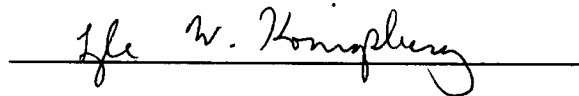


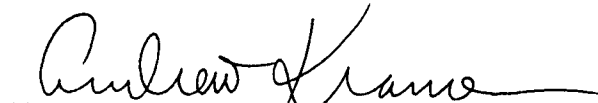
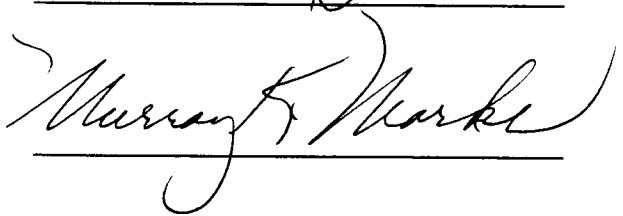
To the Graduate Council:

I am submitting herewith a thesis written by Ashley Hyatt McKeown entitled "Cranio-metric Sex Estimation Using a Finite Mixture Missing Data Model." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Anthropology.



Dr. Lyle W. Konigsberg, Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:



Associate Vice Chancellor and
Dean of The Graduate School

CRANIOMETRIC SEX ESTIMATION USING A
FINITE MIXTURE MISSING DATA MODEL

A Thesis

Presented for the

Master of Arts

Degree

The University of Tennessee, Knoxville

Ashley Hyatt McKeown

December 1996

ACKNOWLEDGMENTS

I would like to thank each of my committee members, Dr. Murray K. Marks, Dr. Andrew Kramer and my chair, Dr. Lyle W. Konigsberg, for their assistance with this thesis. Dr. Konigsberg selflessly provided his statistical expertise which was invaluable to the completion of this thesis. He patiently explained various aspects of the statistical method utilized in this thesis and guided me through the pitfalls of thesis writing with good humor. Drs. Marks and Kramer provided me with sage advice that greatly enhanced the final product found here. I am deeply indebted to all my committee members for their time and valuable comments.

I also need to thank Dr. R.J.A. Little for the FORTRAN program which was used in this thesis.

On a personal note, I would like to acknowledge all those who lovingly cheered me on while I worked on this thesis, including Todd Ahlman and my dog, Shelby. My parents, Pat and Carolyn, and my brother, Chris, have always encouraged and supported me, and this thesis is no exception. I owe them a great debt of gratitude and much love for all they have given me and made possible for me.

The data collection on the Averbuch material was funded by National Science Foundation Grant SBR-9307693 to Dr. Lyle W. Konigsberg.

ABSTRACT

Accurate estimation of sex from human skeletal material is crucial for generating inferences regarding prehistoric populations. Metric approaches to sex assessment can provide an objective evaluation of skeletal material. However, commonly used procedures such as discriminant analysis require a reference sample or training set which may not be an appropriate analogy for prehistoric samples with differing levels of sexual dimorphism and/or overall size. Additionally, since most statistical analyses are developed for cases with a complete data set, such methods are hindered by unobservable measurements. This thesis tests the utility of applying finite mixture analysis, a form of cluster analysis that does not require a reference sample and also permits missing data, to craniometrics for the purpose of sex estimation.

Finite mixture analysis is applied to two of Howells' known sex samples after measurements were randomly removed from the data set simulating missing data patterns found in 154 crania from the Late Mississippian Averbuch site. The analysis provided levels of accuracy comparable to discriminant analysis for the relatively dimorphic sample from Northern Japan both with and without missing data. However, the analysis performed poorly for the less dimorphic sample from South Africa both with and without missing data. When applied to an actual archaeological sample consisting of 94 Averbuch crania, the analysis performed relatively well considering the prevalence of missing measurements.

Table of Contents

CHAPTER 1	
INTRODUCTION	1
CHAPTER 2	
LITERATURE REVIEW	4
Visual and discriminant estimation of sex	4
Anthropological applications of finite mixture analysis	11
CHAPTER 3	
MATERIAL AND METHODS	15
CHAPTER 4	
RESULTS	30
CHAPTER 5	
DISCUSSION	39
CHAPTER 6	
CONCLUSIONS	48
LITERATURE CITED	51
APPENDIX	57
VITA	62

List of Tables

Table 1. Howells (1973) samples used in finite mixture analysis simulation	16
Table 2. Craniometrics used in finite mixture analysis simulation of the Howells (1973) sample	18
Table 3. Mahalanobis distance (D^2), Burnaby's size distance (D^2_M) and percent of D^2 that is due to "size" difference (% "size") between males and females	19
Table 4. Observed vs. predicted sex classification for the restricted Averbuch sample when probability of the more likely sex was greater than 0.5	34
Table 5. Observed vs. predicted sex classification for the restricted Averbuch sample when probability of the more likely sex was greater than 0.8	36
Table 6. Observed vs. predicted sex classification for the restricted Averbuch sample when probability of the more likely sex was greater than 0.9	38
Table 7. Summary of results for finite mixture missing data simulation	42
Table 8. Summary of the percent correct classification for the finite mixture analysis of the Averbuch sample at the three levels of significance	45

List of Figures

- Figure 1. Vector of cumulative frequencies and the matrix **M** for missing data simulations 25
- Figure 2. Results of the finite mixture with missing data simulation incorporating the Hokkaido sample from Northern Japan. 31
- Figure 3. Results of finite mixture analysis with missing data simulation incorporating the Zulu sample from South Africa. 32

CHAPTER 1

INTRODUCTION

Estimation of sex from human skeletal material has long been a focus of physical anthropology. Approaches include visual assessment of skeletal material (see Bass 1987; Krogman 1962; Stewart 1979; Ubelaker 1989) and metric methods which classify individuals based on analysis of a variable or set of variables (see Giles and Elliot 1962, 1963). The most accurate region of the human skeleton for determining sex is the pelvis due to morphological requirements related to childbirth (Krogman 1962; Ubelaker 1989). The pelvic elements of the female possess primary sex characteristics that allow for discrimination between males and females. However, in both the archaeological and forensic contexts pelvic elements are often unrecoverable or scavenged requiring other elements to be utilized for sex estimation. The cranium is generally regarded as the second best region of the skeleton for determining sex (Krogman 1962; Ubelaker 1989) as secondary sex characteristics typically related to size are exhibited by the cranium. However, these size dependent characters vary from population to population and standards from one population should not be applied to another population.

Research (Weiss 1972; Meindl et al. 1985) has demonstrated that both males or females can be overclassified by traditional sexing methodologies (i.e. Bass 1987; Giles and Elliot 1962, 1963; Krogman 1962; Stewart 1979; Ubelaker 1989). Weiss (1972)

noted that the overclassification of males resulted in biased sex ratios which probably do not accurately represent the mortality patterns of past populations. Reviewing the sex ratios of several contemporary pre-industrial societies revealed that they exhibit the balanced sex ratio (1 male: 1 female) seen in modern industrial populations (Weiss 1972). This indicates that accepting a skewed sex ratio for prehistoric populations is unsupportable and that methods which produce unbalanced sex ratios are probably biased.

Metric methods have been developed in an effort to provide more objective approaches to determining sex from skeletal material (see Cunha and van Vark 1991; Black 1978; DiBernardo and Taylor 1979; Giles and Elliot 1962, 1963; Henke 1977; Keen 1950). However, metric methods generally require a known sex sample which may not be a good analogy for prehistoric series. Research reported by Meindl et al. (1985) has shown that metric methods are population specific and are not appropriate for universal application, particularly for prehistoric samples. Additionally, multivariate approaches are designed to employ a complete data set and missing values can introduce complications into the analysis. Therefore other methods for accurately assessing sex of human skeletal remains need to be evaluated.

Finite mixture analysis does not require a reference sample or training set and also allows for missing data while it classifies individuals into subsets of a larger group. Various forms of the method have been applied to anthropological questions (Brant et al. 1990; Pearson et al. 1992; Godfrey et al. 1993; Plavcan 1994; Josephson et al. 1996; Kramer and Konigsberg 1996). However, it has yet to be tested on a human skeletal sample. The main objective of this thesis is to test the utility of a finite mixture missing

data model for classifying individuals on the basis of craniometrics. In order to evaluate the procedure, analyses were performed with two known sex samples and a missing data simulation was enacted to create data sets that mimicked the missing data patterns one might encounter in an archaeological sample. Then finite mixture analysis was applied to an actual prehistoric sample of crania from the Averbuch site where the patterns of missing data employed in the missing data simulation were found. Comparison of the results of these analyses reveals much about the applicability of finite mixture analysis for estimating sex for human skeletal series.

CHAPTER 2

LITERATURE REVIEW

Visual and discriminant estimation of sex

The sex of human skeletal material can be estimated either by visual assessment of morphological criteria or via metric methods that incorporate various measurements in order to classify the individual under investigation. Initial research regarding the assessment of sex from skeletal remains generally involved the subjective observation of characters of the cranium and/or pelvis (i.e. Bass 1987; Krogman 1962; Stewart 1979). Such traditional criteria remain a mainstay of the bioarchaeological and osteological repertoire for sex assessment. More recently much emphasis has been placed on developing metric methods (i.e. Black 1978; Cunha and van Vark 1991; DiBernardo and Taylor 1979; Giles and Elliot 1962, 1963; Henke 1977) that result in more objective approaches that are easier to learn and potentially more reproducible (Thieme and Schull 1957; Washburn 1948).

Researchers contend that traditional visual assessment of the skull by an experienced observer evaluating levels of rugosity can achieve accuracy from 80-90% (Krogman 1962; Ubelaker 1989). Morphological differences in the pelvic girdle can be used to determine sex with an accuracy of 95% (Krogman 1962). When the morphology of the pelvis (namely the *os coxae*) is used in conjunction with the skull, Krogman (1962)

claimed accuracy levels approaching 98%. Meindl et al. (1985) conducted a study to investigate these claims in which they independently assessed the skulls and pelves of 100 adult skeletons from the Hamann-Todd Collection to determine sex. They reported that when the skull was used alone, two above average assessors were able to achieve 90% accuracy and when sex assessment was based solely on pelvic criteria, an accuracy of 93% was achieved. In accordance with previous research (Krogman 1962), when the assessors combined the skull and pelvis from the same individual they were able to correctly classify 97% of the sample. Meindl and coworkers (1985) noted that most of their errors were males misclassified as females, contrary to a previous study by Weiss (1972) that found a bias in the direction of males from mortuary contexts. Additionally, Meindl et al. (1985) concluded that familiarity with the level of sexual dimorphism in the population under investigation is crucial for accurate sex assessment but that perhaps the most important factor for determining accuracy in such assessment is the completeness of the remains.

