To the Graduate Council:

I am submitting herewith a thesis written by Samantha Marie Hens entitled "The Use of Ordinal Categorical Variables in Skeletal Assessment of Sex." 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 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

THE USE OF ORDINAL CATEGORICAL VARIABLES IN SKELETAL ASSESSMENT OF SEX

A Thesis

Presented for the

Master of Arts

Degree

The University of Tennessee, Knoxville

SAMANTHA MARIE HENS

December, 1994

DEDICATION

This thesis is lovingly dedicated to the members of my family: Joe, Monica, David and especially my mother, Elaine. They have stood by me through many years and provided me with encouragement, support and love. For this, no simple thank you could ever be enough. I miss you and love you all.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee: Dr. Andrew Kramer, Dr. John Philpot and my chair, Dr. Lyle Konigsberg. They have provided me with endless guidance and assistance in many aspects of Anthropology and Statistics. I appreciated their patience with my last minute preparations and their valuable comments on an earlier version of this manuscript.

Anyone in the Anthropology department will recognize my debt to Dr. Lyle Konigsberg, without whom which this thesis would never have come to be. His knowledge of statistics and his patience in explaining it all to me seemed to be endless.

I cannot forget to thank Ms. Parvene Hamzavi who assisted me greatly as the second observer for the interobserver repeatability study.

Lastly, my thanks go to Ms. Ann Reed of the U.T. Computing Center whose help with the statistical programming was essential.

This thesis was funded in part by a National Science Foundation grant to Dr. Lyle Konigsberg.

ABSTRACT

Ordinal categorical variables are frequently used as a basis for assessing the sex of adult archaeological skeletal remains. In this thesis, seven commonly used pelvic and cranial characteristics are scored on a five point ordinal scale by two independent observers from a sample of 148 individuals taken from the Averbuch skeletal collection (40DV60). The scores are compared using ordinal log-linear models appropriate for measuring interobserver agreement. The results show that most of the trait observations are highly correlated between the two observers, but that a few traits do not exhibit marginal homogeneity (ie. the observers disagree in the proportions of individuals that should be placed in each ordinal class.)

The next step involved considering the seven characteristics across "known" sex as determined using Phenice's pubic bone characteristics, and through the addition of a modern collection of known sex. Logistic regression of sex onto the skeletal characters is used in order to find equations that provide posterior probabilities that a skeleton with a given set of characters was male as versus female.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION..... 1

2. MATERIALS AND METHODS..... 7

3. RESULTS..... 20

4. DISCUSSION..... 49

5. SUMMARY AND CONCLUSIONS..... 59

LIST OF REFERENCES..... 63

APPENDICES..... 68

APPENDIX A: EXAMPLE PHOTOGRAPHS OF EACH TRAIT SCORED ON STRONG MALES AND STRONG FEMALES..... 69

APPENDIX B: CROSS-CLASSIFICATION OF OBSERVER RATINGS ON EACH VARIABLE FROM THE AVERBUCH SITE (40DV60) N=148..... 75

APPENDIX C: SAS PROGRAM FOR THE LOGISTIC REGRESSION ON THE DATA USED IN THIS THESIS..... 88

APPENDIX D: RAW DATA FOR THE LOGISTIC REGRESSION ANALYSIS..... 90

VITA..... 101

LIST OF TABLES

TABLE PAGE

1. SAMPLE SIZES FOR EACH TRAIT SELECTED FROM THE AVERBUCH SKELETAL SERIES..... 9
2. SUMMARY OF THE CRITERIA USED IN SCORING MALES VERSUS FEMALES..... 10
3. SUMMARY OF THE RESULTS FROM THE TEST OF THE AGREEMENT PLUS UNIFORM-ASSOCIATION MODEL AGAINST THE SATURATED MODEL (WITH 14 DF)..... 21
4. MAXIMUM LIKELIHOOD PARAMETER ESTIMATES FOR THE AGREEMENT PLUS UNIFORM-ASSOCIATION MODEL..... 23
5. SUMMARY OF THE RESULTS FROM THE TEST OF THE UNIFORM-ASSOCIATION MODEL AGAINST THE SATURATED MODEL (WITH 15 DF)..... 25
6. SUMMARY OF THE RESULTS OF THE TEST OF THE DIAGONAL MODEL AGAINST THE SATURATED MODEL (WITH 15 DF)..... 26
7. SUMMARY OF THE RESULTS FROM THE TEST OF THE INDEPENDENCE MODEL AGAINST THE SATURATED MODEL (WITH 16 DF)..... 28
8. SUMMARY OF THE RESULTS FROM THE TEST OF THE UNIFORM-ASSOCIATION MODEL VS. THE AGREEMENT PLUS UNIFORM-ASSOCIATION MODEL (WITH 1 DF)..... 30
9. TEST OF MARGINAL HOMOGENEITY..... 31

LIST OF TABLES - CONTINUED

TABLE		PAGE
10.	ESTIMATES FOR THE FORCED MODEL, ALL VARIABLES, COMBINED DATA SET (WITH 1 DEGREE OF FREEDOM).....	33
11.	STEPWISE SELECTION ON ALL VARIABLES FOR THE COMBINED DATA SET.....	34
12.	VARIABLES NOT FOUND IN TABLE 11.....	34
13.	STEPWISE SELECTION ON THE CRANIAL VARIABLES FOR THE COMBINED DATA SET.....	35
14.	VARIABLES NOT FOUND IN TABLE 13.....	35
15.	STEPWISE SELECTION ON ALL VARIABLES FOR THE AVERBUCH SAMPLE ONLY.....	37
16.	VARIABLES NOT FOUND IN TABLE 15.....	37
17.	STEPWISE SELECTION ON THE CRANIAL VARIABLES FOR THE AVERBUCH SAMPLE.....	38
18.	VARIABLES NOT FOUND IN TABLE 17.....	38
19.	OBSERVED VS. PREDICTED VALUES OF ALL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P=0.5.....	41
20.	OBSERVED VS. PREDICTED VALUES OF ALL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.8 OR P<0.2..	42
21.	OBSERVED VS. PREDICTED VALUES OF ALL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.9 OR P<0.1..	43
22.	OBSERVED VS. PREDICTED VALUES OF CRANIAL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P=0.5.....	45

LIST OF TABLES - CONTINUED

TABLE PAGE

23. OBSERVED VS. PREDICTED VALUES OF CRANIAL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.8 OR P<0.2.. 47

24. OBSERVED VS. PREDICTED VALUES OF CRANIAL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.9 OR P<0.1.. 48

25. CONDENSED CHART OF MARGINAL HOMOGENEITY..... 51

CHAPTER 1

INTRODUCTION

Sex determination in skeletal samples is a focal point of bioarchaeological research. Anthropologists are frequently concerned with mortality and morbidity patterns by sex, as well as biases that may occur in sex ratio estimates. Results may depend heavily on the completeness of the skeletal remains. Krogman (1962) recognized this and reported results varying from 90% accuracy from the skull alone, to 95% accuracy from the pelvis alone, and 98% correct sexing when both the skull and pelvis are used.

Traditionally, sex assessment of skeletal remains is through visual (subjective) or metric (objective) methods. Metric methods are argued to be easier to teach, more reliable and more repeatable than visual methods (Giles and Elliot, 1963; Washburn, 1948; Weiss, 1972). However, proponents for the visual techniques of sex determination argue that metric methods usually require excellent skeletal preservation and limited or no artificial cranial deformation (Meindl et al., 1985). Limitations of both the visual and metric methods are with the morphological and morphometric standards used in sex differentiation, which may differ with the population samples being studied.

Standards are not normally interchangeable and should be used with reference to the group upon which they are based (Krogman, 1962).

A survey of some of the more significant visual and metric studies that have been done would be relevant at this point. Meindl and coworkers (1985) examined pelvic and cranial characteristics of one hundred adult skeletons from the Hamann-Todd Collection. They found a 96% correct sexing rate on the pelvis alone and a 92% correct sexing rate on the skull alone. When the two anatomical regions were taken together, there was 97% accuracy in determining sex. The authors suggest that it is helpful if the sexual dimorphism of the particular population under study is known.

Other visual sexing techniques include the work by Houghton (1974) with the pre-auricular groove and Phenice (1969) with a visual assessment of the os pubis. Houghton examined the pre-auricular groove inferior to the auricular surface of the pelvis in 119 skeletons. He distinguished between the groove of pregnancy (GP) and the groove of the ligament (GL). Eighty-one percent of his male sample had grooves and all of those were GL. However, while 92% of the female sample had grooves, 71% were GP and 21% were GL. He concluded that the GL is the normal imprint of the ligament on pelves of both sexes, while GP is the result of the physiological and pathological changes that occur during

pregnancy and labor. Studies by Holt (1978) and Suchey et al. (1979) on dorsal pubic pitting have shown that the defects have been found in females who have never been pregnant. The scars were shown to have a statistically significant correlation to age regardless of parity.

Phenice described a method of sexing the pelvis which is accurate, rapid and objective enough for the beginning researcher. Phenice examined a sample of 275 adult individuals from the Terry Collection and estimated sex according to three critera: the ventral arc, subpubic concavity and ischio-pubic ramus. He found an accuracy of 96% for the total sample when all three characteristics were combined. Sutherland and Suchey (1991) expanded Phenice's study with a sample of 1,281 pubic bones which included adolescents. Their focus was on the ventral arc which, alone, provided 96% accuracy in sex determination.

Recent research by Rogers and Saunders (1994) reports on the accuracy and intraobserver reliability of seventeen individual morphological traits commonly used in sex determination of the pelvis. They examined the accuracy of the traits in determining sex, intraobserver error and direction of sex bias. The sample (N=49) was taken from a nineteenth century cemetery from the St. Thomas Anglican Church in Belleville, Canada where the individual sex was known from names on coffin plates and parish registers. The

overall degree of intraobserver error was 11.3% for all of the pelvic traits in combination. This suggests difficulty in observing some traits. The authors found 95-98% accuracy in sexing with different combinations of two and three traits together.

Metric sexing of the skeleton has been done through both univariate and multivariate procedures. Black (1978) analyzed 114 skeletons from the Libben site in Ohio using midshaft femoral circumference for the determination of sex. He found 85% accuracy and believed the technique might be an important aid in sexing fragmentary skeletal remains where such fragile areas as the cranium and pelvis may be broken or worn. Two other researchers (DiBennardo and Taylor, 1979) tested this method on a sample of North American Whites (N=115) and found 82% accuracy in sex determination. They believed femoral shaft circumference might have a usefulness beyond just a limited archaeological scope.

When context and preservation allow for the recovery of complete os coxae or crania, discriminant analysis has provided sensible and accurate methods for sex determination. Several researchers have utilized these techniques. Schmidt (1985) designed a set of measurements for the pelvic bone which provided 95-98% accuracy in sexing from a sample of one hundred white skeletons.

The work of Giles (1964) and Giles and Elliot (1963) on

the mandible and cranium is well known in the anthropological literature for determining sex and race. Their 1963 article found an accuracy of 82-89% for sexing with a combination of American Negro and White material regardless of race. The authors do caution, however, that in some cases the male-female dividing point may need to be adjusted to fit the population at hand.

In a test of the Giles and Elliot methodology, Birkby (1966) examined 104 adult human crania from American Indian and Eskimo groups. He found that regardless of deformation, the male crania were at least as accurately sexed with metric methods as with visual methods. Unfortunately, the female crania were incorrectly sexed almost 50% of the time. Birkby supports the notion that sex (and race) determinations must be made on crania from the populations where the functions have been established.

Realizing the often fragmentary or deformed nature of human skeletons in the archaeological record, which would negate the use of any of the above discriminant function analyses, Holland (1986) tackled the idea of determining sex from the cranial base. A total of nine measurements were taken on a sample of one hundred skeletons from the Terry Collection. Holland found he could predict the sex of the sample with 71-90% accuracy, but he suggests that the technique needs to be tested on larger, more geographically

diverse populations before it can be utilized across differing skeletal samples.

Other metric studies in sex determination include, but are not limited to, the analysis of multiple variables on pelvic bones (Gomez and Fernandez, 1992), the diameter of the radial head (Berrizbeitia, 1989), measurements on Chinese femora (Wu, 1989), sexing from metacarpals and the first proximal phalanx (Scheuer and Elkington, 1993),arm bone measurements (Holman and Bennett, 1991) and Japanese tibiae (Iscan, et al., 1994).

The above findings require that many commonly used sex characteristics be more critically examined for their use as indicators in sex determination. This thesis investigates the interobserver repeatability for visual indicators and secondly, objective methods for combining these various indicators into a single assessment of sex. Finally, this thesis attempts to consider the characters across known sex samples in order to find equations that provide posterior probabilities that a skeleton with a given set of characters was male as versus female.

