the super population distribution in each stratum is normal with mean $\mu_h$ and variance $\sigma_h^2$, $h = 1,2, \ldots, k$. Also, suppose the prior distribution of $\mu_h$ and $\sigma_h^2$, $h = 1,2, \ldots, k$, is a Jeffrey vague prior. Let $n_h$, $h = 1,2, \ldots, k$, be an allocation for the first-phase sample such that

$$C > \sum_1^k n_h^g c_h, \quad g > 0.$$

Then (5.11) is minimized subject to (5.1) if

$$m_h = \left\{ \frac{c(W_h^2 p_h c_h^{-1})^a}{\sum_1^k (W_h^2 p_h c_h^{-1})^a c_h} \right\}^{\frac{1}{g}} - n_h, \quad h = 1,2, \ldots, k \qquad (5.12)$$

where

$$W_h = \frac{N_h}{N},$$

$$a = \frac{g}{g+1},$$

and

$$p_h = \frac{n_h - 1}{n_h - 3} s_{hx}^2 .$$

Proof: The result is obtained by applying the method of Lagrange multipliers to (5.11) and (5.1).

The discussion following Theorem 5.1 pertaining to formula (5.7) also applies to formula (5.12). In addition, note that formula (5.11) requires $n_h \geq 4$, $h = 1,2, \ldots, k$.

Parametric Functions in Stratified Sampling

Previously in this chapter, we were concerned with estimating the overall finite population mean. In this section, we will consider the more general problem of estimating $r \leq k$ linear functions

$$L_i = \sum_{h=1}^{k} \ell_{ih} \overline{Y}_h, \quad i = 1,2, \ldots, r$$

of the $k$ stratum means where the coefficients, $\ell_{ih}$, are known.

Des Raj [6] considered this problem from the classical approach with the stratum variances known. We will assume the super population distribution in each stratum is a normal with mean $\mu_h$ and variance $\sigma_h^2$. Also, we will assume that the joint prior distribution on $\mu_h$ and $\sigma_h^2$, $h = 1,2, \ldots, k$, is a Jeffrey vague prior. With these assumptions, we will utilize the two-phase sampling scheme to determine the optimum second-phase allocation for various restrictions. We will assume that an allocation for the first-phase has been determined such that

$$C > \sum_1^k n_h^g c_h, \quad g > 0.$$

We will first consider the minimization of cost plus total expected loss based on the first-phase sample where the loss function is of the form

$$\mu_i[L_i - E(L_i)]^2, \quad i = 1,2, \ldots, r,$$

and $\mu_i$ is a constant. Hence, the function to be minimized is

$$G = \sum_1^k (m_h + n_h)^g c_h + \sum_1^r \mu_i E[V(L_i)].$$

Using Lemma 5.3 and Lemma 5.4, we can express $G$ as

$$G = \sum_1^k (m_h + n_h)^g c_h + \sum_{h=1}^k q_h \left\{ p_h \left( \frac{1}{m_h+n_h} - \frac{1}{N_h} \right) \right\}$$

where $p_h$ is defined in Theorem 5.2 and

$$q_h = \sum_{i=1}^{r} \mu_i \ell_{ih}^2 .$$

It is easy to verify that

$$(m_h + n_h)^{g+1} = \frac{q_h p_h}{g\ c_h} \tag{5.13}$$

minimizes G. Hence, the second-phase allocation formula is

$$m_h = \left( \frac{q_h p_h}{g\ c_h} \right)^{\frac{1}{g+1}} - n_h .$$

Now consider minimizing the expected loss based on the first-phase sample subject to a fixed cost. That is, minimize

$$G = \sum_{1}^{r} \mu_i E[V(L_i)]$$

subject to

$$C = \sum_{1}^{k} (m_h + n_h)^g\ c_h .$$

Expressing G as

$$G = \sum_{1}^{k} q_h p_h \left( \frac{1}{m_h + n_h} - \frac{1}{N_h} \right)$$

and using the method of Lagrange multipliers, we obtain

$$m_h = \left\{ \frac{C(q_h p_h c_h^{-1})^a}{\sum_1^k (q_h p_h c_h^{-1})^a c_h} \right\}^{\frac{1}{g}} - n_h \qquad (5.14)$$

where $q_h$ and $p_h$ are as previously defined. We note that if

$$r = 1, \quad \mu_1 = 1, \quad \text{and} \quad \ell_{1h} = \frac{N_h}{N}$$

then (5.14) is equivalent to (5.12).

Now suppose we wish to minimize the cost subject to a fixed expected variance based on the first-phase sample. That is, minimize

$$C = \sum_1^k (m_h + n_h)^g c_h$$

subject to

$$E[V(L_i)] = a_i, \quad i = 1,2, \ldots, r,$$

where $a_i$ are fixed constants. If we apply Lagrange multipliers, we have the system of equations

$$(m_h + n_h)^{g+1} = p_h \left( \sum_1^r \lambda_i \ell_{ih}^2 \right) (gc_h)^{-1}, \quad h = 1,2, \ldots, k$$

$$\sum_{h=1}^k \ell_{ih}^2 p_h (m_h + n_h)^{-1} = a_i + \sum_{h=1}^k \ell_{ih}^2 p_h N_h^{-1}, \quad i = 1,2, \ldots, k$$

We note that these equations are not algebraic in $m_h$, $h = 1,2, \ldots, k$, and $\lambda_i$, $i = 1,2, \ldots, r$. Hence, these equations would have to be solved by an iterative process to determine the second-phase allocation.