In a successful attempt to define specific morphological criteria of the pelvis that can reliably be employed for sex estimation, Phenice (1969) described three aspects of the *os pubis* that can be used to achieve accuracy exceeding 95%. By assessing the morphology of the ventral arc, subpubic concavity, and the medial aspect of the ischio-pubic ramus, Phenice (1969) claimed that this technique can be used quickly and objectively. Sutherland and Suchey (1991) continued this approach with their research regarding the efficacy of employing the ventral arc of the *os pubis* as the sole criterion. Utilizing a sample of 1284 well documented pubic bones, Sutherland and Suchey (1991) determined that the use of the ventral arc alone can provide 96% accuracy.

Other aspects of the *os coxae* have been investigated regarding their efficiency for sex estimation. Houghton (1974) documented the occurrence of two forms of the preauricular groove employing two series of *os coxae*, one from the Department of Anatomy, University of Ontario and the other of unknown origin. He defined two different manifestations of this groove, the groove of pregnancy (GP) and the groove of ligament (GL), and the absence of either form of the groove (NG). Houghton (1974) reported that while the GP is exhibited solely by females, only 71% of the females sampled possessed that particular form of the preauricular groove. Another 21% of the female pelvis exhibited the GL (no GP) and 8% had neither form of the groove. Eighty-one percent of the male pelvis in the sample exhibited a preauricular groove classified as GL while the remaining 19% exhibited no groove (NG). Houghton attributed the GP to parturition and suggested that it definitively classified female individuals. However, he noted that individuals exhibiting the GL are not necessarily male since both sexes can possess this form of the preauricular groove.

Recently, Loth and Henneberg (1996) described a characteristic of the mandibular ramus which correctly predicted sex in 99% of the study sample. They found that males exhibit flexure on the posterior border of the mandibular ramus at the level of the occlusal surface of the molars; whereas, females either exhibit a straight border or flexure either above or below the occlusal surface of the molars. Based on this criterion, the authors correctly classified 99% of the original study sample of 200 normal known sex mandibles of South African origin, while 91% of 100 mandibles composing the pathologic sample were correctly diagnosed. When applied blind to a sample (N=247) of American whites,

blacks and Amerinds, accuracy ranged from 91% for the blacks and whites to 92% for the Amerinds. Additionally, the authors suggest that the trait is evident in fossil hominids and may be feasible for sex determination in such specimens.

Metric methods may provide more objective approaches to assessing sex of human skeletal remains. Keen (1950) sought to define differences between the male and female skull for use in determining sex of unknown individuals. Employing a collection of known sex skulls from the Cape Coloured population of South Africa, Keen evaluated previously described measurements and morphological criteria and their value in differentiating males and females. Subsequently, he outlined a set of 4 measurements that when used in conjunction with 3 morphological characters could determine sex with 85% accuracy. The measurements combined with the visual criteria were used to classify individuals into one of three categories: male, questionable, and female. However, Keen emphasized that this method should not be applied to other populations since levels of sexual dimorphism are different from one population to the next.

Hanihara (1959) published discriminant functions employing 9 measurements of the skull for diagnosing sex in Japanese individuals. From a sample of 105 modern Japanese individuals, Hanihara derived discriminant functions with theoretical accuracies ranging from 83% (calvarium only) to 90%. Giles and Elliot (1962) reported discriminant functions for determining race from 8 cranial measurements, 5 of which the researchers noted could be used for determining sex. In a subsequent publication expanding on this methodology, Giles and Elliot (1963) provided several discriminant functions developed solely for determining sex using 9 cranial measurements in varying combinations.

Measurements from 75 known sex skulls from the Terry and Todd Collections were employed to derive 21 functions considered applicable to American blacks and whites. The resulting functions were tested on 333 known sex skulls from the same collections and accuracy ranged from 82-89%. According to the authors this level of accuracy was comparable to visual sexing. In further testing, the researchers included 3 prehistoric Native American series: Pecos Pueblo, Indian Knoll and Florida. They used a portion of each sample that had been securely sexed based on visual morphological assessment to derive new sectioning points for the existing discriminant functions. Upon application to the remaining portions of each sample, accuracies ranging from 75-92.7% were achieved with a majority falling in the 80% range, basically providing results as accurate as those discriminant functions when applied to the black and white material.

Birkby (1966) attempted to apply the discriminant functions derived by Giles and Elliot (1962) to 95 American Indians and 9 Labrador Eskimos in order to test their applicability for sex determination. The males in the sample were classified as accurately as could be expected for visual sexing regardless of deformation; however, the female series had much lower levels of correct classification. In particular, 80% of one series of females were misclassified. From these results, Birkby (1966) concluded that discriminant functions derived from a single Native American sample cannot be uniformly used to determine sex in other Native American series.

Kajanoja (1966) utilized a known sex sample of Finnish crania to derive two multivariate discriminant functions employing 8 variables each. When the first function was applied back to the series of 232 known sex Finnish crania, 79.4% of the males and

79.1% of the females were correctly classified. However, when Kajanoja applied a discriminant function analogous to function number 19 from Giles and Elliot (1963) to the Finnish material, only 65% of the crania were correctly sexed. Snow and coworkers (1979) also conducted a test of the Giles and Elliot discriminant functions for race and sex employing 52 crania from a series of forensic cases. The discriminant function for sex incorporating 5 cranial measurements as outlined in Giles and Elliot (1962) correctly classified 88% of the forensic sample.

Henke (1977) employed the Westerhus series of known sex crania to calculate multivariate discriminant functions which were then applied to samples of varying antiquity and geography. Actual misclassification rates varied depending on the particular function being applied, the method by which the sectioning point was found, and the series under investigation. Henke noted that when the Westerhus sectioning point was used within a discriminant function, it was most accurate for those groups most similar to the Westerhus series.

In response to problems applying metric methods to fragmented and deformed crania, Holland (1986) developed 6 regression models incorporating measurements of the cranial base for determining sex. When reapplied to the reference sample of 100 crania from the Terry Collection, Holland achieved accuracies ranging from 71-90%. When the equations were applied to a separate sample of 20 skulls from the Terry Collection, 70-85% of the series was correctly classified depending on the equation used.

Recently, Cunha and van Vark (1991) investigated the applicability of employing a large series of known sex skulls for the purpose of calculating multivariate discriminant

functions. Using a sample of 570 known sex skulls collected between 1896 and 1903 housed at the University of Coimbra, Portugal, the authors derived a multivariate discriminant function with which they reclassified the reference sample using the "leaving one out (L.O.O.)" based method. The function correctly classified only 80.1% of the collection. Trying to determine where the heterogeneity existed within the series, they separated the crania into three samples based on geography: North, Central and South Portugal. Using the L.O.O. based method, the samples were reclassified resulting in correct classifications of 75.2% for the North sample, 80.3% for the Central sample and 79.8% for the South sample. Cunha and van Vark tentatively attributed these poor results to relatively low sexual dimorphism related to heterogeneity in the sample. They concluded that a series such as this particular sample would not adequately serve as a reference sample for widely applicable discriminant functions.

Both univariate and multivariate metric methods have also been applied to post cranial elements. Iscan and Miller-Shaivitz (1986) examined the expression of sexual dimorphism in the femur and tibia and calculated discriminant functions for whites and blacks for these elements based on a sample from the Hamann-Todd Collection. For whites the functions ranged from 87-94% in accuracy and for blacks a similar range of 83-93% accuracy was achieved. A study by Black (1978) had already recognized the potential of the femur for sex estimation. In this research, Black described a univariate discriminant function employing the femoral circumference at the midshaft that classified 85% of 114 skeletons from the Libben Site, Ontario County, Ohio consistent with other methods of sex estimation. DiBernardo and Taylor (1979) tested Black's (1978)

discriminant function using the femoral circumference at the midshaft on 115 American white femurs of known sex. They were able to correctly classify 82% of the test sample and determined that the midshaft circumference was as accurate as other criteria used to sex the femur. Other research regarding sex estimation from the femur and tibia has been conducted and reported by Wu (1989) and Iscan et al. (1994), respectively.

Various other methods for determining sex from postcranial elements include, but are not limited to, discriminant functions employing the talus and/or calcaneus (Steele 1976), diameter of the radial head (Berrizbeitia 1989), measurements of the arm and wrist (Holman and Bennett 1991) and measurements of the metacarpals and the first proximal phalanx (Scheuer and Elkington 1993).

A theme common to many of these studies is the population specific nature of both morphological criteria and metric methods. As Meindl et al. (1985) emphasized, levels of sexual dimorphism vary from population to population and prehistoric populations should not be evaluated with modern standards. This is an important aspect of current studies concerning sex assessment from skeletal material which will be addressed in this thesis through the use of finite mixture analysis.

Anthropological applications of finite mixture analysis

Finite mixture analysis has been considered and applied to anthropological research questions. Brant, Morrell and Pearson first presented a summary of the approach as a potential tool for anthropologists in a 1990 AAPA symposium. Pearson and coworkers (1992) discuss finite mixture analysis as a univariate method for exploring complex

distributions and identifying commingled subpopulations within genetic, epidemiological and anthropological problems. The authors presented two examples of finite mixture application. The first involved distinguishing between healthy individuals and individuals with subclinical diseases using white blood cell counts. A second application used mesiodistal breadth of mandibular first molars of African dryopithecines in an effort to discern if two or three (as the prevailing taxonomic classification contends) taxonomically distinct species exist within the genus. For this example alternative interpretations of the cluster analysis are discussed as the mixture of three normals (*Dryopithecus africanus*, *D. nyanzae*, and *D. major*) is not a significantly better fit than the mixture of two normals (*D. major* and combined *D. africanus* and *D. nyanzae*). While they present an overview of finite mixture and its potential applications in anthropological research, they only address univariate analysis.