CHAPTER 2

MATERIALS AND METHODS

The Averbuch site (40DV60) is a late Mississippian village located near the Nashville Basin of Middle Tennessee. The site consisted of a stockaded village with three separate cemeteries which have provided the largest, professionally excavated skeletal collection from Middle Tennessee. The Averbuch site is from the Middle Cumberland Culture and the people occupied the site between the thirteenth and fifteenth centuries AD. They vanished prior to contact with the first settlers to that region in the eighteenth century. The three cemeteries had a total of 645 graves (887 individuals) that were recovered by the University of Tennessee and the Tennessee Division of Archaeology in 1977. Destruction by cultivation, construction and lack of excavating time caused an inability to recover approximately 409 individuals (Berryman, 1981).

A sample of 148 adult individuals from the Averbuch skeletal collection was examined for a set of seven commonly used morphological characteristics of the skull and pelvis. Unfortunately, working with prehistoric skeletal material is often a challenge due to the incompleteness of fragmented or worn remains. Of the 148 skeletons examined, not all traits

were present for observation on each individual, reducing sample sizes (see Table 1).

The seven variables used in sex determination were (1) chin form; (2) preauricular sulcus (Houghton, 1974); (3) sciatic notch; (4) superciliary arch; (5) mastoid process; (6) eye orbits; and (7) nuchal crests. Both left and right sides were observed where possible. These traits were scored on a five point ordinal scale, from (1) "M"-very male; (2) "M?"-somewhat male; (3) "?"-questionable (or "in-between"); (4) "F?"-somewhat female; and (5) "F"-very female. Table 2 shows the criteria used for distinguishing between males and females. Photographs of the individual traits for strong males and strong females are provided in Appendix A for all traits except the female nuchal crest.

It is well known that many ordinal scales are quite subjective. There is rarely perfect agreement between raters due to differing perceptions about the meaning of the category labels and due to intrarater variability (Agresti, 1988). In this case, two observers independently scored the material on the same ordinal categorical scale. A square contingency table with the same categories in each dimension was arranged to display the joint ratings of the two observers for each trait (see Appendix B). With this type of display, the main diagonal will represent observer agreement for the ratings which, in turn, measures the

TABLE 1

SAMPLE SIZES FOR EACH TRAIT SELECTED FROM THE AVERBUCH SKELETAL SERIES

Variable	Size
Chin form	142
Preauricular sulcus (left)	118
Preauricular sulcus (right)	112
Sciatic notch (left)	116
Sciatic notch (right)	122
Superciliary arch (left)	104
Superciliary arch (right)	104
Mastoid process (left)	133
Mastoid process (right)	125
Eye orbit (left)	114
Eye orbit (right)	115
Nuchal crest (left)	128
Nuchal crest (right)	125

TABLE 2

SUMMARY OF THE CRITERIA USED IN SCORING

MALES VERSUS FEMALES

Variable	Male	Female
Chin Form	square	round
Preauricular Sulcus	straight, narrow and shallow	wide, deep and scalloped
Sciatic Notch	narrow	wide
Superciliary Arch	large, projecting	small or absent
Mastoid Process	large, projecting	small and narrow
Eye Orbits	dull or blunt	sharp
Nuchal Crest	a large ridge	smooth, no ridge

reliability of the category. An analysis of the differences in the marginal distributions allows us to determine whether classifications by one rater are consistently higher than those by the other rater (Agresti, 1988; 1990).

Cohen (1960) devised the Kappa coefficient for summarizing the data from contingency tables. However, several researchers point out many unsatisfactory features when using kappa, including: (1) loss of information due to summarizing a table with one number; (2) sensitivity of the value to the marginal distributions; and (3) dangers in comparing kappa values between two tables (Agresti, 1988; Darroch and McCloud, 1986).

To solve these problems Darroch and McCloud (1986) and Tanner and Young (1985a) have suggested the use of models for describing the tables in nominal data situations. Further work on ordinal variables was carried out by Tanner and Young (1985b) and Agresti (1988). The work of Agresti will be especially relevant to this study.

Kennedy (1992) describes log linear models which are most appropriate when all of the variables are categorical. Historically, log linear models were designed for use by sociological and survey researchers in academia. The parameters utilized in these models, called lambda parameters, are estimated using the maximum likelihood method. This method is the most statistically efficient

method for the estimation of proportional responses.

Each specified model in this thesis uses chi-square procedures to assess the agreement between the frequencies generated by the model and those actually observed (goodness of fit). When selecting a model, it should have the most parsimonious fit, ie. it should contain only those terms that have been shown to be statistically deemed necessary to provide an acceptable fit (Kennedy, 1992).

Agresti (1988) discusses four models designed for use on ordinal catagorical scales. Each model was applied to the data in this thesis to see which had the most parsimonious fit. The saturated model was also included for comparison. The models were as follows:

(1) independence model

$$\ln m_{ij} = \mu + \lambda_i^A + \lambda_j^B \quad (\beta=0, \delta=0) \quad df = (r-1)^2$$

(2) diagonal model

$$\ln m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \delta\,(i,j) \quad (\beta=0) \quad df = (r-1)^2 - 1$$

(3) uniform-association model

$$\ln m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta\,\mu_i\,\mu_j \quad (\delta=o) \quad df = (r-1)^2 - 1$$

(4) agreement plus uniform-association model

$$\ln m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \beta\mu_i\,\mu_j + \delta\,(i,j) \quad df = (r-1)^2 - 2$$

(5) saturated model

$$\ln m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad df = 0$$

In each of these models, let A and B represent the two raters and m_{ij} represent the expected frequency of rating i by the first rater and rating j by the second rater. The μ denotes the grand mean of the population cell means, while λ_i represents the difference between the mean belonging to the i^{th} level and the grand mean (μ_i-μ) and $\lambda_j$denotes the difference in the j^{th} level from the grand mean (μ_j-μ).

The independence model (1) tests whether the two observers (A and B) are independent, as opposed to being associated, related or correlated. Independence between A and B implies that the two raters have not responded in approximately the same manner when scoring the sex traits. If A and B are found to be independent then there are material differences in the way in which the observers have scored the categories. However, if A and B are found not to be independent, then the pattern of scoring responses for one observer will not be different than for the other observer (Kennedy, 1992). The independence model looks only at the marginal totals and ignores the influence of the main diagonal (δ) and the relationship between the off diagonal elements (β).

The diagonal model (2) is an improvement over the previous model in that the diagonal model does take into consideration the effects of the main diagonal. However, it does not consider the relationship between the off-diagonal

elements. The parameter δ is included here to represent the main diagonal cells and shows the agreement beyond that expected by chance, ie. beyond what would be expected when the raters are independent.

Model (3), the uniform-association model, is the one proposed by Goodman (1979). This model adds the parameter β, which represents the relationship between the off-diagonal elements. Agresti (1988) feels that although this model works well for describing association between ordinal classifications, it does not take into consideration any behavior which is particular to the diagonal. Thus model (3) would have limited use in modeling agreement between raters.

Agresti (1988) suggests the use of model (4), the agreement plus uniform-association model, will provide a reasonable fit to the data because it allows for behavior particular to the main diagonal as well as off-diagonal elements. In this model, the overall agreement is partitioned into three parts: (1) chance agreement, (2) agreement due to a baseline association, and (3) an extra increment that shows agreement beyond that of chance or baseline association.

The saturated model (5) will always fit the data set. However, it is the least parsimonious model because it contains as many terms as there are cells in the contingency

table. Many of these terms have not been shown to be statistically necessary to fitting the data. This model also has zero degrees of freedom.

The results from the fit of these five models to the data in this section of the thesis will provide information of the joint agreement of the raters (interobserver agreement) as well as a test of independence. The data were also tested for marginal homogeneity following Agresti (1990):

$$G^2(S|CS) = G^2(S) - G^2(CS) \qquad df=1$$

where S represents symmetry, CS represents conditional symmetry and G^2 is a likelihood ratio chi-square with one degree of freedom. For an ordinal scale, single-degree-of-freedom statistics are more sensitive to certain departures from the null hypothesis. When results are significant, then CS will fit better than S. This equation analyzes the differences in marginal distributions which allows us to determine whether one rater consistently scored traits higher than the second rater.

The second section of this thesis involves the use of objective methods for combining the various indicators into a single assessment of sex. The Averbuch skeletal collection (40DV60) was used as a "known" sex sample where sex was determined using Phenice's (1969) characteristics. The use of these techniques has proven to correctly identify

the sex of an individual 96% of the time (Phenice, 1969). The individuals analyzed using Phenice's methods were only included when there was agreement between the traits scored.

To this archaeological sample was added an additional sample of known sex individuals taken from the Donated Collection. This collection is housed at The University of Tennessee at Knoxville and was assembled under the direction of Dr. William M. Bass. The collection is composed of modern peoples from the Tennessee area that have willed their remains to the Anthropology Department for use in research.

Multiple regression analysis and discriminant analysis are two types of multivariate statistical techniques which might come to mind when a dependent variable is to be predicted from a set of independent variables. However, when the dependent variable is binary, ie. an individual is to be classified into one of two populations, problems may arise with certain assumptions needed for these techniques (Norusis and SPSS, Inc.; 1993). The discriminant function analysis may lead to an underestimation of the coefficients when all of the X's are categorical (Afifi and Clark, 1984).

The logistic regression model requires far fewer assumptions and performs well in these situations. With logistic regression the probability of an event occurring is directly estimated and the logistic model allows the use of

categorical X variables. Due to the high correlations between sides for the bilateral traits, the data for this thesis was condensed to make seven independent predictor variables rather than the original thirteen. This was accomplished by eliminating one side from the bilateral variables. When both sides (left and right) were missing, the observation was dropped from the data set. Where only one side was present, that side was used in the analysis and where both sides were present, one side was randomly chosen for the data set. Also at this point, the observer scores which were (1) "M", (2) "M?", etc. were centered around zero so that: "M" was coded as -2, "M?" was coded as -1, "?" as 0, "F?" as 1 and "F" was coded as 2.

Model selection in logistic regression follows the ideas presented earlier in this thesis, ie. the model should contain only those terms statistically deemed necessary for an acceptable fit. Unlike linear regression, where model parameters are estimated using the method of least squares, the model parameters are estimated using the maximum-likelihood method. In this method, the coefficients that make the observed results the most "likely" are selected for the model (Norusis and SPSS, Inc; 1993).

All seven variables were originally forced to create a full model for the complete data set containing observations from the Averbuch and Donated Collections. This combined

sample had 106 males and 47 females of known sex. It was subsequently considered beneficial to identify that subset of independent variables which would be the best predictors of sex. Thus, the stepwise selection procedure was applied to the seven variables and a combination of indicators which provided the best estimate of sex was chosen. This procedure was done on the full data set (106 males and 47 females) and repeated on the observations from the Averbuch collection only (31 males and 30 females) to the exclusion of the Donated material. There is no assurance that the same model will be selected if a different sample from the population is used (Norusis and SPSS, Inc.; 1993). It was expected, in fact, that the selection procedure would choose different variables due to the differing sex ratios and sexual dimorphism found between the two collections.

The analysis was repeated using the cranial variables only and excluding the two pelvic traits (preauricular sulcus and sciatic notch). Again, the sex was regressed on the traits for the entire collection (109 males and 47 females) and the Averbuch collection only (34 males and 30 females). Appendix C shows the SAS® program (SAS Institute, Inc.,1990) used to complete all of the above regression equations while Appendix D lists the raw data used in the analysis.

Finally, following the identification of the logistic

regression equation best suited to the data sets, posterior probabilities were sought which would identify whether a skeleton with a given set of characteristics was male as versus female. For more than one independent variable, the model is:

$$Prob\,(male) = \frac{e^{z}}{1+e^{z}}$$

where Z is the linear combination:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

This model provides the posterior probabilities needed to estimate the probability of an event occurring when a given set of characters are present, ie. we can predict the sex of an individual when certain clusters of traits are known. In general, if the estimated probability for a male is greater than 0.5 we would predict that the specimen will be a male; whereas, if the probability is less than 0.5 we predict that the specimen will be a female.

CHAPTER 3

RESULTS

In this section each model described above has been fitted to the data and tested. This allows us to choose the most parsimonious model and evaluate the independence of the raters along with the interobserver agreement.