# CHAPTER VI

## SUMMARY AND EXTENSIONS

Our study is devoted to the application of predictive densities to sample surveys utilizing the super population concept as given by Cochran [3], [4]. This approach is applied to three general areas of sampling theory, namely, (i) estimation of the finite population mean in simple random sampling, (ii) estimation of the finite population mean utilizing available auxiliary information, and (iii) allocation of sampling units among strata when estimating a linear function of the stratum means is of interest.

Estimators of the finite population mean are derived in Chapter II assuming the super population is (i) a Bernoulli, (ii) an exponential, and (iii) a normal distribution. Also, a measure of the precision of these predictors is obtained. Auxiliary information is utilized in Chapter III to derive regression and ratio type estimators of the finite population mean.

The results of Chapters II and III are extended in Chapter IV to obtain estimators of the overall finite population mean when it is feasible to stratify the total finite population. These results are then used in Chapter V to derive formulas for allocating the sampling units among the strata. In particular, allocation formulas are derived when estimating several linear combinations of stratum means and only vague prior information is available for the vector parameter $\theta$ of the super population.

Since only vague prior information is available, we allocated a portion of the total resources to the first-phase sample and then, based on the results of the first-phase sample, the remaining resources were allocated among the strata. A problem for future consideration would be to determine a best method of allocating the total resources between the first-phase sample and the second-phase sample. In addition, the allocation problem for a stratified population utilizing auxillary information would be of interest.

It would be of interest to apply the technique we have used to other areas of sample surveys. For instance, our technique could be adapted to cluster sampling. Also, it may be of interest to apply this procedure to sampling with probability proportional to size.

Another area of interest for future study is to compare finite population parameters utilizing predictive densities. Geisser [2] has suggested how this could be done.

# BIBLIOGRAPHY


[1] Kalbfleisch, J. D. and D. A. Sprott, "Applications of Likelihood and Fiducial Probability to Sampling Finite Populations," New Developments in Survey Sampling, New York, Wiley-Interscience, 358-389, 1969.

[2] Geisser, Seymour, "The Inferential Use of Predictive Distribution," Foundations of Statistical Inference, New York, Holt, Rinehart and Winston, 456-469, 1971.

[3] Cochran, W. G., "The Use of Analysis of Variance in Enumeration by Sampling," Journal American Statistical Association, 34, 492-510, 1939.

[4] Cochran, W. G., "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," Annals of Mathematical Statistics, 17, 164-177, 1946.

[5] Draper, Norman R. and Irwin Guttman, "Some Bayesian Stratified Two-Phase Sampling Results," Biometrika, 55, 131-139, 1968.

[6] Raj, Des, "On Estimating Parametric Functions in Stratified Sampling Designs," Sankhyā, 17, 361-366, 1957.

[7] Zellner, Arnold. An Introduction to Bayesian Inference In Econometrics. New York: John Wiley and Sons, Inc., 38-53, 1971.

[8] DeGroot, Morris H. Optimal Statistical Decisions. New York: McGraw-Hill Book Co., 157-174, 1970.

[9] Feller, William. An Introduction to Probability Theory and Its Applications, Vol. 1, Third Edition. New York: John Wiley and Sons, Inc., 1968.

[10] LaValle, Irving H. An Introduction to Probability, Decision, and Inference. New York: Holt, Rinehart, and Winston, Inc., 1970.

[11] Lindley, D. V. Introduction to Probability and Statistics From a Bayesian Viewpoint, Part 2, Inference. Cambridge at the University Press, 1965.

[12] Cochran, William G. Sampling Techniques. New York: John Wiley and Sons, Inc., 1967.

# VITA

Melvin Valone Foster

Candidate for the Degree of

Doctor of Philosophy

Thesis: BAYESIAN PREDICTIVE DENSITIES AND THEIR APPLICATION TO SAMPLING THEORY

Major Field: Statistics

Biographical:

Personal Data: Born in Butterfield, Arkansas, February 20, 1934, the son of Arthur and Blanche Foster.

Education: Graduated from Magnet Cove High School, Malvern, Arkansas, in May, 1952; received the Bachelor of Science degree from Southern State College, Magnolia, Arkansas, in January, 1956, with majors in chemistry and mathematics; received the Master of Science degree from the University of Arkansas, Fayetteville, in January, 1963, with a major in mathematics; attended University of Missouri at Rolla, Rolla, Missouri in the summer of 1964 as a National Science Foundation Institute participant; attended the University of South Carolina, Columbia, South Carolina the academic term of 1965-66; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May, 1973.

Professional Experience: Chemist, Crossett Paper Mills, Crossett, Arkansas, 1956-57; Chemist, Arkansas Chemical Corporation, Malvern, Arkansas, 1957-58; Chemist, Arkansas Louisiana Chemical Corporation, Magnolia, Arkansas, 1958-61; Graduate Assistant, University of Arkansas, Fayetteville, Arkansas, 1961-62; Instructor of Mathematics, Memphis State University, Memphis, Tennessee, 1962-65; Instructor of Mathematics, University of South Carolina, Columbia, South Carolina, 1965-66; Assistant Professor of Mathematics, Southwest Missouri State University, Springfield, Missouri (on leave of absence 1969-73), 1966- ; Graduate Assistant, Oklahoma State University, Stillwater, Oklahoma, 1969-73; Research Assistant, Los Alamos Scientific Laboratories, Los Alamos, New Mexico, summers of 1971 and 1972.

Professional Organizations: Member of American Statistical Association.