Godfrey et al. (1993) utilized univariate finite mixture analysis in an attempt to discern patterns of sexual dimorphism as exhibited in extinct large-bodied lemurs. Since most modern lemurs consistently exhibit low levels of sexual dimorphism, they are poor analogies for extinct lemurs. The authors utilized samples of extinct large-bodied Malagasy lemurs that vary considerably in skull length (the single variable) in conjunction with finite mixture analysis to determine if they could detect sexual dimorphism patterns embedded in the unimodal distribution. Application of finite mixture analysis and two other approaches revealed that low levels of dimorphism exist amongst the skull lengths for the subfossil lemurs similar to that observed in their modern relatives. In this instance, finite mixture was more effective than the other analyses employed (division at the median

or mean) for extant species even when dimorphism was relatively low. However, when levels of true dimorphism were very low, all three methods provided poor estimations of the potential bimodal structure of the population.

Plavcan (1994) utilized computer modeling experiments to evaluate four methods, including finite mixture analysis as described by Godfrey and coworkers (1993), for estimating sexual dimorphism for fossil species from skeletal and dental measurements. Since these methods could be affected by sample size, sex ratio, or the amount of intrasexual variability, Plavcan (1994) controlled these parameters throughout the course of the computer simulations and assessed the various methods. His results demonstrated that finite mixture analysis was most effective for estimating the maximum sexual dimorphism for unimodally distributed populations. However, when dimorphism was low and intrasexual variability decreased, finite mixture analysis actually underestimated the dimorphism present in the sample.

Recently, Josephson et al. (1996) employed finite mixture analysis using the method-of-moments, in an attempt to estimate sexual dimorphism in distributions where the underlying structure is unknown. Using a single metric trait, the method-of-moments approach worked well for distributions where there was little overlap between males and females, and even had relative success with distributions where substantial overlap existed. However, Josephson and coworker's application of the method assumed fixed mixing proportions ($P(\text{male}) = P(\text{female})$) which limits the effectiveness of their application to define the true structure of the distribution.

In an application of finite mixture analysis similar to that tested in this thesis, Kramer and Konigsberg (1996) utilized the Expectation Maximization algorithm to group large bodied hominoids by species. To replicate missing data patterns similar to those encountered in the fossil samples, measurements were stochastically removed from the complete data sets consisting of 6 craniometrics for 64 chimps, 60 humans, 64 gorillas and 29 orangutans. For the complete data sets finite mixture analysis correctly classified 100% of the sample composed of the chimps and humans and misclassified only one individual in the sample composed of the gorillas and orangutans. Over 200 runs the finite mixture missing data model correctly classified 99% of the chimp-human sample and 89% of the gorilla-orangutan sample. These results indicate that finite mixture analysis may provide a heuristic tool for assessing group phylogeny.

Given the relatively recent introduction of finite mixture analysis to anthropological problems and the lack of application in the bioarchaeological arena, this thesis tests the utility of employing the procedure for estimating sex from human skeletal material. Since it does not necessitate a reference sample (which may have no particular relationship to the series under investigation) and allows for missing data, finite mixture analysis should prove useful for estimating the sex of individuals from prehistoric samples.

CHAPTER 3

MATERIAL AND METHODS

In order to evaluate the utility of finite mixture analysis for determining sex from craniometrics, two known sex samples compiled by W.W. Howells as part of a much larger collection for multivariate analysis (Howells 1973) were employed. The Hokkaido sample from Northern Japan consists of 55 males and 32 females of known sex. Housed in the Anatomy Department of the Medical School at Hokkaido University, Sapporo, Japan, this dissecting room collection was measured by W.W. Howells in 1968 (Howells 1973). The Zulu sample from South Africa is part of the Raymond A. Dart Collection housed at the University of Witwatersrand, Johannesburg, R.S.A. Composed of 55 males and 46 females of known sex, this sample is largely drawn from the dissecting room of the Department of Anatomy. In the process of collecting the measurements, Howells eliminated one female cranium from the Zulu sample because as the largest skull in the collection it was judged to be aberrant (Howells 1973). The composition of these samples is shown in Table 1.

Howells (1973) collected a large series of measurements for all individuals in his samples. For the purpose of this study, the following 6 measurements were used: glabella-occipital length (GOL), maximum cranial breadth (XCB), bifrontal breadth (FMB),

Table 1. Howells (1973) samples used in finite mixture analysis simulation

Howells Sample	Male	Female	Total
Hokkaido, Northern Japan	55	32	87
Zulu, South Africa	55	46	101

bizygomatic breadth (ZYB), basion-bregma height (BBH) and basion-nasion length (BNL) (see Table 2).

These particular samples were selected for two reasons. First, both of the samples consist of specimens of known sex, which is critical for evaluating the accuracy of any method of sex estimation. Second, the samples differ in levels of sexual dimorphism as demonstrated by the Mahalanobis distance (D^2) between the males and females for each group on the basis of the 6 craniometrics and Burnaby's (1966) "size" difference (D^2_M) between the sexes. The much higher values of D^2 and D^2_M calculated for the Hokkaido sample from Northern Japan indicates that the sample is more dimorphic than the Zulu sample from South Africa. Additionally, the percentage of the total Mahalanobis distance due to "size" (% "size") indicates that the degree of sexual dimorphism expressed as "size" is much less in the Zulu sample than in the Hokkaido sample. These values are shown in Table 3.

Finite mixture analysis is a form of cluster analysis that creates groups or clusters within a distribution without *a priori* knowledge regarding the structure of the distribution (McLachlan and Basford 1988). Additionally, finite mixture analysis can incorporate specimens with missing data into the multivariate analysis. By employing the Expectation-Maximization (EM) algorithm, this analysis can be simply performed and posterior probabilities that an individual specimen belongs to a particular group within the distribution computed (Little and Rubin 1987; Little and Schluchter 1985).

Table 2. Craniometrics used in finite mixture analysis simulation of the Howells (1973) sample

Craniometric	Definition
Glabella-occipital length (GOL)	Greatest length, from the glabellar region, in the median sagittal plane (Howells 1973:170).
Maximum cranial breadth (XCB)	The maximum breadth across the perpendicular to the median sagittal plane (above the supramastoid crests) (Howells 1973:172).
Bifrontal breadth (FMB)	The breadth across the frontal bone between frontomale anterior on each side, i.e. , the most anterior point on the fronto-malar suture (Howells 1973:177).
Bizygomatic breadth (ZYB)	The maximum breadth across the zygomatic arches, wherever found, perpendicular to the median plane (Howells 1973:173).
Basion-bregma height (BBH)	Distance from bregma to basion, as defined (Howells 1973:172).
Basion-nasion length (BNL)	Direct length between nasion and basion (Howells 1973:171).

Table 3. Mahalanobis distance (D^2), Burnaby's size distance (D^2_M) and percent of D^2 that is due to "size" difference (% "size") between males and females

Howells Sample	D^2	D^2_M	% "size"
Hokkaido, Northern Japan	8.23	5.91	71.8%
Zulu, South Africa	3.54	1.49	42.1%

Unlike other forms of multivariate analysis, finite mixture does not require a reference sample for determining parameters for classification. This is an important aspect of its application since prehistoric populations exhibit differing levels of sexual dimorphism and/or overall size than that observed in modern populations used as training sets or reference samples. A commonly used multivariate analysis among anthropologists is discriminant analysis. However, this approach necessitates a training set which is often a modern sample or a subset of the series under investigation where specimens are “securely” sexed through morphological assessment and/or death records. Both types of reference samples are subject to error in application either due to a differing level of dimorphism as in the case of the modern sample training set or through misclassification of specimens included in the subset for the reference sample.

Generally, standard statistical procedures are designed for complete data sets and missing data often complicates analysis (Little and Schenker 1995). However, missing elements or portions of elements which result in incomplete data sets is common with archaeological specimens and can pose a problem when attempting statistical analysis.

Common approaches to dealing with missing data include complete-case analysis, available-case analysis and unconditional mean imputations (Little and Schenker 1995). The complete-case analysis only utilizes cases with complete data sets, whereas the available-case analysis uses those cases that contain a particular variable or set of variables and unconditional mean imputation involves replacing missing variables with sample means (Little and Schenker 1995). However, these are naive approaches in that they introduce potential biases into the statistical process.

Utilizing only cases that have complete data sets wastes information and in the case of an archaeological data set potentially results in a very limited sample. Samples restricted to complete cases may not be representative, and does not necessarily constitute a random subsample of the original sample. This often leads to biased interpretations from the analysis (Little and Rubin 1987; Little and Schenker 1995). In the second approach, available-case analysis, means are calculated variable by variable, so that the sample sizes differ amongst variables. The covariances are calculated using "pairwise-deletion" (van Vark and Schaafsma 1992). For a univariate application this method involves using a data set that is composed of all cases that have a particular variable available (Little and Rubin 1987; Little and Schenker 1995). From a single sample, analysis of each variable or subset of variables on an available case basis would employ a different subsample (Little and Rubin 1987; Little and Schenker 1995; van Vark and Schaafsma 1992) making interpretations regarding the entire sample difficult at best.

A third approach to missing data is unconditional sample mean imputation. While this method completes previously incomplete data sets and allows for multivariate analysis, inconsistent parameters can arise resulting in distorted associations between variables under investigation. These inconsistencies in turn affect the covariance matrix potentially producing biased results (Little and Rubin 1987; Little and Schenker 1995). Additional adverse effects of imputing sample means to replace missing values can include the underestimation of variances and covariances (Santos 1981).

This review of methods for dealing with missing data indicates the need for alternative approaches. Little and Rubin (1987) have extensively discussed maximum

likelihood estimators that can appropriately deal with incomplete data sets for multivariate analysis. Within this body of methodology, they have specifically dealt with the finite mixture case by defining a model for the data and missing data mechanism and employing maximum likelihood to estimate the parameters. These parameter estimates can then be used to generate values to impute, replacing those that are missing, and complete the data set via the EM algorithm (Little and Schluchter 1985). This approach reduces the potential for biased results as seen in the previously described methods for dealing with missing data by generating an unbiased estimation of the missing observation.