The first test was of the agreement plus uniform-association model in comparison to the saturated model (see Table 3). In only one case, the right eye orbit (p-value= .0197, chi-square= 26.9221), was a significant result found and the model was rejected in favor of the saturated model. For the remaining twelve traits, the agreement plus uniform-association provided an acceptable fit. With 14 fewer parameters than the saturated model, the agreement plus uniform-association model is much more parsimonious and is therefore the model of choice for those twelve traits.

Since the agreement plus uniform-association model takes into account any behavior particular to the diagonal, we can use this model to test the interobserver agreement for each variable. Where the model is rejected, we do not see good agreement between the raters. So we can conclude that the two observers did not agree on sex assessment of the right eye orbit, but were not shown to disagree on any

TABLE 3

SUMMARY OF THE RESULTS FROM THE TEST OF THE AGREEMENT PLUS UNIFORM-ASSOCIATION MODEL AGAINST THE SATURATED MODEL (WITH 14 DF)

Variable	chi-square	p-value
Chin Form	14.7956	.3923
Preauricular Sulcus (l)	14.8918	.3856
Preauricular Sulcus (r)	17.9529	.2089
Sciatic Notch (l)	14.5020	.4130
Sciatic Notch (r)	16.0322	.3114
Superciliary Arch (l)	14.3078	.4270
Superciliary Arch (r)	19.6538	.1414
Mastoid Process (l)	23.4056	.0540
Mastoid Process (r)	21.9232	.0802
Eye Orbit (l)	12.0220	.6045
Eye Orbit (r)	26.9221	.0197 *
Nuchal Crest (l)	10.7699	.7040
Nuchal Crest (r)	13.2782	.5047

* represents a significant value

other variables.

Table 4 represents the maximum likelihood parameter estimates for the agreement plus uniform-association model with asymptotic standard errors. Using these estimates, the beyond-chance agreement can be summarized. For example, using the chin form variable, for i=1,2,3,4, the odds that the score of observer A is i+1 rather than i is estimated to be $\exp(\beta+2\delta)=2.84$ times higher when the score of observer B is i+1 than when it is i. This reasoning can be followed for the remaining twelve variables to summarize the beyond-chance agreement.

For the comparison of the uniform-association model to the saturated model (see Table 5), only two traits (the right superciliary arch and right eye orbit) show a significantly better fit of the saturated model as versus the more parsimonious uniform-association model. Conversely, for the remaining eleven traits, the uniform-association model provided an acceptable fit. With 15 fewer parameters than the saturated model, the uniform-association model is much more parsimonious and is therefore the model of choice for those eleven traits.

Table 6 shows the likelihood-ratio chi-squared values and associated p-values for the test of the diagonal model compared to the saturated model. These comparisons have 15 degrees of freedom, as the diagonal model contains ten

TABLE 4

MAXIMUM LIKELIHOOD PARAMETER ESTIMATES

FOR THE AGREEMENT PLUS UNIFORM-ASSOCIATION MODEL

Variable	β	a.s.e.
Chin Form	.4410	.0856
Preauricular Sulcus (l)	.9954	.1941
Preauricular Sulcus (r)	.7689	.1847
Sciatic Notch (l)	.7115	.1639
Sciatic Notch (r)	.5779	.1381
Superciliary Arch (l)	.7520	.2019
Superciliary Arch (r)	.6839	.1812
Mastoid Process (l)	.3327	.0729
Mastoid Process (r)	.3209	.0781
Eye Orbit (l)	.6169	.1272
Eye Orbit (r)	.3696	.0928
Nuchal Crest (l)	.5388	.1221
Nuchal Crest (r)	.6464	.1336

TABLE 4 continued

Variable	δ	a.s.e.	Odds
Chin Form	.3021	.2259	2.8440
Preauricular Sulcus (l)	-.4137	.3367	1.1828
Preauricular Sulcus (r)	.7664	.3052	9.9911
Sciatic Notch (l)	.4364	.2928	4.8754
Sciatic Notch (r)	.8095	.2898	8.9880
Superciliary Arch (l)	.8880	.3023	12.5301
Superciliary Arch (r)	.8955	.2886	11.8816
Mastoid Process (l)	.1974	.2370	2.0698
Mastoid Process (r)	.3947	.2560	3.0355
Eye Orbit (l)	.5693	.2878	5.7863
Eye Orbit (r)	1.0448	.2800	11.6947
Nuchal Crest (l)	.6213	.2312	5.9381
Nuchal Crest (r)	.2849	.2319	3.3746

TABLE 5

SUMMARY OF THE RESULTS FROM THE TEST OF THE UNIFORM-ASSOCIATION MODEL AGAINST THE SATURATED MODEL (WITH 15 DF)

Variable	chi-square	p-value
Chin Form	16.5727	.3450
Preauricular Sulcus (l)	16.5437	.3467
Preauricular Sulcus (r)	24.1306	.0629
Sciatic Notch (l)	16.7503	.3340
Sciatic Notch (r)	24.1821	.0621
Superciliary Arch (l)	23.3434	.0771
Superciliary Arch (r)	29.6687	.0132 *
Mastoid Process (l)	24.0974	.0635
Mastoid Process (r)	24.2965	.0602
Eye Orbit (l)	15.9692	.3841
Eye Orbit (r)	41.3506	.0003 *
Nuchal Crest (l)	18.1430	.2552
Nuchal Crest (r)	14.8053	.4655

* represents a significant value

TABLE 6

SUMMARY OF THE RESULTS OF THE TEST OF THE DIAGONAL MODEL AGAINST THE SATURATED MODEL (WITH 15 DF)

Variable	chi-square	p-value
Chin form	56.4326	<.0001
Preauricular Sulcus (l)	90.1810	<.0001
Preauricular Sulcus (r)	57.9554	<.0001
Sciatic Notch (l)	61.9101	<.0001
Sciatic Notch (r)	51.6281	<.0001
Superciliary Arch (l)	49.7694	<.0001
Superciliary Arch (r)	52.6334	<.0001
Mastoid Process (l)	51.0398	<.0001
Mastoid Process (r)	44.2602	<.0001
Eye Orbit (l)	61.1243	<.0001
Eye Orbit (r)	50.0737	<.0001
Nuchal crest (l)	45.1949	<.0001
Nuchal crest (r)	59.4237	<.0001

parameters (it has one additional parameter over the independence model that is the term for the diagonal association). Again, we see very high chi-squared values paired with low p-values. The diagonal model is rejected in all cases. Thus, the saturated model provides a significantly better fit to the data for this thesis than does the diagonal model.

Table 7 shows the same statistics for the test of the independence model against the saturated model. These comparisons have 16 degrees of freedom, as the saturated model contains 25 parameters, while the independence model contains nine parameters. All variables had a high chi-squared value and a very low p-value, with significant results. The null hypothesis for independence was rejected in all cases. The raters did not show independence in their observations.

Choosing the most parsimonious model to explain the data set at this point comes down to a decision between the uniform-association model and the agreement plus uniform-association model. Both the independence model and the diagonal model were soundly rejected in favor of the saturated model. However, the uniform-association model was favored for eleven out of the thirteen variables while the agreement plus uniform-association model was favored in twelve out of the thirteen variables over the saturated

TABLE 7

SUMMARY OF THE RESULTS FROM THE TEST OF THE INDEPENDENCE MODEL AGAINST THE SATURATED MODEL (WITH 16 DF)

Variable	chi-square	p-value
Chin form	99.9044	<.0001
Preauricular sulcus (l)	140.7965	<.0001
Preauricular sulcus (r)	157.0388	<.0001
Sciatic Notch (l)	149.5321	<.0001
Sciatic Notch (r)	157.1193	<.0001
Superciliary Arch (l)	152.9941	<.0001
Superciliary Arch (r)	151.3097	<.0001
Mastoid Process (l)	74.5098	<.0001
Mastoid Process (r)	75.8398	<.0001
Eye Orbit (l)	127.5065	<.0001
Eye Orbit (r)	132.6871	<.0001
Nuchal Crest (l)	108.4978	<.0001
Nuchal Crest (r)	103.6769	<.0001

model. Table 8 provides the likelihood-ratio chi-squared values and associated p-values when the two models are compared to each other. In the case of the right eye orbit, we have already seen that the saturated model is favored, while for the right superciliary arch, the uniform-association model was rejected and the agreement plus uniform-association model was accepted over the saturated model. Thus, these two variables have already been decided and are not included in Table 8. Significance was found for five of the remaining eleven variables (right preauricular sulcus, right sciatic notch, left superciliary arch, left eye orbit and left nuchal crest). In these cases, the agreement plus uniform-association model had a better fit than the uniform-association model. In the remaining six variables, where the uniform-association model was favored, nothing was gained by using the agreement plus uniform-association model.

The results of the test of the marginal homogeneity are shown in Table 9. Significance was found for the chin form, preauricular sulci, and the eye orbits. Thus, these variables displayed marginal heterogeneity. Interpretations of these results are included in the Discussion section of this thesis.

The second objective of this thesis was to determine which subset of variables could predict sex the best on the

TABLE 8

SUMMARY OF THE RESULTS FROM THE TEST OF THE UNIFORM-ASSOCIATION MODEL VS. THE AGREEMENT PLUS UNIFORM-ASSOCIATION MODEL (WITH 1 DF)

Variable	chi-square	p-value
Chin Form	1.7771	.1825
Preauricular Sulcus (l)	1.6518	.1987
Preauricular Sulcus (r)	6.1776	.0129 *
Sciatic Notch (l)	2.2483	.1338
Sciatic Notch (r)	8.1499	.0043 *
Superciliary Arch (l)	9.0356	.0026 *
Mastoid Process (l)	0.6918	.4056
Mastoid Process (r)	2.3733	.1234
Eye Orbit (l)	3.9472	.0469 *
Nuchal crest (l)	7.3731	.0066 *
Nuchal Crest (r)	1.5271	.2165

* represents a significant value

TABLE 9

TEST OF MARGINAL HOMOGENEITY

Variable	chi-square	p-value
Chin Form	10.1237	.0015 *
Preauricular Sulcus (l)	6.7043	.0096 *
Preauricular Sulcus (r)	11.9153	.0006 *
Sciatic Notch (l)	0.4104	.5128
Sciatic Notch (r)	<.0001	.0000
Superciliary Arch (l)	1.1583	.2818
Superciliary Arch (r)	1.9995	.1574
Mastoid Process (l)	1.0900	.2965
Mastoid Process (r)	1.0758	.2996
Eye Orbit (l)	28.7779	<.0001 *
Eye Orbit (r)	22.5026	<.0001 *
Nuchal Crest (l)	0.1536	.6951
Nuchal Crest (r)	0.7661	.3814

* represents a significant value

combined sample and the Donated sample by itself. The results of the logistic regression follow.

The parameter estimates, chi-squares and p-values are shown in Table 10 for the forced model. This model contains all seven variables and is based on the combined data set of the Averbuch and Donated skeletal collections. Only the coefficients for the preauricular sulcus and sciatic notch are significant in this full model. This suggests that the full model contains parameters which have not been deemed necessary to predict sex.

Table 11 provides the results of the stepwise selection procedure of all seven variables on the combined sample. The selection procedure has indicated that the combination of variables including preauricular sulcus, sciatic notch and superciliary arch is the best subset of variables for predicting sex. All of these variables remained significant when placed in the model. Table 12 shows the p-values of the remaining four variables not included in the model.

When the combined collection sample was analyzed for cranial variables, only the superciliary arch and mastoid process were found to be significant for predicting sex in this sample (see Table 13). The chin form had been added to the model but was removed by the selection procedure when its model p-value rose to .0517, a non-significant value. Table 14 shows the p-values of the remaining three cranial

TABLE 10

ESTIMATES FOR THE FORCED MODEL

ALL VARIABLES, COMBINED DATA SET

(WITH 1 DEGREE OF FREEDOM)

Variable	Parameter Estimate	Chi-square	P-Value
Intercept	-0.2903	0.1940	.6596
Preauricular Sulcus	-1.3558	9.4777	.0021 *
Sciatic Notch	-0.7920	6.4554	.0111 *
Superciliary Arch	-0.4792	1.9559	.1620
Mastoid Process	-0.5321	2.1046	.1469
Eye Orbit	-0.0832	0.0748	.7845
Nuchal Crest	0.2463	0.3247	.5688
Chin Form	-0.3507	0.9902	.3197

* represents a significant value

TABLE 11

STEPWISE SELECTION ON ALL VARIABLES

FOR THE COMBINED DATA SET

Variable	Parameter Estimate	P-Value
Intercept	0.3113	.4549
Preauricular Sulcus	-1.2422	.0010
Sciatic Notch	-0.9589	.0009
Superciliary Arch	-0.7608	.0047

TABLE 12

VARIABLES NOT FOUND IN TABLE 11

Variable	P-Value
Mastoid Process	.1314
Eye Orbit	.6360
Nuchal Crest	.9081
Chin Form	.3870

TABLE 13

STEPWISE SELECTION ON THE CRANIAL VARIABLES

FOR THE COMBINED DATA SET

Variable	Parameter Estimate	P-Value
Intercept	-0.1457	.7123
Superciliary Arch	-1.2789	.0001
Mastoid Process	-0.7740	.0010

TABLE 14

VARIABLES NOT FOUND IN TABLE 13

Variable	P-Value
Eye Orbit	.4901
Nuchal Crest	.9957
Chin Form	.0470

traits which were not included in the model.