While the finite mixture analysis employed in this thesis is similar to the previous applications addressed in Chapter 2 of this thesis in that it does not require a reference sample, it differs from most of the studies in that it incorporates multiple variables for an individual and unbiased imputation of missing values. The Expectation-Maximization (EM) algorithm as described by Little and Schluchter (1985) is a straightforward way to fit the finite mixture missing data model providing maximum likelihood estimates for parameters which can be used as the basis for a cluster analysis. The EM algorithm itself is subdivided into two steps, the expectation (E) step and the maximization (M) step. For the E step, the investigator establishes initial parameters for cell means, cell probabilities and the covariance matrix. For the purpose of this study, the initial parameters are defined as follows. Sample cell means are scaled up and down by 10% to represent the male and female vectors of means. Mixing proportions of 0.5 and 0.5 for a sex ratio of 1:1 are initial estimates for cell probabilities. Finally, the pooled within-group variance-covariance matrix begins as a diagonal matrix with the observed total sample variances along the

diagonal. The initial parameters are used to compute the expected values of the unobserved data conditional on the observed data and to calculate discriminant functions for any pattern of observed data. These discriminant functions are used to calculate current estimates of the posterior probabilities of being male and female. Any missing values are then replaced using multivariate regressions of missing on observed data, weighting the regression by the probability that a particular case is male or female.

The subsequent M step computes the new maximum likelihood parameter estimates based on the now complete data sets as created by the E step. Then the algorithm returns to the E step to recompute the expected values for the unobserved data and the posterior probabilities that an individual is male and female based on the new maximum likelihood parameters. The algorithm cycles between the two steps until convergence occurs. Based on the posterior probabilities that emerge after convergence of the E and M step, clustering of individuals within the distribution can occur. In this particular case, they are classified as male or female.

A FORTRAN listing was obtained from Dr. R.J.A. Little that runs the analysis as described above. With modification, the program was compiled and run in conjunction with the Howells data sets.

In order to simulate a missing data pattern that approximates the pattern of missing measurements one might encounter in an archaeological sample of human crania, the patterns of missing measurements associated with 154 crania from the late Mississippian Averbuch site were employed. Five measurements, glabella-occipital length (GOL), maximum cranial breadth (XCB), bizygomatic breadth (ZYB), basion-bregma height

(BBH) and basion-nasion length (BNL), were collected on the Averbuch crania per the definitions provided in Howells (1973) (see Table 2) during the inventory process. However, bifrontal breadth (FMB) was not collected. Instead, minimum frontal breadth (WFB), "the direct distance between the two frontotemporale" (Moore-Jansen and Jantz 1989) was collected on the Averbuch series. For the purposes of this study and the missing data simulation, WFB is substituted for FMB. Due to the fragmentary nature of the Averbuch crania, measurements are frequently unobservable. After confirming that this pattern of missing data conforms to the MCAR or "missing completely at random" assumption as defined by Rubin (1976), a simulation is enacted whereby crania from the Howells samples are assigned patterns of missing measurements that correspond to patterns observed in the Averbuch crania.

This simulation originates with a "mask" matrix **M** where each row is a missing data pattern from the Averbuch series beginning with the most common pattern and concluding with the least common patterns. Ones (1) in the matrix represent observed variables, while zeros (0) indicate unobserved variables. A vector of cumulative frequencies is also formed for the patterns observed in Averbuch to be used in conjunction with the matrix **M** (see Figure 1). For example, according to the matrix **M** the first row represents a pattern where the first 3 measurements are observed while the last 3 measurements are unobserved. According to the vector of cumulative frequencies, this pattern occurs in 30% of the Averbuch crania. Similarly, the second row in the matrix **M** represents the second most common pattern, where all 6 measurements are observed,

Cum. Freq.	GOL	XCB	FMB	ZYB	BBH	BNL
0.30	1	1	1	0	0	0
0.47	1	1	1	1	1	1
0.62	1	1	1	0	1	1
0.72	0	0	1	0	0	0
0.81	0	1	1	0	0	0
0.86	0	1	0	0	0	0
0.90	1	0	1	0	0	0
0.92	1	1	0	0	0	0
0.94	1	0	0	0	0	0
0.95	0	0	1	1	1	1
0.96	1	0	1	0	1	1
0.97	1	0	0	1	0	1
0.98	0	1	1	1	0	0
0.99	0	1	0	0	1	0
1.00	0	0	0	1	0	0

Figure 1. Vector of cumulative frequencies and the matrix **M** for missing data simulations

which occurred in 17% of the crania as indicated by the .47 entry in the cumulative frequency vector.

For each cranium from the Howells data sets a random uniform deviate between 0 and 1 is generated. This value is compared to the cumulative frequency vector and the corresponding row from the matrix **M** is chosen. The missing data pattern associated with the row from the matrix **M** is then applied to the cranium from the Howells data set. For example, if a random uniform deviate of .35 is generated for a particular cranium from the Howells samples, then the missing data pattern associated with the second row of the matrix **M** would be applied to that skull. Hence, all measurements would be retained. Then, say for the next cranium in the Howells series if a random uniform deviate of 0.75 is generated, the corresponding pattern in the **M** matrix would be applied eliminating GOL, ZYB, BBH, and BNL from the data set for that cranium. One thousand data sets with missing data were generated for each of the two Howells samples following this procedure.

In order to test the utility of finite mixture analysis under a missing data model, each of the 1,000 data sets are subjected to analysis. Clustering is accomplished through classification by the posterior probability (> 0.50) that a specimen is male or female. The percentages of correct classification by the finite mixture analysis both with complete data sets and with missing data is compared to a discriminant function created from the Howells samples themselves and tested on a bootstrap sample derived from the sample itself. The correct classification for the discriminant analysis is taken from Efron's (1983) 0.632 estimator, a weighted sum of the apparent error rate which underestimates the true

error rate and the bootstrap error rate which overestimates the true error rate. Utilizing the 0.632 estimator eliminates bias due to reclassification in a small sample.

In a further test of the utility of finite mixture analysis for use with archaeological human skeletal material, the method is applied to crania from the Averbuch site. The Averbuch site (40DV60) located in Davidson County in the Nashville basin belongs to the Late Mississippian period (Berryman 1981). Excavations occurring from 1977-1978 recovered 887 individuals from 645 graves, the majority of which were located in 3 cemeteries associated with an adjacent village. Recent recalibration of 8 of the 9 radiocarbon (C_{14}) dates obtained from structures and non-burial features indicates that the site may have been occupied between the fourteenth and fifteenth centuries A.D. (Konigsberg and Frankenberg 1995).

Subsequent to their excavation, the burials were transported to the University of Tennessee for analysis and curation. One hundred fifty-four (154) of the Averbuch crania were sufficiently intact to permit observation of at least 1 of the following 6 craniometrics: glabella-occipital length (GOL), maximum cranial breadth (XCB), bizygomatic breadth (ZYB), basion-bregma height (BBH), basion-nasion length (BNL) as defined by Howells (1973) and minimum frontal breadth (WFB) as defined by Moore-Jansen and Jantz (1990). Any of the 6 craniometrics that were observable were collected, and the resultant data set investigated to determine if it is viable for analysis with the finite mixture missing data model.

One of the assumptions associated with finite mixture analysis under a missing data model is that the missing data are missing completely at random (MCAR) (Little 1988;

Little and Schluchter 1985). According to Rubin (1976) missing data is MCAR if for all observed values of y (y_{obs}) and missing values of y (y_{mis}):

$$f(\mathbf{r}|\mathbf{y}_{obs}, \mathbf{y}_{mis}, \psi) = f(\mathbf{r}|\psi)$$

where $\mathbf{r}$ is defined as a missingness indicator matrix such that $r_{ij} = 1$ if y_{ij} is missing and 0 otherwise and ψ represents the unknown parameters for the distribution. Basically this equation simply means that missing data is MCAR when “missingness does not depend on the observed or missing values of y ” (Little 1988). In other words, whether or not a particular measurement is missing does not depend on the value of that missing measurement. If the missing data for a particular multivariate analysis conforms to the MCAR definition, then the missing data mechanism is considered ignorable for likelihood-based inferences (Rubin 1976). In order to determine if the MCAR assumption can be met for a particular data set with missing values, Little (1988) proposed the following quadratic test:

$$d^2 = \sum_{j=1}^J \mathbf{m}_j (\bar{\mathbf{y}}_{obs, j} - \boldsymbol{\mu}_{obs, j}) \mathbf{V}_{obs, j}^{-1} (\bar{\mathbf{y}}_{obs, j} - \boldsymbol{\mu}_{obs, j})^T$$

$$\text{with df} = \sum p_j - p$$

where $\boldsymbol{\mu}$ and $\mathbf{V}$ are estimated via the EM algorithm, $\mathbf{m}$ is the number of cases with a particular missing data pattern and p is the number of observed variables for those cases. When the missing data pattern from the Averbuch sample was subjected to this test ($d^2 =$

45.87, $p\text{-value} = 0.1781$), the null hypothesis that the missing data conform to the MCAR assumption could not be rejected at the .15 level.

After confirming the validity of the MCAR assumption for the pattern of missing data found in the sample of 154 Averbuch crania, the sample for analysis was restricted to a subsample of 94 crania with associated *os coxae* that permitted the observation of at least one of the three Phenice (1969) criteria for determining sex from the *os pubis*. Based on this evaluation the 94 crania were assigned an “actual” or “known” sex resulting in a sample of 49 males and 45 females. The finite mixture missing data model described previously is applied to the incomplete data set associated with the 94 crania of “known” sex to determine sex for individuals from the restricted sample. The accuracy of the classification of a particular specimen is evaluated by comparing it to the “actual” sex as determined by the Phenice (1969) criteria.

CHAPTER 4

RESULTS

Finite mixture analysis of the 1000 simulated data sets with missing data for the Hokkaido sample from Northern Japan and comparison of those results with discriminant and finite mixture analysis with a complete data set are illustrated in Figure 2. The histogram that illustrates the distribution for the entire sample (Overall) also includes the correct classification for finite mixture and discriminant analysis with complete data sets. For the Hokkaido females, correct classification for the finite mixture missing data model across 1000 runs ranged from 34.4% to 100% with an average of 74.6%. Correct classification for the Hokkaido males ranged from 67.3% to 100% with an average of 92.2%. Overall, correct classification across the 1000 runs ranged from 72.4% to 94.3% with an average of 85.7%. This compares quite favorably to the complete data finite mixture correct classification of 88.5% and the discriminant correct classification of 88.4%.

However, the results for the Zulu sample from South Africa are not as promising. Finite mixture analysis of the 1000 simulated data sets with missing data for the Zulu sample and comparison of those results with discriminant and finite mixture analysis with a complete data set are illustrated in Figure 3. The histogram that illustrates the correct classification distribution for the entire sample (Overall) includes the correct classification

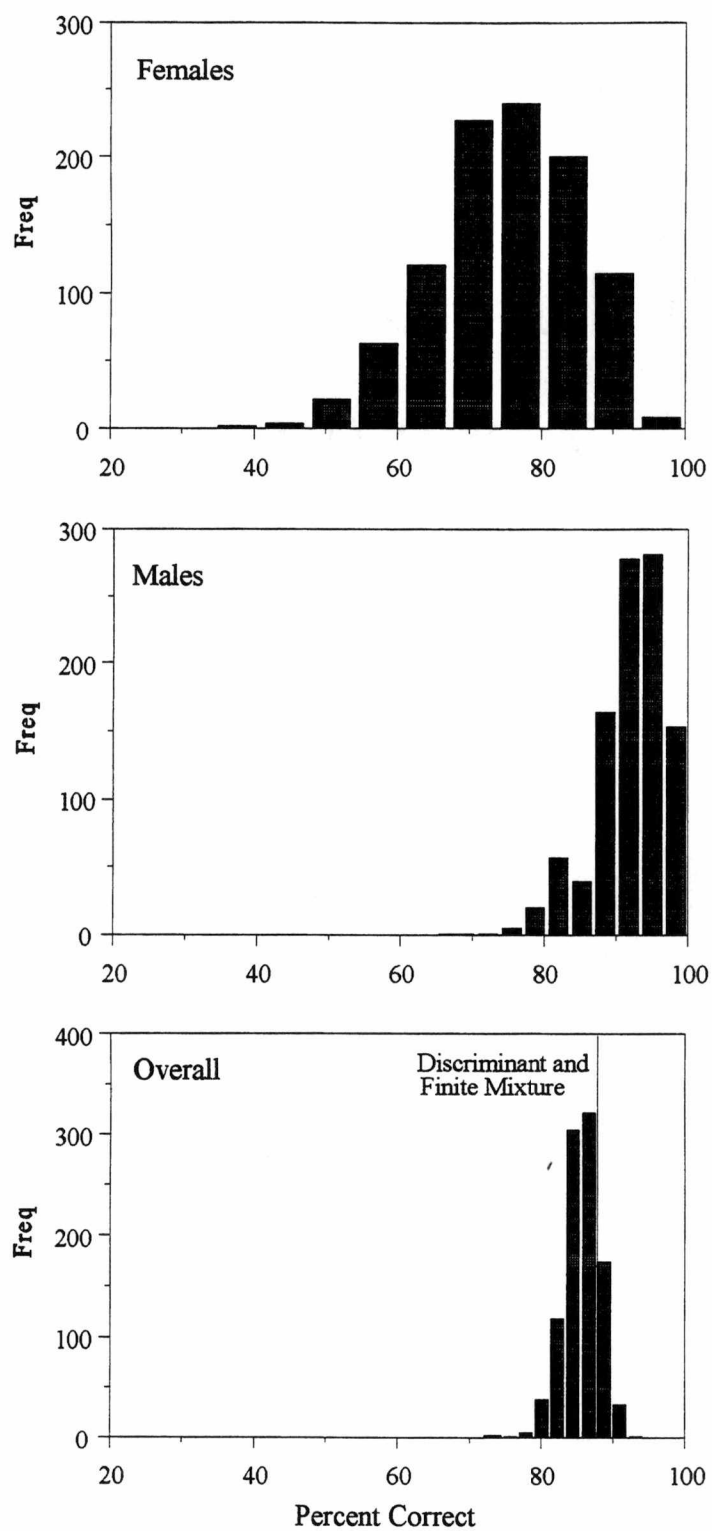


Figure 2. Results of the finite mixture with missing data simulation incorporating the Hokkaido sample from Northern Japan.

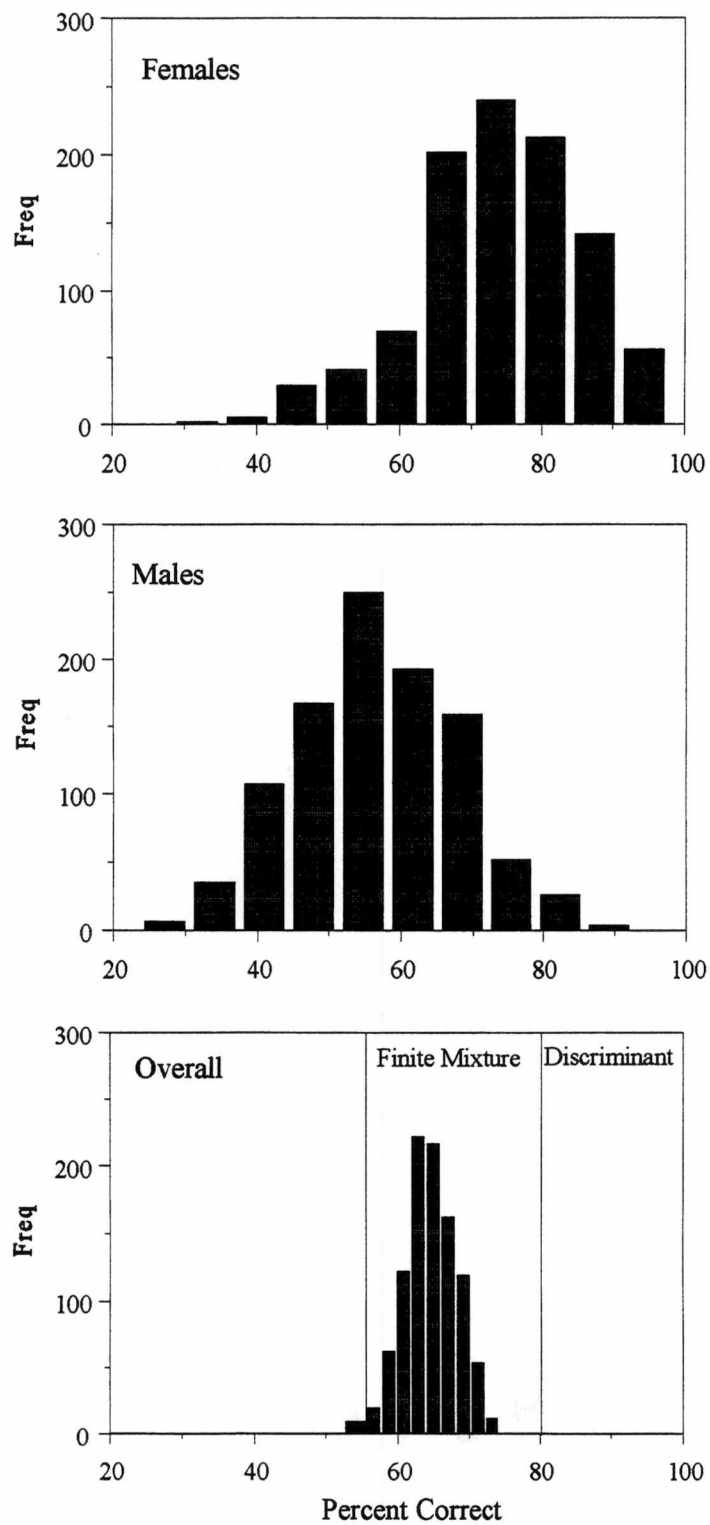


Figure 3. Results of finite mixture analysis with missing data simulation incorporating the Zulu sample from South Africa.

for finite mixture and discriminant analysis with the complete data sets. For the Zulu females, correct classification of the finite mixture missing data model over 1000 runs ranges from 28.3% to 97.8% with an average of 73.7%. The correct classification for the Zulu males ranges from 23.6% to 92.7% with an average of 56.8%. Overall, the correct classification across a 1000 runs ranges from 53.5% to 74.2% with an average of 64.5%, much lower than the discriminant correct classification of 80.1% but better than the complete data finite mixture correct classification of 55.4%. The implications of these seemingly contradictory results are dealt with in the Discussion section of this thesis.

The results of finite mixture analysis with missing data for the restricted Averbuch sample are evaluated at three different levels of significance (0.5, 0.8 and 0.9) for the posterior probabilities. As mentioned previously, subsequent to finite mixture analysis, individuals within a group are clustered by the posterior probability that they belong to a particular subgroup, in this instance male and female. The appendix provides a complete listing of the Averbuch crania and their individual posterior probabilities. Table 4 illustrates the structure of the Averbuch sample as predicted by finite mixture analysis with a general dividing point of 0.5 compared to the composition of the sample as determined by "actual" sex using the Phenice (1969) criteria. When using 0.5 as a dividing point for classifying individuals based on posterior probabilities, the probability (male or female) that is greater than 0.5 determines the sex of the individual. At the 0.5 level the differences between the actual and predicted number of males and females indicate the females are marginally overclassified and the males are marginally underclassified. This

Table 4. Observed vs. predicted sex classification for the restricted Averbuch sample when probability of the more likely sex was greater than 0.5

	Predicted males	Predicted females	Total
Actual males	35	14	49
Actual females	10	35	45
Total	45	49	94

overclassification of females results in a predicted sex ratio of 1.09 females: 1 male versus the real sex ratio of 1.09 males: 1 female. However, according to the McNemar test for correlated proportions (Thomas 1976), the difference between the predicted and actual sex ratio is not statistically significant ($p=.54$). At this level, 71% of the males and 78% of the females are correctly classified. Overall, 74% of Averbuch sample is correctly classified by finite mixture analysis. According to a Fisher's Exact Test the probability of these results occurring if the Phenice (1969) criteria and finite mixture analysis are independent is $p = 0.000002$.