The analysis so far has considered the combined data sets. Table 15 provides the results of the regression when all variables were used on the Averbuch collection only. Interestingly, the subset of variables found to be significant was slightly different than the subset found for the combined sample. When only the archaeological collection was used, the preauricular sulcus and chin form traits were found to be significant. The sciatic notch was placed in the model but was later removed by the selection procedure when its p-value in the model rose to a non-significant value of .0511. Table 16 shows the p-values of the five traits not included in the model.

The last selection procedure to be run was on the cranial variables of the Averbuch skeletal sample (see Table 17). The results here indicate that significance was found for the superciliary arch and the chin form but not for the remaining variables (see Table 18).

Finally, these logistic regression equations were used to obtain posterior probabilities. These probabilities predict the outcome of an event given certain sets of characters. For the combined skeletal sample with all variables used, the posterior probabilities are listed below. For example, the first statement should be read, "the probability that an individual is a male given that all

TABLE 15

STEPWISE SELECTION ON ALL VARIABLES

FOR THE AVERBUCH SAMPLE ONLY

Variable	Estimate	P-Value
Intercept	-1.7591	.0324
Preauricular Sulcus	-1.6359	.0014
Chin Form	-1.1986	.0058

TABLE 16

VARIABLES NOT FOUND IN TABLE 15

Variable	P-Value
Sciatic Notch	.0309
Superciliary Arch	.0993
Mastoid Process	.0506
Eye Orbit	.2998
Nuchal Crest	.3053

TABLE 17

STEPWISE SELECTION ON THE CRANIAL VARIABLES

FOR THE AVERBUCH SAMPLE

Variable	Parameter Estimate	P-Value
Intercept	-0.6550	.1939
Superciliary Arch	-1.0301	.0009
Chin Form	-0.9226	.0058

TABLE 18

VARIABLES NOT FOUND IN TABLE 17

Variable	P-Value
Mastoid Process	.2421
Eye Orbit	.3091
Nuchal Crest	.9897

traits were scored as -2 is 99.8%," and so on.

(1) p(male¦all -2)=.9980

(2) p(male¦all -1)=.9635

(3) p(male¦all 0)=.5772

(4) p(male¦all 1)=.0659

(5) p(male¦all 2)=.0036

When the pelvic traits are deleted and just the cranial variables are used for both skeletal collections, the probabilities become:

(1) p(male¦all -2)=.9813

(2) p(male¦all -1)=.8707

(3) p(male¦all 0)=.4636

(4) p(male¦all 1)=.0999

(5) p(male¦all 2)=.0140

The posterior probabilities were also calculated for all of the variables in the Averbuch collection. These probabilities, in turn, can predict the sex of an individual from the Donated collection given certain values of traits are present. The probabilities are:

(1) p(male¦all -2)=.9803

(2) p(male¦all -1)=.7456

(3) p(male¦all 0)=.1469

(4) p(male¦all 1)=.0100

(5) p(male¦all 2)=.0006

When these seven variables from the Averbuch collection

are used to predict sex in the Donated collection the results will vary depending on what level of significance is chosen for the cut off between male and female. Tables 19, 20 and 21 show the differences between observed and predicted values for three different significance levels. Table 19 shows probabilities at a general dividing point of 0.5; any values above this level would be classified as male while values below 0.5 would be classified as female. It is obvious from the differences in actual and predicted totals that males are underclassified and females overclassified at this level. The percentage of males and females that were correctly classified at this level was 75%, while the number of males correctly classified was 69.3% and the number of females correctly classified was 100%. At this level, the real sex ratio was 4.41 males:1 female and the estimated sex ratio was 1.3 males:1 female. However, when we require the probability to be greater than 0.8 to classify an individual as a male and less than 0.2 for a female, we see some improvements. The overall correct classification for both sexes from Table 20 is 89.5%. The percentage of males who were correctly classified was 86% and females were correctly classified 100% of the time. The sex ratios also showed a slight improvement. The real sex ratio was 2.94 males:1 female and 1.79 males:1 female was the estimated sex ratio.

Further still, as the requirements become even more

TABLE 19

OBSERVED VS. PREDICTED VALUES OF ALL VARIABLES

FROM THE AVERBUCH COLLECTION WHEN P=0.5

	Predicted males	Predicted females	Total
Actual males	52	23	75
Actual females	0	17	17
Total	52	40	92

TABLE 20

OBSERVED VS. PREDICTED VALUES OF ALL VARIABLES

FROM THE AVERBUCH COLLECTION WHEN P>0.8 OR P<0.2

	Predicted males	Predicted females	Total
Actual males	43	7	50
Actual females	0	17	17
Total	43	24	67

TABLE 21

OBSERVED VS. PREDICTED VALUES OF ALL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.9 OR P<0.1

	Predicted males	Predicted females	Total
Actual males	28	5	33
Actual females	0	15	15
Total	28	20	48

stringent, where p must be greater than 0.9 to classify an individual as a male and less than 0.1 for females, the results again look better as the predicted values approach the actual values. Both males and females are correctly classified 89.6% of the time at this level. Males are correctly classified 84.8% of the time and females are again classified correctly 100% of the time. The real sex ratio improved to 2.2 males for every 1 female and the estimated sex ratio was 1.4 males for every 1 female.

The same analysis on the Averbuch collection using just the cranial traits yields the following posterior probabilities for predicting sex in the Donated sample:

(1) p(male¦all -2)=.9627

(2) p(male¦all -1)=.7854

(3) p(male¦all 0)=.3419

(4) p(male¦all 1)=.0686

(5) p(male¦all 2)=.0103

When these probabilities are used to predict sex there is a difference in observed and predicted values for different cut off levels of p. When a probability level of 0.5 is used to distinguish the two groups, males are underclassified and females are overclassified (see Table 22). We see a poor overall correct classification of 72.8%. The males are correctly classified 68% of the time and the females 94.1% of the time. The real sex ratio is 4.41

TABLE 22

OBSERVED VS. PREDICTED VALUES OF CRANIAL VARIABLES

FROM THE AVERBUCH COLLECTION WHEN P=0.5

	Predicted males	Predicted females	Total
Actual males	51	24	75
Actual females	1	16	17
Total	52	40	92

males:1 female and the estimated sex ratio is 1.3 males for every 1 female. Unfortunately, requiring males to exhibit a probability of greater than 0.8 and females less than 0.2 does not do much to bring the predicted values closer to the actual values (see Table 23). The overall correct classification remains low at 75%, while males are correctly classified 67.3% of the time and females are correctly classified 100% of the time. The real sex ratio remains high at 3.27 males for every female and the estimated sex ratio is 1.06 males for every female. Only in Table 24, when males require a p greater than 0.9 and females require a p of less than 0.1 do we see the predicted values approach the actual observed values. The overall correct classification of sex rises to 87.2%, while males are correctly classified 82.7% of the time and females remain at a 100% correct classification level. The real sex ratio inproves a little at 2.9 males for every female and the estimated sex ratio is 1.6 males for every female.

TABLE 23

OBSERVED VS. PREDICTED VALUES OF CRANIAL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.8 OR P<0.2

	Predicted males	Predicted females	Total
Actual males	33	16	49
Actual females	0	15	15
Total	33	31	64

TABLE 24

OBSERVED VS. PREDICTED VALUES OF CRANIAL VARIABLES FROM THE AVERBUCH COLLECTION WHEN P>0.9 OR P<0.1

	Predicted males	Predicted females	Total
Actual males	24	5	29
Actual females	0	10	10
Total	24	15	39

CHAPTER 4

DISCUSSION

Most papers dealing with skeletal demography accept the sexing of the specimens they use and attempt to explain the observed sex ratios in demographic or cultural terms rather than first questioning the accuracy of the sexing procedures themselves (Weiss, 1972). This thesis has attempted to question the accuracy of visual methods of sex determination and define more objective methods for visual sexing.

The interobserver repeatability section has shown several things. First, the observers were not independent; this was the desired outcome. If the observers had been independent then they would have had different criteria in mind when they examined and scored the variables. However, their dependence means that they were agreeing on what a particular trait should look like, ie. square or round, sharp or dull, wide or narrow, etc.

Secondly, the two observers did show agreement in scoring the variables. This was shown by the acceptance of the agreement plus uniform-association model for twelve of the thirteen variables. This model takes into consideration the behavior particular to the main diagonal, which is the number of times the observers agree exactly. The model also

considers any off-diagonal associations, ie. if one observer scored a 1 ("M") and the other observer scored a 2 ("M?"), there is much more association than if the first observer had scored a 1 ("M") and the second observer had scored a 4 ("F?") or 5 ("F") for the same trait. The only problem was with the right eye orbit. For this variable, the agreement plus uniform-association model was rejected in favor of the saturated model. It is likely that a model which is more parsimonious than the saturated model, not included among those tested in this thesis, would have the most reasonable fit for this variable.

An analysis of the marginal distributions may help to explain this problem. In Table 9, the results of the test for marginal homogeneity are depicted. The right eye orbit and four other variables (left eye orbit, preauricular sulci and chin form) were found to be significant. Significance for the right eye orbit suggests that the conditional symmetry model (not discussed in this thesis) would better represent this variable. Significance also indicates that the five variables display marginal heterogeneity. Marginal heterogeneity occurs when the column and row totals are significantly different for the two observers. Table 25 represents a condensed version of the number of individuals scored in the three overall categories available for each trait. The categories were condensed by combining all "M"

TABLE 25

CONDENSED CHART OF MARGINAL HOMOGENEITY

	Observer A			Observer B		
Variable	M	?	F	M	?	F
Chin Form	73	17	52	55	35	52
Preauricular Sulcus (l)	56	17	45	46	21	51
Preauricular Sulcus (r)	49	21	42	43	21	48
Sciatic Notch (l)	51	11	54	52	4	60
Sciatic Notch (r)	57	5	60	48	10	64
Superciliary Arch (l)	58	7	39	58	7	39
Superciliary Arch (r)	59	8	37	62	7	35
Mastoid Process (l)	91	8	34	80	22	31
Mastoid Process (r)	90	5	30	82	21	22
Eye Orbit (l)	49	11	54	61	6	47
Eye Orbit (r)	46	8	61	58	6	51
Nuchal Crest (l)	71	21	36	67	23	38
Nuchal Crest (r)	71	19	35	69	19	37

and "M?" scores into "M" and all "F" and "F?" scores into "F" while leaving the "?" category as it was. Hopefully, this condensed chart will allow for more ease in interpretation. For the five variables just listed, it can be seen that there exists a larger difference in the marginal totals between the observers. In the case of the right eye orbit, observer A scored 46 males, 8 undecided and 61 females, while observer B scored more males (58), about the same in the undecided category (6) and fewer females (51). These differences were found to be significant. While an analysis of marginal distributions can be very beneficial, it disregards any association that may occur between the cells of the contingency table. Thus, although column and row totals could be equivalent, this does not require association. Therefore, the importance of a complete model containing the essential parameters needed to describe the data set cannot be understated.

Finally, Agresti (1988) discussed four models presented and tested in this thesis. This final part to the interobserver repeatability study involved finding the most parsimonious model for the data. The independence model and the diagonal model were both rejected. Neither of these models contained the appropriate number of parameters to provide an acceptable fit. The independence model only tested the marginal distributions, which, as we have just

seen, are insufficient for describing the data because any association and agreement that may occur between the variables is ignored. The diagonal model, which tests the exact interobserver agreement, unfortunately disregards the influence of any association between the off-diagonal elements. The last two models, the uniform-association model and the agreement plus uniform-association model fit the data best. Each of these models was selected to have the best fit for about half of the variables (see Table 8). Where the uniform-association model was selected for over the agreement plus uniform-association model, it suggests that the addition of the extra parameter for behavior particular to the main diagonal in the agreement plus uniform-association model was unnecessary and did not add any additional, relevant information to the model.