When the significance level is increased to 0.8, individuals are classified as male or female when the corresponding posterior probability is greater than 0.8. In other words, if the posterior probability of being a given sex was between 0.2 and 0.8, then the individual was excluded at this level. Table 5 demonstrates the differences between the predicted and "actual" classification for the Averbuch sample when $P=0.8$. In this instance, the predicted sex ratio improves slightly as does the accuracy of the finite mixture analysis classification. The predicted sex ratio is 1.04 females: 1 male compared to the actual ratio of 1.04 males: 1 female. Once again, the McNemar test for correlated proportions (Thomas 1976) indicates that this difference is not statistically significant ($p=1.0$). At this level, 79% of the males and 81% of the females are correctly classified. The percentage of the entire sample that is correctly classified is 80%. According to a Fisher's Exact Test the probability of these results occurring if the Phenice (1969) criteria and finite mixture analysis are independent is $p = 0.000008$.

Table 5. Observed vs. predicted sex classification for the restricted Averbuch sample when probability of the more likely sex was greater than 0.8

	Predicted males	Predicted females	Total
Actual males	22	6	28
Actual females	5	22	27
Total	27	28	55

When the requirements become even more stringent by increasing the significance level to 0.9, accuracy of the predicted classifications also improves. At this level, if the posterior probability of being a given sex was between 0.1 and 0.9, then the individual was excluded. Table 6 illustrates the predicted classification compared to the "actual" classification for the Averbuch sample when $P=0.9$. At this level, 83% of the males and 89% of the females are correctly classified. For the entire sample, correct classification occurs for 85% of the specimens. According to a Fisher's Exact Test the probability of these results occurring if the Phenice (1969) criteria and finite mixture analysis are independent is $p = 0.000005$. The predicted sex ratio of 1.05 males: 1 female is comparable to the actual ratio of 1.27 males: 1 female, since the McNemar test for correlated proportions (Thomas 1976) indicates that there is no statistically significant difference between the two ratios ($p=.68$).

Table 6. Observed vs. predicted sex classification for the restricted Averbuch sample when probability of the more likely sex was greater than 0.9

	Predicted males	Predicted females	Total
Actual males	19	4	23
Actual females	2	16	18
Total	21	20	41

CHAPTER 5

DISCUSSION

Estimation of sex from the cranium depends on secondary sex characteristics generally involving differences in size. Weiss (1972) and Meindl et al. (1985) attribute bias in sex ratios to this dependence on size differences for visually diagnosing archaeological specimens. However, they contradict one another regarding the direction of error; Weiss' research indicates that males are overclassified, and Meindl and coworkers contend from their study that females are overclassified.

Metric approaches such as discriminant and finite mixture analysis utilize differences between males and females in a more objective manner to classify individuals. However, as mentioned, discriminant analysis generally requires a reference sample or training set which may not exhibit the same sexual dimorphism and/or overall "size" as the series under investigation. Finite mixture analysis alleviates the need for a reference sample by utilizing the differences observed in the data to create a discriminant function to classify specimens. Therefore, either method may not be as accurate for skeletal material that exhibits low sexual dimorphism particularly when the differences between males and females are low. Similarly, finite mixture should prove to be as accurate as other methods for sex estimation when levels of sexual dimorphism as expressed by differences in "size"

are high. The results of the analysis incorporating the two Howells samples presented in the previous chapter confirm these expectations.

The utility of employing finite mixture analysis with incomplete data sets was investigated in this thesis. It might be expected that analyses involving data sets with missing variables would be less accurate than analyses incorporating complete data sets for the same sample. However, since the EM algorithm allows for an unbiased estimation of those missing variables, the level of accuracy ought not decrease substantially.

As noted in Chapter 3, the Howells samples employed for the finite mixture missing data simulation differ in their levels of sexual dimorphism. The Hokkaido sample from Northern Japan exhibits a much higher level of "size" related sexual dimorphism than the Zulu sample from South Africa. This difference is borne out in the contrasting results for the finite mixture analyses both with and without missing data. Comparing the average correct classification for finite mixture analysis with complete data sets for both samples (88.5% for the Hokkaido and 55.4% for the Zulu) illustrates the effect of "size"-related sexual dimorphism on the accuracy of the method. The more dimorphic Hokkaido sample is correctly classified via finite mixture analysis at a percentage comparable to the discriminant analysis (88.4%) for the sample. Considering that the correct classification for the discriminant analysis is best considered an optimistic estimate of how such an analysis would work with an archaeological sample, the finite mixture analysis truly performed well. On the other hand, the poor correct classification rendered by finite mixture analysis of the Zulu sample when compared to the discriminant analysis (80.1%) is partially due to the low level of dimorphism exhibited by the sample. However, these poor

results are also related to the discriminatory power of the craniometrics used in the finite mixture analysis which will be discussed later.

The results of the missing data simulation over 1000 runs also reflects the differing levels of dimorphism observed in the two samples (see Table 7 for a summary). While the Hokkaido females were only correctly classified marginally better than the Zulu females, a dramatic difference is noted in the percent of Hokkaido and Zulu males correctly classified. This difference between the sample males is reflected in the difference of the overall correct classifications for the two samples where the Hokkaido sample is classified much more accurately than the Zulu sample. The higher degree of sexual dimorphism in the Hokkaido sample allowed for a much higher level of correct classification. The low level of dimorphism exhibited by the Zulu sample inhibited accuracy in classifying individuals and results in classification not much better than random chance. While this aspect of the differing classificatory accuracy seems rather straightforward it is necessary to note that the direction of error is different for the two samples.

The contradictory direction of error for the two samples may also be related to the differing levels of sexual dimorphism. For the Hokkaido sample, 25% of the females were misclassified as males while less than 8% of the males were misclassified as females. This direction of error in sex estimation resulting in females being misclassified as males is similar to the pattern reported by Weiss (1972). Contrary to that pattern, the Zulu sample exhibits error in the opposite direction with more males being misclassified as females. In this sample, 26% of the females were misclassified as males while 43% of the males were

Table 7. Summary of results for finite mixture missing data simulation

	Average Percent Correct Classification		
	Females	Males	Overall
Hokkaido	74.6%	92.2%	85.7%
Zulu	73.7%	56.8%	64.5%

misclassified as females. In their study, Meindl et al. (1985) noted that errors in visual sex determination tend to result in males being misclassified as females.

As mentioned above, a quite intriguing aspect of the finite mixture analysis results is the poor performance of the complete data set analysis for the Zulu sample. For the missing data simulation over 1000 runs, 64.5% of the entire sample was correctly classified. However, for the complete data finite mixture analysis only 55.4% of the sample was correctly classified. While multivariate discriminant analysis generally discriminates between groups better than univariate discriminant analysis (Godfrey and Marks 1991), this is not necessarily true for finite mixture analysis. Since finite mixture analysis relies on the differences found within the data set itself, variables that are relatively uniform across the distribution do not aid in the discrimination between groups in the mixture. In the case of the Zulu, many of the craniometrics exhibit very little dimorphism, or very little difference in "size" between males and females exists for those craniometrics. Including these craniometrics with low discriminatory power inhibits finite mixture analysis' ability to create discriminant functions that correctly classify individuals. To illustrate this point, consider the following situation. Finite mixture analysis is performed on the Zulu sample utilizing only bifrontal breadth (FMB) and bizygomatic breadth (ZYB) correctly classifying 76% of the sample. This is already better than the previous correct classification of 55.4% for the sample with all 6 craniometrics employed. However, if bifrontal breadth (FMB) is removed from the analysis, leaving bizygomatic breadth (ZYB) as the sole variable under consideration, the correct classification increases to 79% which compares favorably to the discriminant analysis correct classification for the

entire data set of 80.1% . By eliminating variables that due to their low discriminatory power contribute little to the separation of groups within a sample, finite mixture analysis becomes more accurate even for samples exhibiting low levels of sexual dimorphism. As it happens, the patterns of missing data from the Averbuch sample used in the missing data simulations often eliminated variables that contributed little information resulting in more accurate classification by finite mixture analysis.

The application of finite mixture analysis for the purposes of determining sex to prehistoric skeletal material from the Averbuch site provide promising results. Table 8 presents a summary of the percent correct classification for the finite mixture analysis for the three levels of significance (0.5, 0.8 and 0.9). As the level of significance increases so does the accuracy of classification for males, females and the entire sample. At each level, more males are misclassified as females than vice versa. This direction of error is similar to that noted by Meindl et al. (1985). At the 0.5 level which includes the entire sample of 94 crania females are overclassified and males are underclassified. However, the sex ratio that results from the predicted classifications is not significantly different than the actual sex ratio. Even though 26% of the sample was misclassified, the sex ratio that was predicted by finite mixture analysis is comparable to the actual sex ratio.

As the posterior probability for classification is required to be greater than 0.8, the classification accuracy for males and females increases. The percent correct classification for the entire sample increases to 80%. At this level, one more male is misclassified as a female than vice versa (see Table 5). Once again, while a percentage of the sample was misclassified the predicted sex ratio is not significantly different than the actual sex ratio.

Table 8. Summary of the percent correct classification for the finite mixture analysis of the Averbuch sample at the three levels of significance

Significance Level	Percent Correct Classification		
	Males	Females	Overall
0.5	71%	78%	74%
0.8	79%	81%	80%
0.9	83%	89%	85%

However, using this level for classification restricts the sample to 55 crania. The final increase in the required posterior probability for classification to 0.9 also results in the increase in percent correct classification for male and females. Overall, 85% of the limited sample of 41 crania were correctly classified. As with the two previous classification levels, more males are misclassified as females resulting in a slight overclassification of females. Even though 15% of the sample was misclassified, the predicted sex ratio is not significantly different than the actual ratio.

It should be noted that 78% of the 94 Averbuch crania exhibit deformation in the form of occipital flattening due to cradle boarding. Generally the deformation is mild but in a few cases the alteration is more severe. Hence, the occipital flattening associated with the Averbuch crania potentially affects the maximum cranial breadth measurements (XCB) in a small percentage of the crania. This could, in turn, affect the discriminatory power of this variable when utilized with finite mixture analysis.

While the baseline assessment of the restricted Averbuch sample at 0.5 only provides a correct classification of 74%, it should be considered that the sample exhibits the missing data patterns utilized in the missing data simulation with the Howells samples. Increasing the level of significance necessary for classifying individuals improves the correct classification to 80% at 0.8 and 85% at 0.9, both comparable to the levels of correct classification for prehistoric material achieved by Giles and Elliot (1963). When the results for the Averbuch material are compared to the results for the Howells samples with missing data, a more accurate picture of the efficacy of finite mixture analysis is developed. Since the Howells samples were all classified at the 0.5 level, the comparison

incorporates the Averbuch results at the 0.5 level as well. For the more dimorphic Hokkaido sample, on average 85.7% of the entire sample was correctly classified and on average 64.5% of the entire less dimorphic Zulu sample was correctly classified. The Averbuch sample with an overall correct classification of 74% falls in between the two Howells samples.

CHAPTER 6

CONCLUSIONS

Sex and age-specific mortality estimates are the basic data upon which many paleodemographic inferences are built (Buikstra and Mielke 1985). The accuracy of such estimations has grave implications for the appropriateness of any interpretations regarding the population under investigation. Correctly assessing the sex of individual skeletal remains and predicting the sex ratio is critical to understanding the structure of the series. Past research (Weiss 1972; Meindl et al. 1985) has suggested that both males and females are susceptible to underclassification that results in skewed sex ratios. Current methods and techniques for determining sex from human skeletal material introduce bias into the assessment process, whether they are visual or metric in application. Both approaches generally require previous knowledge of expected expression of sexual dimorphism and too often this knowledge is generated from research with other samples not necessarily comparable in levels of dimorphism and/or overall size. Additionally, metric approaches are generally designed for complete data sets and naive methods for dealing with missing data can introduce their own biases. As an alternative to the more commonly multivariate analyses employed in anthropology, finite mixture analysis does not necessitate a reference sample and allows for missing data. Instead finite mixture analysis utilizes the "size"

differences that exist within the data to define parameters for classifying specimens and generates unbiased estimations of missing values to impute where variables are absent.

When this analysis was applied to two of Howells (1973) known sex samples with and without missing data, accuracy depended on the level of size related sexual dimorphism exhibited by the sample. Finite mixture analysis with and without missing data of the more dimorphic of the two samples, the Hokkaido from Northern Japan, provided levels of accuracy comparable to discriminant analysis. However, analysis of the less dimorphic Zulu from South Africa resulted in dismal levels of accuracy for the missing data simulation and even worse results for the complete data analysis. The inclusion of craniometrics with low discriminatory power in the analysis of the complete Zulu data set rendered it unable to predict sex much better than random chance. When the analysis was restricted to a single variable, the accuracy increased to levels comparable to the percent correct classification achieved through discriminant analysis of the entire Zulu data set. For samples with low levels of sexual dimorphism eliminating non-dimorphic metrics could improve the accuracy of the classification.

While finite mixture analysis misclassified a percentage of the Averbuch sample, it predicted a sex ratio comparable to the actual sex ratio. Considering the prevalence of missing data with the Averbuch data set, the overall correct classification of 74% with the general dividing point of 0.5 for the posterior probabilities is respectable. Additionally, a majority of the Averbuch crania were deformed which potentially affected one of the measurements and hindered the finite mixture discriminatory process.

Overall, finite mixture analysis works well for samples that express relatively high levels of sexual dimorphism and when that dimorphism is primarily due to size. For such a sample, finite mixture provides classifications nearly as accurate as discriminant analysis derived from the sample itself both with a complete data set and with missing data. However, when dimorphism is relatively low in a sample, finite mixture does not provide very accurate classification. For samples where pelvic elements are unavailable for creating a training subset as necessitated by discriminant analysis, finite mixture may prove to be a more appropriate and equally accurate method which does not require applying standards from known sex material that bears no particular relationship to the archaeological sample under investigation.

Potential future applications include sex estimation from postcranial measurements and species identification from a variety of metric variables. Devising approaches for determining which variables should be included in finite mixture analysis to achieve the most accurate results possible is another avenue for future research. As always, alternative methods for determining sex from human skeletal material that eliminates biases inherent in current methodologies need to be developed and evaluated.

LITERATURE CITED

LITERATURE CITED

- Bass WM (1987) *Human Osteology: A Laboratory and Field Manual*. Columbia: Missouri Archaeological Society.
- Berrizbeitia EL (1989) Sex determination with the head of the radius. *J. For. Sci.* 34:1206-1213.
- Berryman HE (1981) *The Averbuch Skeletal Series: A Study of Biological and Social Stress at a Late Mississippian Period Site from Middle Tennessee*. Unpublished Doctoral Dissertation, Department of Anthropology, University of Tennessee, Knoxville.
- Birkby WH (1966) An evaluation of race and sex identification from cranial measurements. *Am. J. Phys. Anthropol.* 24:21-28.
- Black III TK (1978) A new method for assessing the sex of fragmentary skeletal remains: femoral shaft circumference. *Am. J. Phys. Anthropol.* 48:227-232.
- Brant LJ, Morrell CH and Pearson JD (1990) Mixture models for investigating complex distributions. *Am. J. Phys. Anthropol.* 81:198 (abstract).
- Buikstra JE and Mielke JH (1985) Demography, Diet and Health. In RI Gilbert and JH Mielke (eds.): *The Analysis of Prehistoric Diets*. New York: Academic Press. pp. 359-422.
- Burnaby TP (1966) Growth-invariant discriminant functions and generalized size differences. *Biometrics* 22:96-110.
- Cunha E and van Vark GN (1991) The construction of sex discriminant functions from a large collection of skulls of known sex. *Intl. J. of Anthropol.* 6:53-66.
- DiBernardo R and Taylor JV (1979) Sex assessment of the femur: a test of a new method. *Am. J. Phys. Anthropol.* 50:635-638.
- Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist. Assoc.* 78:316-331.
- Giles E and Elliot O (1962) Race identification from cranial measurements. *J. For. Sci.* 7:147-157.

- Giles E and Elliot O (1963) Sex determination by discriminant function analysis of crania. *Am. J. Phys. Anthropol.* 21:53-68.
- Godfrey LR, Lyon SK and Sutherland MR (1993) Sexual dimorphism in large-bodied primates: the case of the subfossil lemurs. *Am. J. Phys. Anthropol.* 90:315-334.
- Godfrey L and Marks J (1991) The nature and origins of primate species. *Yrbk. of Phys. Anthropol.* 34:39-68.
- Hanihara K (1959) Sex diagnosis of Japanese skulls and scapulae by means of discriminant function. *J. Anthrop. Soc. Nippon.* 67:191-197.
- Henke W (1977) On the method of discriminant function analysis for sex determination of the skull. *J. Hum. Evol.* 6:95-100.
- Holland TD (1986) Sex determination of fragmentary crania by analysis of the cranial base. *Am. J. Phys. Anthropol.* 70:203-208.
- Holman DJ and Bennett KA (1991) Determination of sex from arm bone measurements. *Am. J. Phys. Anthropol.* 84:421-426.
- Howells WW (1973) *Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference among Recent Human Populations.* Papers of the Peabody Museum of Archaeology and Ethnology. Vol. 67. Cambridge, Mass: Harvard University.
- Houghton P (1974) The relationship of the pre-auricular groove of the ilium to pregnancy. *Am. J. Phys. Anthropol.* 41:381-390.
- Iscan MY and Miller-Shaivitz P (1986) Sexual Dimorphism in the Femur and Tibia. In K Reichs (ed.): *Forensic Osteology: Advances in the Identification of Human Remains.* Springfield, IL: Charles C. Thomas. pp. 101-111.
- Iscan MY, Yoshino M and Kato S (1994) Sex determination from the tibia: standards for contemporary Japan. *J. For. Sci.* 39:785-792.
- Josephson SC, Juell KE and Rogers AR (1996) Estimating sexual dimorphism by method-of-moments. *Am. J. Phys. Anthropol.* 100:191-206.
- Kajanoja P (1966) Sex determination of Finnish crania by discriminant function analysis. *Am. J. Phys. Anthropol.* 24:29-34.

- Keen JA (1950) A study of the differences between male and female skulls. *Am. J. Phys. Anthropol.* 8:65-79.
- Konigsberg LW and Frankenberg SR (1995) Applications of Monte Carlo methods to radiocarbon dating problems in the Southeast. Paper presented at the 52nd Annual Meeting of the Southeastern Archaeological Conference, Knoxville, TN.
- Kramer A and Konigsberg LW (1996) Recognizing species diversity among hominoids: a simulation test using missing data finite mixture analysis. *Am. J. Phys. Anthropol. Suppl.* 22:142 (Abstract).
- Krogman WM (1962) *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C. Thomas.
- Little RJA (1988) A test of missing completely at random for multivariate data with missing values. *J Am. Stat. Assoc.* 83:1198-1202.
- Little RJA and Rubin DB (1987) *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.
- Little RJA and Schenker N (1995) Missing Data. In G Arminger, CC Clogg and ME Sobel (eds.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press. pp. 39-69.
- Little RJA and Schluchter MD (1985) Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika* 72:497-512.
- Loth SR and Henneberg M (1996) Mandibular ramus flexure: a new morphologic indicator of sexual dimorphism in the human skeleton. *Am. J. Phys. Anthropol.* 99:473-485.
- McLachlan GJ and Basford KE (1988) *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Drekker, Inc.
- Meindl RS, Lovejoy CO, Mensforth RP and Don Carlos L (1985) Accuracy and direction of error in the sexing of the skeleton: implications for paleodemography. *Am. J. Phys. Anthropol.* 68:79-85.
- Moore-Jansen PH and Jantz RL (1989) *Data Collection Procedures for Forensic Skeletal Material*. Report of Investigations No. 48. Knoxville: University of Tennessee, Department of Anthropology.