Logistic regression has been described as the best method to use whenever an individual is to be classified into one of two populations, and it works well with categorical data (Afifi and Clark, 1984). Utilization of this technique for regressing the sex of the individual onto the traits provided regression equations which could best estimate the sex of an individual. The forced (full) model has indicated that a subset of variables would better predict sex, as not all of the variables stayed significant when placed in the model. Thus, four reduced models were

run on: all variables for the combined sample and Averbuch sample and for the cranial variables on the combined sample and Averbuch sample.

When all seven variables were run on the combined sample, the reduced model contained estimates for the preauricular sulcus, sciatic notch and superciliary arch. However, only two of the seven variables were found to be significant on the Averbuch sample alone: the preauricular sulcus and chin form. The support found here for the preauricular sulcus as a sex determinant in skeletal samples sustains Houghton's (1974) conclusions that this landmark has a different appearance in males as opposed to females. It is interesting that the sciatic notch was not added to the model for the Averbuch sample because pelvic traits are usually considered superior to cranial traits in sex determination of a skeleton (Meindl et al., 1985). But, the sciatic notch was just barely found to be non-significant at .0511 when placed in the model.

In the analysis of the cranial variables on the combined sample, the superciliary arch and mastoid process were found to be significant. Interestingly enough, though, the regression carried out for the Averbuch collection alone selected the superciliary arch and the chin form to fit the reduced regression model. The exclusion of the mastoid process from the archaeological sample is thought to be due

to the sexual dimorphism found in the Averbuch collection. Throughout the sample, it was observed that the "female" crania had very large mastoid processes, which is considered to be a male characteristic. When the samples were combined, however, the overproportion of males in the Donated collection may have countered this effect, allowing the mastoid process to be estimated for the reduced model there. It may be important to note here, that when dealing with subjective traits, some may be more subjective than others. This may especially be the case when size is a factor rather than shape. It may be easier to distinguish between a square shaped chin versus a rounded chin form than it is to distinguish between a small versus a large mastoid process. These differences could be magnified when one is working with a fragmented archaeological population where the sexual dimorphism is not known.

Posterior probabilities express the probability of belonging to a particular population after performing the analysis. They are a valuable method of interpreting classification results (Afifi and Clark, 1984). It is common for an investigator to classify only those individuals whose probabilities clearly favor one group over another. Judgment can be withheld for observations where the posterior probabilities are close to 0.5.

When the combined sample was used and all traits

analyzed, there is a 99.8% chance that the individual is a male when all of the traits are scored as -2 ("M"). This value decreases somewhat to 96% when all traits are scored as -1 ("M?"). The use of the cranial variables alone in the combined sample decreases these probabilities. There is a 98% chance that the individual is a male when all cranial traits are scored as a -2 ("M") and an 87% chance that the individual is a male when all cranial traits are scored as a -1 ("M?"). These probabilities reflect the notion that pelvic traits are better visual indicators of sex than are cranial traits. The high probabilities of getting a male when the characters are scored as a 0 ("?") (.5772 for all traits and .4636 for cranial traits) may reflect the sex bias of the Donated sample. Since this collection has over a three to one bias in favor of males, the estimates should reflect the reference population.

The posterior probabilities for the Averbuch collection also show a decrease from those of the combined sample. When all traits are scored as a -2 ("M"), the chance that the individual is a male is 98%, but this value decreases to 75% when the characters are scored as a -1 ("M?"). However, a noticeable change occurs when all traits are scored as a 0 ("?"). In this situation, the posterior probability is 15%. This is much lower than the values reported for the combined sample. The sex ratio for the Averbuch collection is nearly

even.

The probabilities changed somewhat when analyzing the cranial traits from the Averbuch collection. Here, when the features were scored as a 0 ("?"), the probability rose to a 34% chance of getting a male. This may be related to the "masculine" appearance of many of the "female" crania.

Using the probabilities from the Averbuch sample to predict the sex of individuals in the Donated collection did not provide particularly promising results. When the general dividing point of 0.5 for males and females was used, the males were underestimated and the females overestimated for both the cranial variables alone and the cranial and pelvic traits together (see Tables 19 and 22). Moving to a more stringent dividing point, where males must exhibit a probability greater than 0.8 and females a probability of less than 0.2 helped the predictions somewhat when all variables were used (see Table 20), but had very little effect on the estimates from cranial variables (see Table 23). Boosting the criteria for acceptance to a probability greater than 0.9 for males and less than 0.1 for females resulted in much improved estimates, especially for the cranial variables (see Tables 21 and 24). It would be wise in these situations to classify individuals only when their probabilities clearly favor one group or another.

The poor estimations of sex in the Donated sample from

the Averbuch material is thought to be due to the sex bias in the Donated material. While the prior probability of the Donated sample for males is close to 0.8, the prior probability of the Averbuch sample for males is close to 0.5. The prior probability affects the choice of the dividing point that is used to produce an equal percentage of errors of both types, ie. misclassifying an individual from category 1 into category 2 and vice versa (Afifi and Clark, 1984). Thus, more stringent dividing points would be needed to decrease the probability of error when dealing with a severely biased sample.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The three objectives of this thesis were: (1) to examine interobserver repeatability in the visual scoring of thirteen traits commonly used in sex determination, (2) after reducing the six bilateral traits to one side to make a total of seven characteristics, determine the regression equations which are best suited to predicting sex from all variables and from the cranial variables on both the combined and the archaeological samples and, (3) use these equations to find posterior probabilities that a given individual was male as versus female.

The first section dealing with repeatability was examined by fitting four models to the data set to find the most parsimonious model that would explain the data. It was determined that the observers were not independent in their observations, meaning that when they scored the traits, they were using the same visual criteria and had similar ideas of what a particular trait should look like in order to be scored a particular way. The results also indicated that the observers were scoring the traits in the same way, ie. they were repeatable, for all traits except the right eye orbit. The problems here may have been due to differences

in the marginal totals between the observers. Further testing would show how well the observers agreed, rather than just knowing that they did agree.

Of the four models utilized in the first section, the uniform-association and agreement plus uniform-association models were almost evenly split in their fit to the data set. In some cases the extra parameter for the main diagonal included in the agreement plus uniform-association model was not needed to fit the data and the uniform-association model was the most parsimonious model. In the case of the right eye orbit, the saturated model provided the best fit of the data.

The second section of the thesis identified a full model when all variables were forced into the equation. Only two of the seven variables were significant in the full model suggesting the need for a reduced model. This reduced model would be based on a subset of variables which could best predict sex in a given collection. The differences in the traits selected for the model representing the combined sample and the model representing the Averbuch sample may be due to differing sexual dimorphism between the two populations, one modern and one archaeological. Another very likely reason would be the biased sex ratio of the Donated collection where observed males outnumbered females more than three to one. This situation would cause the

estimates to mimic the reference population.

This second section would be benefitted by the addition of multivariate ordinal regression of the variables on sex. In this case, the known sex would be the independent variable and from that we could predict the probability of getting a certain set of traits. Konigsberg and Frankenberg (in press) described this technique where an indicator is regressed on age. This does not force us into making the assumption that a given skeleton comes from a particular age-at-death distribution. The methodology can be applied here to predict the probability of a certain set of traits given the individual is a male or a certain set of traits given the individual is a female. There are no assumptions here about the sex ratio of the reference sample, which, as in the case of the Donated sample, may be skewed.

The final section of the thesis determined a number of posterior probabilities which can be used to estimate the sex of an individual when it is unknown. Only the simplest probabilities were calculated here, ie. when all traits were assigned the same score. The study could be expanded to include probabilities where traits had different scores. Unfortunately, the probabilities derived from the Averbuch collection did not estimate the sex of individuals in the Donated collection very well unless more stringent dividing points separating males from females were utilized.

However, this was to be expected due to the differences in sexual dimorphism between the two populations and the biased sex ratio of the Donated collection.

It is hoped that this thesis will provide a more methodologically rigorous and objective approach to the subjective method of visual sexing which is so common in Anthropology today.

LIST OF REFERENCES

LIST OF REFERENCES

Afifi AA and Clark V (1984) Computer-Aided Multivariate Analysis. California: Lifetime Learning Publications.

Agresti A (1988) A model for agreement between ratings on an ordinal scale. Biometrics 44:539-548.

Agresti A (1990) Categorical Data Analysis. New York:John Wiley and Sons.

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.

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37-46.

Darroch JN and McCloud PI (1986) Category distinguishability and observer agreement. Austral. J. Statist. 28:371-388.

DiBennardo R and Taylor JV (1979) Sex assessment of the femur: a test of a new method. Am. J. Phys. Anthropol. 50:635-638.

Giles E (1964) Sex determination by discriminant function analysis of the mandible. Am. J. Phys. Anthropol. 22:129-136.

Giles E and Elliot O (1963) Sex determination by discriminant function analysis of crania. Am. J. Phys. Anthropol. 21:53-68.

Goodman LA (1979) Simple models for the analysis of association in cross-classifications having ordered catagories. J. Amer. Stat. Assoc. 74:537-552.

Gomez PL and Fernandez CFJ (1992) Biometry of the anterior border of the human hip bone: normal values and their use in sex determination. J. of Anat. 181:417-422.

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.

Holt CA (1978) A re-examination of partuition scars on the human female pelvis. Am. J. Phys. Anthropol. 49:91-94.

Houghton P (1974) The relationship of the pre-auricular groove of the ilium to pregnancy. Am. J. Phys. Anthropol. 41:381-390.

Iscan MY, Yoshino M, and Kato S (1994) Sex determination from the tibia: standards for contemporary Japan. J. For. Sci. 39:785-792.

Kennedy JJ (1992) Analyzing Qualitative Data: Log Linear Analysis for Behavioral Research. New York: Praeger.

Konigsberg LW and Frankenberg SR (in press) Paleodemography: "not quite dead" yet. Evolutionary Anthropol.

Krogman WM (1962) The Human Skeleton in Forensic Medicine. Springfield, Illinois: Charles C. Thomas.

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.

Norusis MJ and SPSS, Inc. (1993) SPSS for Windows: Advanced Statistics Release 6.0. New York: McGraw Hill.

Phenice TW (1969) A newly developed visual method of sexing the os pubis. Am. J. Phys. Anthropol. 30:297-302.

Rogers T and Saunders S (1994) Accuracy of sex determination using morphological traits of the human pelvis. J. For. Sci. 39:1047-1056.

SAS Institute, Inc. (1990) SAS® Procedures Guide, Version 6, Third Edition. North Carolina: SAS institute, Inc.

Scheuer JL and Elkington NM (1993) Sex determination from metacarpals and the first proximal phalanx. J. For. Sci. 38:769-778.

Schmidt DJ (1985) Determination of sex with a discriminant analysis of new pelvic bone measurements: part II. J. For. Sci. 30:178-185.

Suchey JM, Wisely DV, Green RF, and Noguchi TT (1979) Analysis of dorsal pitting on the os pubis in an extensive sample of modern American females. Am. J. Phys. Anthropol. 51:517-540.

Sutherland LD and Suchey JM (1991) Use of the ventral arc in pubic sex determination. J. For. Sci. 36:501-511.

Tanner MA and Young MA (1985a) Modeling agreement among raters. J. Amer. Stat. Assoc. 80:175-180.

Tanner MA and Young MA (1985b) Modeling ordinal scale disagreement. Psycholo. Bulletin 98:408-415.

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

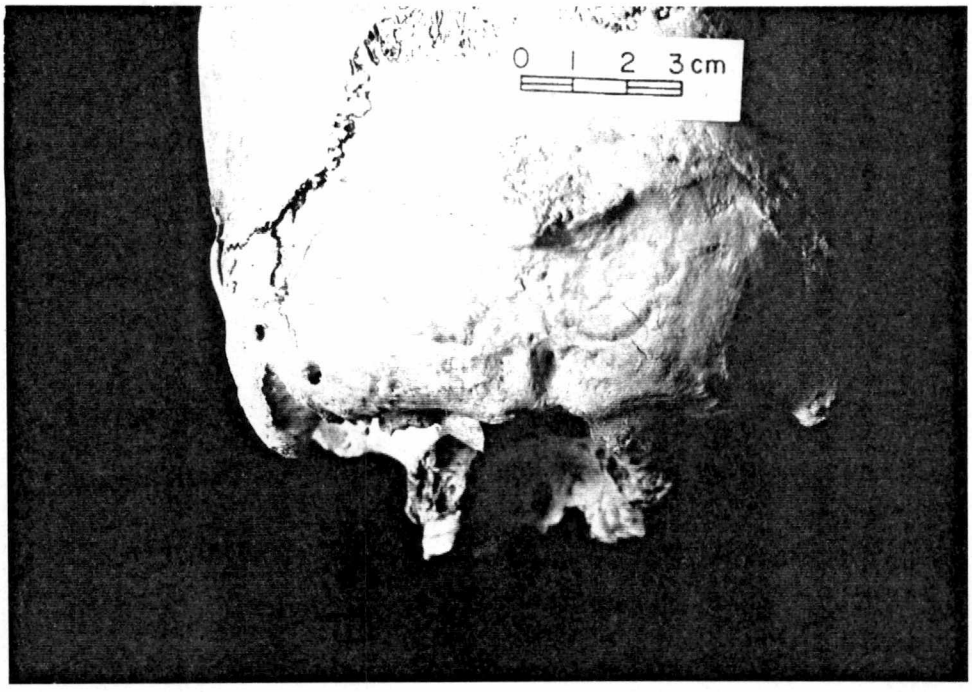
Weiss KM (1972) On the systematic bias in skeletal sexing. Am. J. Phys. Anthropol. 37:239-250.

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

APPENDICES

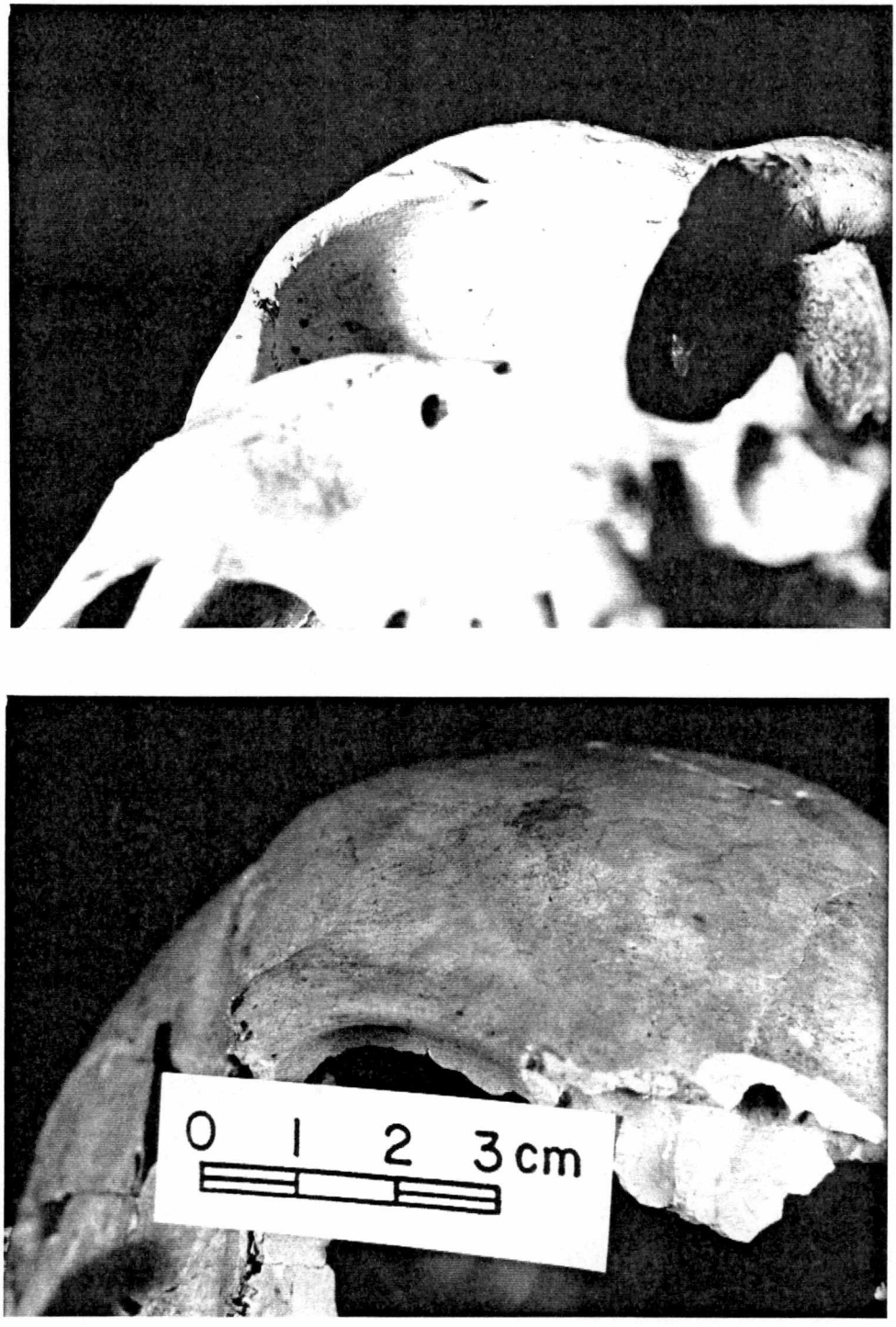
APPENDIX A

EXAMPLE PHOTOGRAPHS OF EACH TRAIT SCORED ON STRONG MALES AND STRONG FEMALES



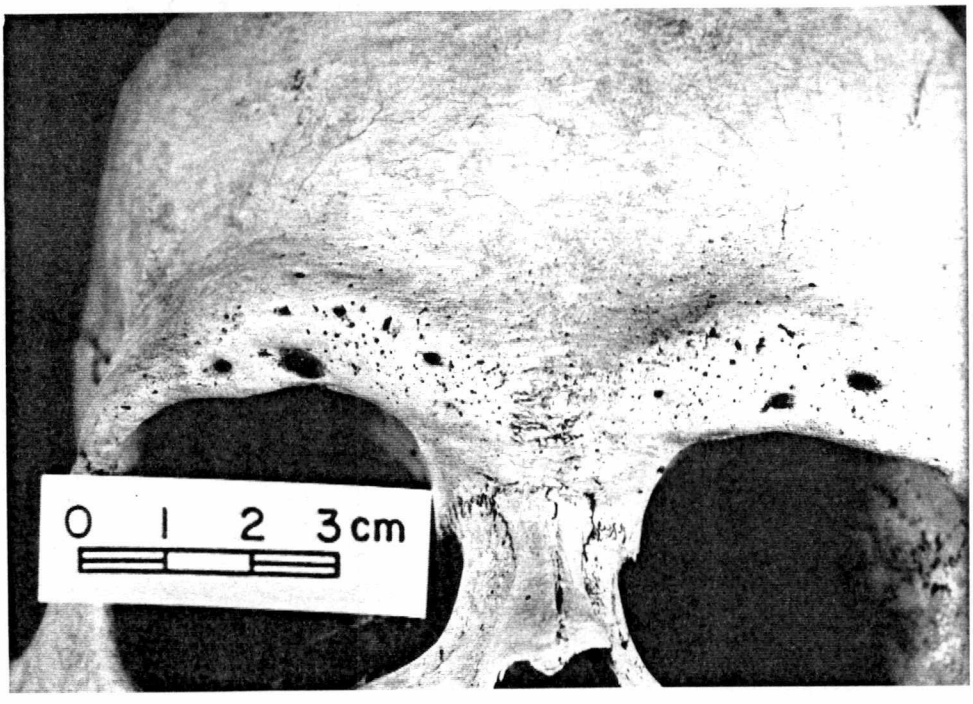
MALE NUCHAL CREST

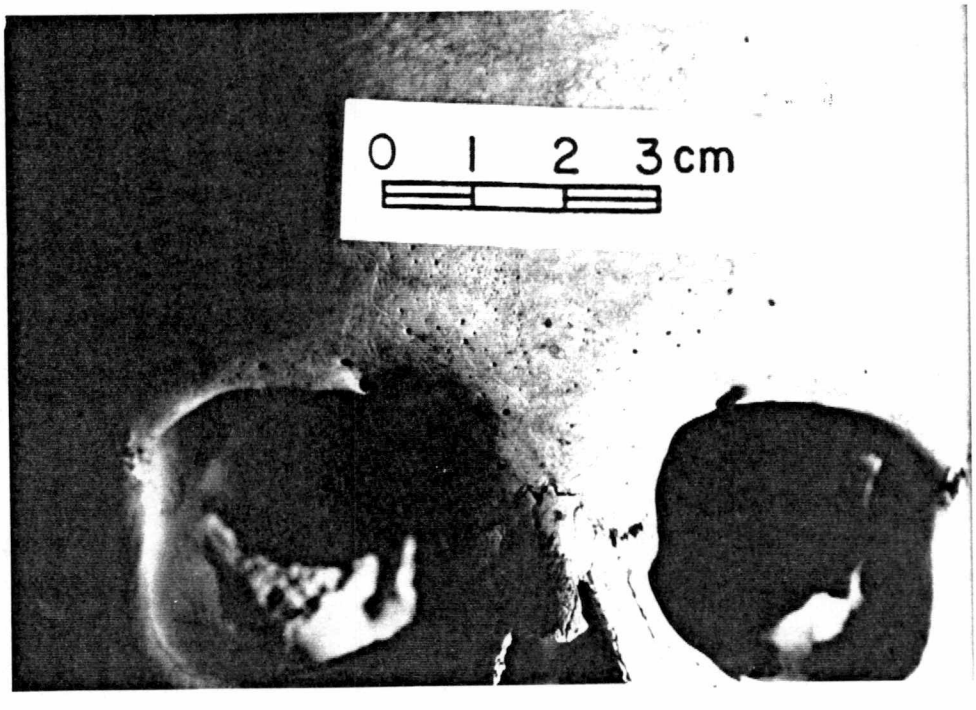
(NO FEMALE SHOWN)