- Pearson JD, Morrell CH and Brant LJ (1992) Mixture models for investigating complex distributions. *J. Quant. Anthropol.* 3:325-345.
- Plavcan JM (1994) Comparison of four simple methods for estimating sexual dimorphism in fossils. *Am. J. Phys. Anthropol.* 94:465-476.
- Phenice TW (1969) A newly developed visual method of sexing the os pubis. *Am. J. Phys. Anthropol.* 30:297-302.
- Rubin DB (1976) Inference and missing data. *Biometrika* 63:581-592.
- Santos R (1981) Effects of imputation on regression coefficients. In *Proc. Survey Res. Meth. Sec., Am. Statist. Assoc.* pp. 140-145.
- Scheuer JL and Elkington NM (1993) Sex determination from metacarpals and the first proximal phalanx. *J. For. Sci.* 38:769-778.
- Snow CG, Hartman S, Giles E and Young FA (1979) Sex and race determination of crania by calipers and computer: a test of the Giles and Elliot discriminant functions in 52 forensic cases. *J. For. Sci.* 24:448-460.
- Steele, DG (1976) The estimation of sex on the basis of the talus and calcaneus. *Am. J. Phys. Anthropol.* 45:581-588.
- Stewart TD (1979) *Essentials of Forensic Anthropology: Especially as Developed in the United States.* Springfield, IL: Charles C. Thomas.
- Sutherland LD and Suchey JM (1991) Use of the ventral arc in pubic sex determination. *J. For. Sci.* 36:501-511.
- Thieme FP and Schull WJ (1957) Sex determination from the skeleton. *Hum. Biol.* 29:242-273.
- Thomas DH (1976) *Figuring Anthropology: First Principles of Probability and Statistics.* New York: Holt, Rinehart and Winston.
- Ubelaker DH (1989) *Human Skeletal Remains: Excavation, Analysis, Interpretation.* Washington, DC: Smithsonian Institution.
- van Vark GN and Schaafsma W (1992) Advances in Quantitative Analysis of Skeletal Morphology. In SR Saunders and MA Katzenberg (eds.): *Skeletal Biology of Past Peoples: Research Methods.* New York: Wiley-Liss, Inc. Pp 225-227.

Washburn S (1948) Sex differences in the pubic bone. *Am. J. Phys. Anthropol.* 6:199-207.

Weiss KM (1972) On the systematic bias in skeletal sexing. *Am. J. Phys. Anthropol.* 37:239-250.

Wu L (1989) Sex determination of Chinese femur by discriminant function. *J. For. Sci.* 34:1222-1227.

APPENDIX

Burial Number	P(male)	P(fem)	Morpho- logical Sex	Cranometrics							Def
				GOL	XCB	WFB	ZYB	BBH	BNL		
101	0.066	0.934	M	153	153	98	0	142	91	Y	
102	0.735	0.265	F	166	148	93	0	0	0	Y	
103	1.000	0.000	M	175	146	101	151	150	102	Y	
105	0.967	0.033	M	183	0	98	0	0	0	Y	
106	1.000	0.000	M	171	151	102	145	147	109	Y	
107A	0.350	0.650	M	156	147	92	0	0	0	Y	
121A	0.491	0.509	M	0	0	92	0	0	0	N	
121B	0.142	0.858	M	158	142	84	0	0	0	Y	
122	0.161	0.839	F	0	0	85	0	0	0	Y	
123A	0.424	0.576	F	0	143	91	0	0	0	Y	
137A	0.698	0.302	M	167	144	94	0	0	0	Y	
156B	0.925	0.075	M	167	153	100	0	0	0	Y	
158A	0.021	0.979	F	147	153	89	0	138	94	Y	
160B	0.008	0.992	M	145	141	86	130	0	0	Y	
164	0.237	0.763	F	0	144	87	0	0	0	Y	
170	0.189	0.811	F	164	133	0	0	0	0	Y	
171	0.592	0.408	M	0	0	88	138	145	96	N	
181	0.950	0.050	M	161	147	97	142	0	0	N	
185A	0.626	0.374	M	0	153	94	0	0	0	Y	
20	0.236	0.764	F	166	130	91	0	0	0	Y	
203A	0.892	0.108	M	168	0	99	0	0	0	N	
207	0.434	0.566	F	0	0	91	0	0	0		
218	0.696	0.304	F	163	147	96	0	0	0	Y	
220B	0.653	0.347	M	168	145	90	0	0	0	Y	

225A	0.933	0.067	M	186	138	92	0	0	0	Y
229B	0.998	0.002	M	160	156	96	144	140	104	Y
255	0.491	0.509	F	0	0	92	0	0	0	Y
264A	0.446	0.554	F	158	146	94	0	0	0	N
264B	0.686	0.314	F	0	156	95	0	0	0	
271A	0.000	1.000	F	156	150	86	126	144	99	Y
277A	0.476	0.524	F	0	146	0	0	0	0	Y
280A	0.099	0.901	F	163	129	83	0	139	101	Y
288A	0.622	0.378	F	156	152	97	0	0	0	Y
291B	0.934	0.066	F	172	152	96	0	0	0	N
293	0.031	0.969	F	166	141	96	0	131	97	Y
297	0.970	0.030	M	177	150	99	0	0	0	Y
301A	0.176	0.824	F	0	144	0	0	137	0	Y
311B	0.362	0.638	M	0	139	0	0	0	0	Y
315A	0.791	0.209	M	162	155	94	0	0	0	Y
328	0.913	0.087	M	181	0	94	0	0	0	Y
332	1.000	0.000	M	162	147	94	148	145	101	N
335A	0.958	0.042	M	163	156	94	0	147	103	Y
341A	0.880	0.120	M	176	144	93	0	0	0	N
349	0.351	0.649	M	156	146	93	0	0	0	Y
353B	0.326	0.674	M	0	0	89	0	0	0	
356	0.000	1.000	F	150	150	94	133	132	94	Y
374	0.236	0.764	M	156	147	94	0	139	97	Y
377	1.000	0.000	M	164	159	99	142	150	105	Y
417A	0.072	0.928	F	159	147	90	0	142	96	Y
433	0.880	0.120	F	171	0	97	0	0	0	Y

490	0.013	0.987	M	172	144	91	129	146	102	
505	0.395	0.605	F	169	133	91	0	0	0	
525	0.424	0.576	M	0	143	91	0	0	0	Y
531	0.914	0.086	F	170	141	93	136	141	104	N
560	0.184	0.816	M	148	0	0	0	0	0	Y
565	0.544	0.456	M	0	150	0	0	0	0	Y
570	0.151	0.849	F	157	145	83	0	0	0	N
403	0.993	0.007	M	166	146	91	0	153	105	Y
524	0.000	1.000	F	151	132	83	122	134	96	Y
597A	0.201	0.799	M	158	146	91	0	145	96	Y
627	0.060	0.940	F	146	149	93	135	0	0	Y
194B	0.000	1.000	F	137	144	92	0	130	86	Y
220A	0.339	0.661	F	165	137	90	0	0	0	Y
16	0.041	0.959	F	152	145	88	0	136	97	Y
394A	0.490	0.510	F	164	143	91	0	0	0	N
539	0.000	1.000	F	145	155	93	133	136	91	Y
498	0.000	1.000	F	162	126	83	127	130	97	N
515	0.781	0.219	M	166	147	93	0	143	102	N
130A	0.227	0.773	F	158	142	89	0	0	0	Y
653	0.491	0.509	M	0	0	92	0	0	0	?
123B	0.977	0.023	M	171	148	98	141	144	102	Y
137B	0.491	0.509	F	161	149	89	0	0	0	Y
142A	0.059	0.941	F	147	140	91	0	0	0	Y
585	0.999	0.001	M	150	147	92	142	145	99	Y
517B	0.011	0.989	F	0	147	88	129	0	0	Y
405	0.709	0.291	M	0	0	96	0	0	0	

455	0.997	0.003	M	160	152	89	146	0	0	Y
480	0.618	0.382	M	0	151	94	0	0	0	Y
463A	0.139	0.861	F	150	146	0	0	0	0	Y
471	0.791	0.209	M	158	157	97	0	0	0	Y
527	0.519	0.481	M	173	138	90	133	0	0	N
546	1.000	0.000	M	161	154	97	148	153	98	Y
341B	0.390	0.610	F	162	0	96	0	145	96	Y
390A	0.623	0.377	F	155	155	89	0	148	99	Y
406	0.236	0.764	F	151	150	91	0	136	99	Y
418	0.820	0.180	F	164	0	98	0	0	0	Y
421	0.000	1.000	F	159	146	82	126	141	97	Y
430A	0.849	0.151	M	173	136	102	0	0	0	Y
442A	0.536	0.464	M	160	153	88	0	0	0	Y
444	0.498	0.502	M	0	148	92	0	0	0	Y
533	0.132	0.868	F	164	142	86	0	132	104	Y
551	0.981	0.019	M	183	130	95	136	145	104	Y
553B	0.803	0.197	F	157	149	92	141	139	98	Y
521	1.000	0.000	M	0	0	92	150	137	107	

VITA

Ashley Hyatt McKeown was born in Atlanta, Georgia on August 22, 1969 and moved to Chapel Hill, North Carolina 10 days later. A few years later, her family moved to Albany, New York where she attended first grade. Eventually, Ashley's family settled in Athens, Georgia where she attended elementary through high school graduating in May of 1987. The following August, Ashley matriculated at the University of North Carolina at Chapel Hill where she majored in Anthropology. Ashley graduated from UNC in May of 1991. In the fall of 1992 she entered the graduate program of the Department of Anthropology at the University of Tennessee, Knoxville. At UT, Ashley specializes in the subdiscipline of physical anthropology.

Upon graduating with a Masters of Arts with a major in Anthropology in December 1996, Ashley will begin the doctoral program in the UT Anthropology department.