TOP: MALE EYE ORBIT

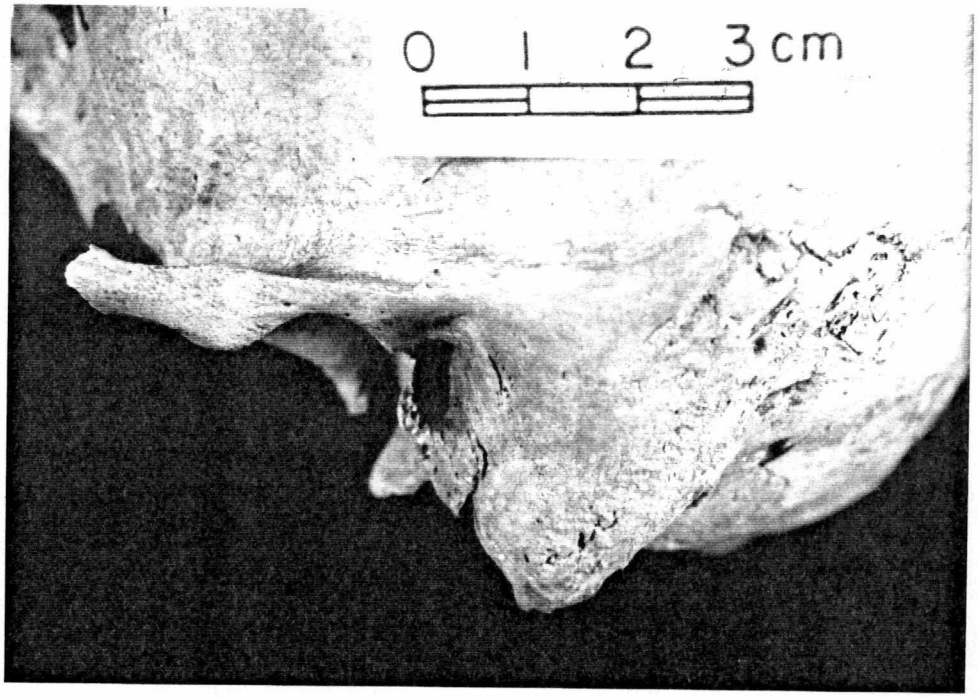
BOTTOM: FEMALE EYE ORBIT

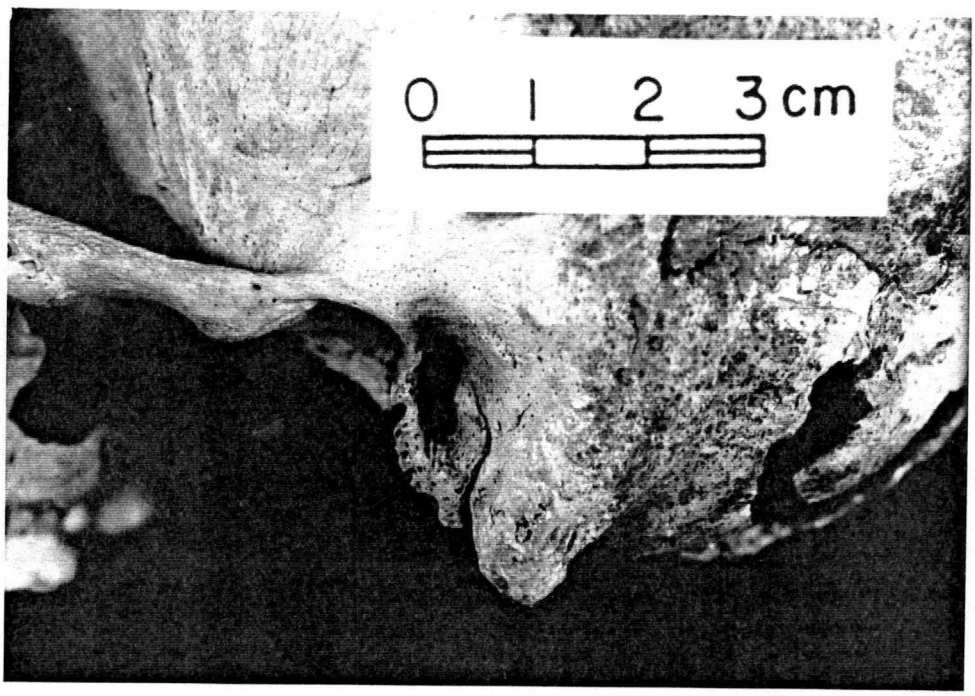




TOP: MALE SUPERCILIARY ARCH

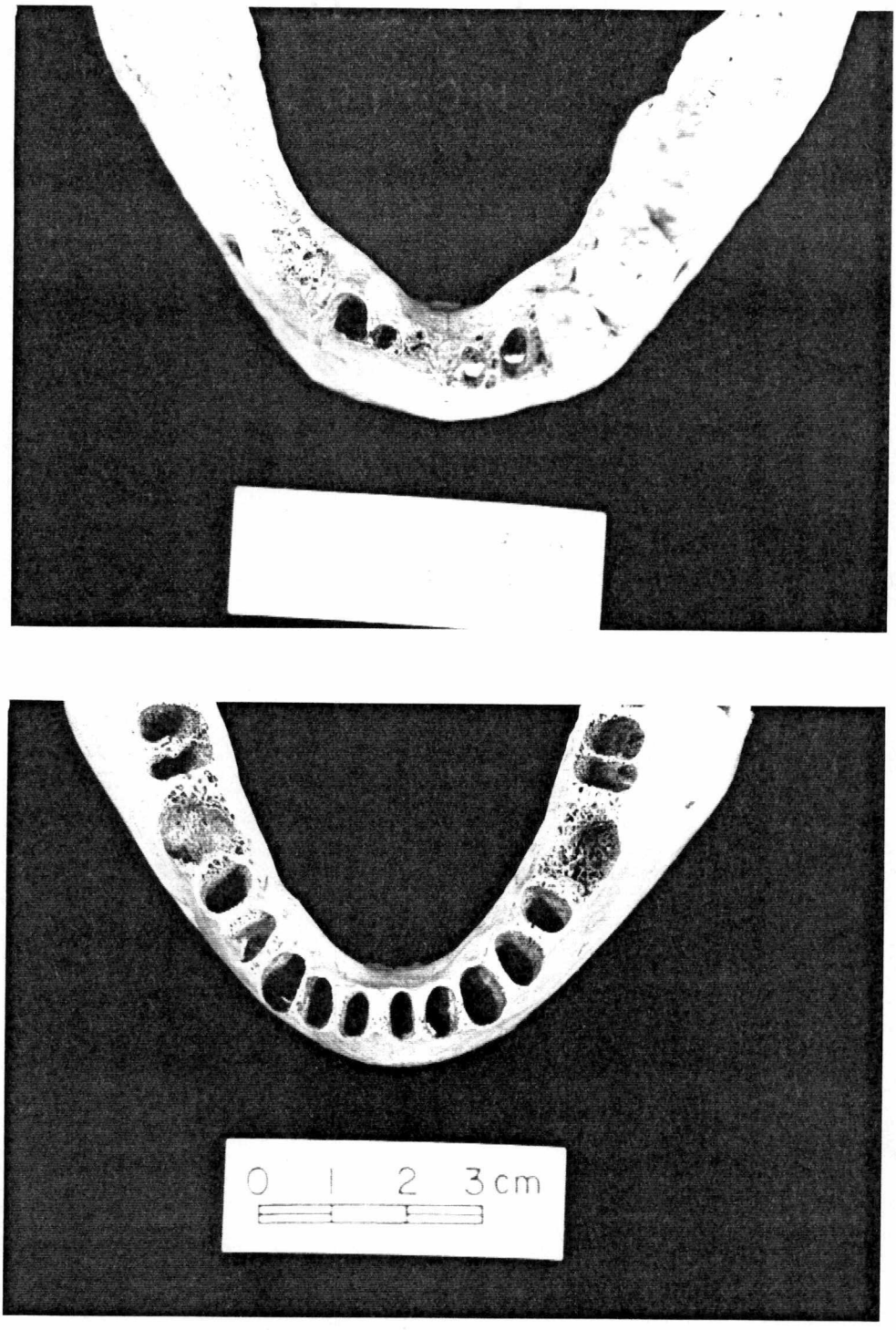
BOTTOM: FEMALE SUPERCILIARY ARCH





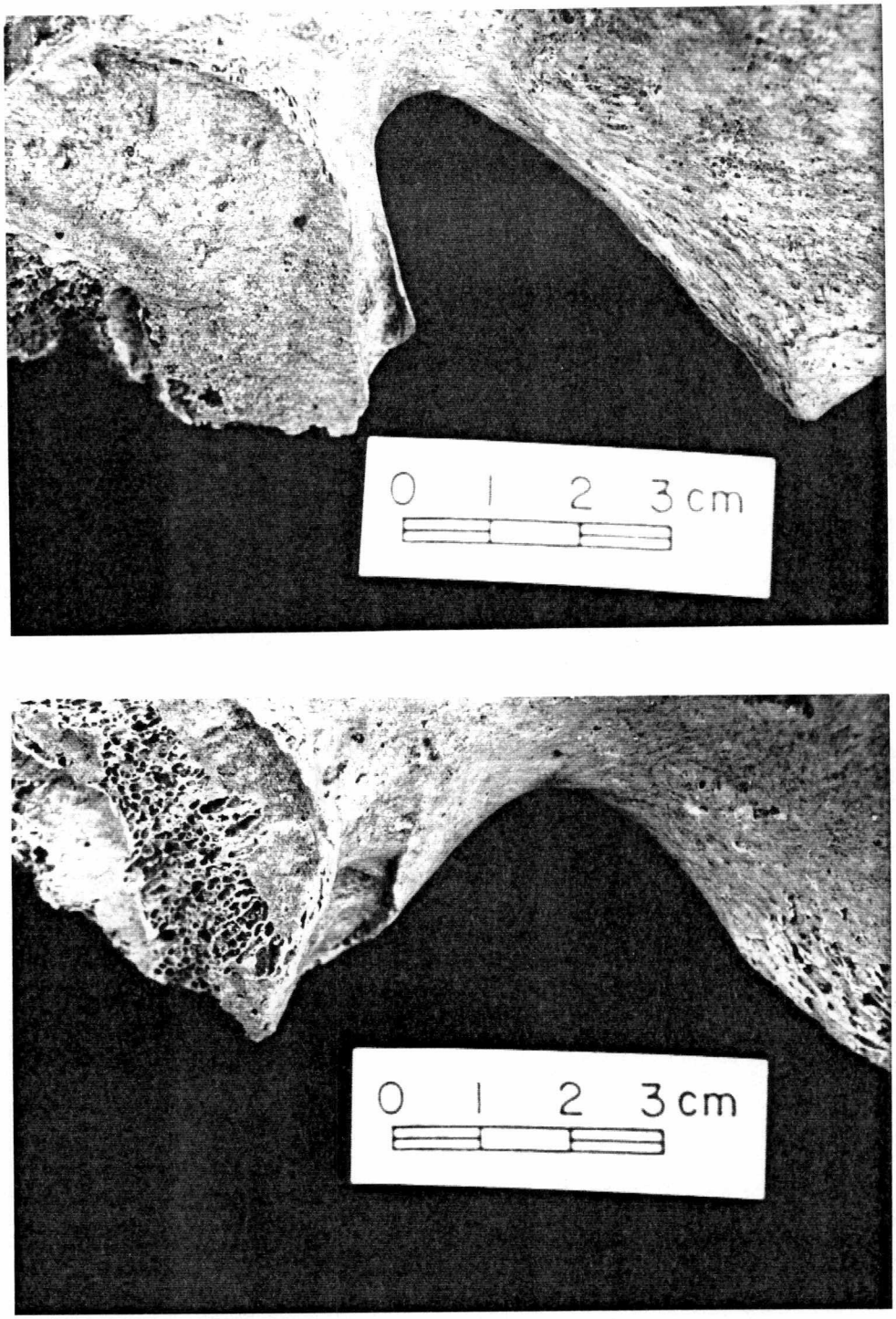
TOP: MALE MASTOID PROCESS

BOTTOM: FEMALE MASTOID PROCESS



TOP: MALE CHIN FORM

BOTTOM: FEMALE CHIN FORM



TOP: MALE SCIATIC NOTCH AND PREAURICULAR SULCUS

BOTTOM: FEMALE SCIATIC NOTCH AND PREAURICULAR SULCUS

APPENDIX B

CROSS-CLASSIFICATION OF OBSERVER RATINGS ON EACH VARIABLE FROM THE AVERBUCH SITE (40DV60) N=148

CHIN FORM

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	23	3	1	1	0	28
M?	14	6	4	1	2	27
?	8	11	7	5	4	35
F?	1	2	1	8	5	17
F	0	5	4	8	18	35
Total	46	27	17	23	29	142

PREAURICULAR SULCUS (LEFT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	32	2	2	0	0	36
M?	6	1	3	0	0	10
?	6	5	5	5	0	21
F?	1	3	4	1	6	15
F	0	0	3	9	24	36
Total	45	11	17	15	30	118

PREAURICULAR SULCUS (RIGHT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	30	1	3	0	0	34
M?	3	4	2	0	0	9
?	3	5	11	2	0	21
F?	0	2	4	5	1	12
F	1	0	1	11	23	36
Total	37	12	21	18	24	112

SCIATIC NOTCH (LEFT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	29	7	2	0	0	38
M?	2	7	3	2	0	14
?	1	1	1	0	1	4
F?	0	4	3	2	7	16
F	0	0	2	5	37	44
Total	32	19	11	9	45	116

SCIATIC NOTCH (RIGHT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	27	9	1	0	0	37
M?	2	7	0	2	0	11
?	2	4	1	2	1	10
F?	0	5	0	6	5	16
F	0	1	3	3	41	48
Total	33	26	5	13	47	122

SUPERCILIARY ARCH (LEFT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	28	6	1	1	0	36
M?	3	16	1	2	0	22
?	0	4	2	0	1	7
F?	0	1	3	4	7	15
F	0	0	0	2	22	24
Total	31	27	7	9	30	104

SUPERCILIARY ARCH (RIGHT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	31	7	0	1	0	37
M?	2	17	3	3	0	25
?	0	2	2	1	2	7
F?	0	2	3	2	3	10
F	0	0	0	3	22	25
Total	31	28	8	10	27	104

MASTOID PROCESS (LEFT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	39	8	0	1	1	49
M?	16	6	2	4	3	31
?	5	6	2	5	4	22
F?	2	6	2	3	5	18
F	0	3	2	0	8	13
Total	62	29	8	13	21	133

MASTOID PROCESS (RIGHT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	45	8	0	0	1	54
M?	13	7	0	4	4	28
?	5	7	1	4	4	21
F?	1	2	2	2	2	9
F	1	1	2	1	8	13
Total	65	25	5	11	19	125

EYE ORBITS (LEFT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	27	14	6	1	1	49
M?	0	5	3	2	2	12
?	0	1	1	3	1	6
F?	0	1	1	6	10	18
F	0	1	0	3	25	29
Total	27	22	11	15	39	114

EYE ORBITS (RIGHT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	27	9	6	5	0	47
M?	0	4	1	4	2	11
?	1	1	1	2	1	6
F?	0	1	0	10	7	18
F	0	3	0	1	29	33
Total	28	18	8	22	39	115

NUCHAL CREST (LEFT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	19	12	2	2	0	35
M?	6	21	3	1	1	32
?	0	9	8	3	3	23
F?	0	3	6	7	5	21
F	0	1	2	2	12	17
Total	25	46	21	15	21	128

NUCHAL CREST (RIGHT)

Observer A	Observer B					
	M	M?	?	F?	F	Total
M	17	14	1	2	0	34
M?	6	21	6	1	1	35
?	0	10	4	3	2	19
F?	0	2	6	6	6	20
F	0	1	2	2	12	17
Total	23	48	19	14	21	125

APPENDIX C

SAS® PROGRAM FOR LOGISTIC REGRESSION ON THE DATA USED IN THIS THESIS

```
OPTIONS LS=72 ERRORS=1;
DATA SK2;
INFILE 'HENS2.DAT';
  INPUT COLLECT$ ID$ SEX PREAURIC SCIATIC BROW MASTOID ORBIT
     NUCHAL CHIN;
PROC LOGISTIC;
     MODEL SEX=PREAURIC--CHIN/SELECTION=NONE DETAILS;
     TITLE 'FORCED MODEL'
PROC LOGISTIC;
     MODEL SEX=PREAURIC--CHIN/SELECTION=STEPWISE DETAILS;
     TITLE 'STEPWISE SELECTION';
PROC LOGISTIC;
     MODEL SEX=BROW--CHIN/SELECTION=STEPWISE DETAILS;
     TITLE 'CRANIAL ONLY:STEPWISE SELECTION';
DATA ABUCH;
  SET SK2;
  IF COLLECT='A';
PROC LOGISTIC;
     MODEL SEX=PREAURIC--CHIN/SELECTION=STEPWISE DETAILS;
     TITLE 'AVERBUCH ONLY, ALL VARS:STEPWISE SELECTION';
```

```
PROC LOGISTIC;
     MODEL SEX=BROW--CHIN/SELECTION=STEPWISE DETAILS;
     TITLE 'AVERBUCH ONLY, CRANIAL ONLY:STEPWISE SELECTION';
RUN;
ENDSAS;
```

APPENDIX D

RAW DATA FOR THE LOGISTIC REGRESSION ANALYSIS

Collection

ID

Sex (0=M, 1=F)

Preauric (-2=M, -1=M?, 0=?, 1=F?, 2=F)

Sciatic

Brow

Mastoid

Orbit

Nuchal

chin

D 10-87	0	-2	-2	1	-2	1	0	0
D 10-88	0	-2	-1	-1	-2	0	0	-2
D 10-89	0	-2	-1	-1	-2	-1	-1	-1
D 10-91	0	-2	0	-1	-2	0	0	0
D 11-87	0	-2	-2	-2	0	-1	0	-2
D 11-89	0	-2	-1	-2	-2	-1	-1	-1
D 11-90	1	1	0	1	-2	2	1	2
D 12-87	0	-2	-2	-1	-2	-2	-1	1
D 12-88	0	-2	-2	-2	-2	0	0	-1
D 12-89	0	-2	-2	-2	-1	-2	-1	-1
D 12-90	0	-2	0	-2	-2	-1	-1	0

D 13-88	0	-2	-1	-2	-2	-1	-1	-2
D 13-91	0	-2	-2	-2	-2	0	-1	-2
D 13-92	0	-2	-2	-1	-2	-1	0	2
D 14-87	0	-2	-2	-2	-2	-2	-1	-2
D 14-88	0	-2	-2	-2	-2	0	-1	-2
D 14-90	0	-2	-2	0	-1	2	0	-1
D 15-87	1	0	2	2	2	2	1	2
D 15-89	0	-2	-2	-1	-2	2	-1	0
D 15-90	0	-2	-2	0	-2	1	-1	0
D 15-91	0	-2	-1	2	-2	2	2	-2
D 16-89	0	-1	-2	-2	-2	-2	-2	0
D 16-90	0	-2	-2	-1	-2	1	-1	-1
D 16-91	0	-2	1	-1	0	1	-1	-1
D 17-88	0	0	0	2	-2	-2	0	1
D 17-90	1	1	1	2	-2	2	-1	2
D 17-91	0	-2	-1	-1	-2	2	-1	-1
D 18-88	0	0	-2	1	-2	1	0	0
D 18-90	0	-2	-2	-1	-2	-2	-2	0
D 19-88	0	-2	-1	0	-2	-1	-1	-1
D 19-90	0	-2	-2	-2	-2	2	0	1
D 19-91	0	-2	-2	-1	-2	2	0	1
D 1-81	0	-2	-1	-2	-2	-1	-2	-2
D 1-82	0	-2	1	-2	-2	1	-1	0
D 1-83	1	0	2	2	0	2	0	1
D 1-84	0	-2	-1	0	-2	2	1	2

D 1-87	0	-1	-1	-1	-2	-1	-1	-2
D 1-88	1	1	1	1	2	2	0	2
D 1-91	0	-2	-2	0	-2	-2	-2	-2
D 1-92	0	-1	-2	1	-1	0	1	0
D 1-93	1	2	1	-1	0	-2	1	2
D 20-88	0	0	1	-2	-2	1	-1	-2
D 20-93	1	0	1	2	2	2	2	2
D 21-90	0	-2	-1	-1	0	0	1	2
D 22-90	0	-2	-2	-2	-1	0	1	2
D 23-88	1	2	2	1	-2	-1	-2	-1
D 24-88	0	1	-1	-1	-1	-2	0	1
D 28-90	1	2	1	1	-2	-1	1	2
D 29-93	0	-1	0	-1	-2	-2	-1	0
D 2-83	0	-2	-2	-1	-2	2	-1	-2
D 2-84	0	-2	2	-2	-1	-1	-1	0
D 2-85	0	-2	-2	-1	0	-1	-1	2
D 2-86	1	1	2	2	-2	-1	1	1
D 2-87	0	-2	-1	-2	-2	2	-2	0
D 2-88	0	-2	-2	-2	1	-2	-1	0
D 2-89	0	0	0	-2	-2	2	-1	-1
D 2-91	0	-1	-1	0	-2	2	0	1
D 2-92	1	1	1	1	0	1	1	0
D 36-93	0	-1	-2	1	-2	2	-1	2
D 3-81	0	-2	-1	-2	-2	2	2	-1
D 3-83	0	-2	0	-2	-2	0	-2	0

D 3-87	0	0	-1	-1	-1	0	-1	0
D 3-89	0	0	0	1	-2	1	1	2
D 3-90	0	-1	-1	-2	-1	0	-2	-1
D 3-91	0	-2	-1	-2	-2	-1	0	-1
D 3-92	0	-2	-1	-1	-1	-1	-1	-2
D 4-81	1	0	1	1	-2	2	-1	0
D 4-83	.	-2	-2	-2	-2	0	-1	-2
D 4-87	0	0	0	-2	-2	-2	-1	-1
D 4-88	0	-2	-1	0	-1	0	0	1
D 4-89	0	-2	-1	-2	-2	-2	-2	0
D 4-90	0	-2	-2	-1	-2	-1	0	1
D 4-92	0	0	1	-2	-2	-2	2	1
D 5-83	0	-2	-1	-1	-2	0	-1	2
D 5-87	1	2	2	0	0	2	-1	-1
D 5-88	0	0	-1	0	-2	2	1	-2
D 5-91	0	-2	-2	0	-2	2	1	1
D 6-87	0	-2	-1	1	-2	2	0	2
D 6-88	1	1	1	2	0	2	0	2
D 6-89	1	1	1	1	2	2	0	1
D 6-91	0	-2	-1	-1	-1	2	-1	0
D 6-92	1	1	2	0	2	0	-1	1
D 7-86	0	-2	-1	-2	-2	-1	0	0
D 7-87	0	-1	0	-2	-2	-1	-1	-2
D 7-89	0	-2	0	-2	-2	1	-1	2
D 7-91	0	-2	-2	-1	-2	1	-1	0

D 7-92	1	0	1	1	1	1	1	0
D 8-87	0	-1	2	-2	-1	-1	-2	-2
D 8-89	0	-2	-1	-2	-2	-1	0	-1
D 8-91	0	-1	-2	-1	0	-2	0	0
D 9-88	0	-1	0	-1	-2	-1	0	-2
D 9-89	0	-2	-2	-2	-2	-1	-2	-2
D 9-93	0	-2	-1	-1	-1	0	2	2
A 6	.	-1	-2	.	-2	-2	-1	-2
A 7A	.	-1	1	-1	-2	2	2	0
A 8A	.	1	2	.	.	.	.	-2
A 9A	.	1	2	2	1	2	2	2
A 10A	.	0	1	2	2	2	1	1
A 11A	.	1	1	.	-1	-1	0	-1
A 12	.	.	.	.	.	.	-1	0
A 19A	.	0	-2	.	-1	.	-1	-1
A 19B	.	-2	-2	.	.	.	-1	2
A 20	1	0	2	-1	2	1	1	2
A 25A	1	2	2	.	2	.	1	2
A 28	.	-2	2	.	-1	.	-2	-1
A 29A	0	-1	-2	.	-1	.	-2	-2
A 32	.	.	.	.	2	.	-1	2
A 33A	.	-1	-1	-2	1	-2	-1	-2
A 35B	1	1	1	2	2	2	1	1
A 100	1	2	2	2	2	2	1	2
A 101	0	-1	-2	-2	-1	-2	0	-2

A 102	1	-2	2	2	2	2	2	2
A 103	0	-1	-1	-2	-2	-2	-2	-2
A 107B	1	2	2	2	-1	-1	-1	-1
A 113	0	.	.	.	1	-2	0	-2
A 118A	1	1	2	1	-1	1	-1	2
A 118B	.	.	-2	.	-1	.	-2	-2
A 121A	0	1	-2	0	-2	-1	-1	-2
A 121B	0	0	-2	-1	-2	-1	-1	-2
A 122	1	2	2	2	1	2	.	-2
A 126A	.	1	2	.	1	.	.	1
A 126B	1	2	2	.	-1	1	-1	-2
A 127	.	1	-1	-2	-2	1	-2	-2
A 137A	0	-1	-2	-2	-2	1	-1	-1
A 139	0	-2	-2	-2	-2	-2	-2	-1
A 141A	.	-2	-2	.	-1	.	0	0
A 141B	.	2	2	1	1	2	2	1
A 143	.	-2	-2	.	-2	.	-1	-1
A 146A	0	-2	-1	2	2	2	-2	-2
A 146B	1	1	2	-1	-2	-1	.	.
A 147A	.	-2	0	-1	-2	1	-1	-2
A 147C	.	.	.	0	-2	-2	0	0
A 148	.	-2	0	-1	-2	-2	-2	0
A 149	1	-2	-1	-2	-2	-2	.	0
A 152A	0	-2	-2	.	0	.	-1	-1
A 152B	.	2	2	.	1	.	2	1

A	156A	.	1	2	2	2	2	2	1
A	156B	0	.	-2	-2	-2	0	-1	0
A	157	.	0	.	-1	-2	-2	-2	0
A	158A	1	2	2	1	-2	2	-1	2
A	160B	0	.	.	-1	-2	2	2	1
A	164	1	2	2	2	1	2	1	2
A	166	.	-2	-1	-2	1	0	-1	0
A	168	.	-2	-2	-2	-2	-1	-1	2
A	170	1	1	1	2	0	2	1	2
A	171	.	-2	-1	-1	-2	1	.	1
A	174	.	2	2	2	2	2	.	1
A	176C	0	2	2	2	-2	2	-1	.
A	180	0	-2	0	-1	-1	-2	-1	-2
A	181	0	-2	-2	-1	-2	0	-2	-2
A	185A	0	-2	0	-1	-2	2	0	-2
A	191	.	.	.	-2	-2	-1	0	-2
A	194A	.	2	2	2	-2	2	2	1
A	194B	1	0	2	0	2	2	-1	-2
A	196	.	-2	0	0	-1	-1	.	-2
A	197A	0	-1	-1	-1	-2	1	-2	-1
A	200	.	1	2	.	1	.	0	.
A	203A	.	-2	-1	-2	-2	1	-2	-2
A	204A	.	-2	-2	.	-2	-2	-1	-2
A	205A	.	-2	-2	-2	-2	0	-2	-2
A	207	1	0	2	-1	-1	-2	.	0

A 212	1	1	1	.	0	.	2	1
A 218	1	2	2	2	-2	2	1	2
A 220A	1	0	2	2	-2	2	2	0
A 220B	0	-2	1	-1	-2	-2	-1	-2
A 225A	0	-2	-2	-2	-2	-2	-1	0
A 225B	.	.	.	-2	-2	-1	0	-2
A 229A	.	0	2	2	2	2	1	2
A 229B	0	-2	-2	-2	-2	0	-1	-2
A 230	.	0	-1	-2	-2	-1	-2	2
A 232B	.	.	.	1	-2	2	2	2
A 236	.	.	.	-1	-2	-1	-1	-1
A 242A	.	1	2	2	-1	2	.	-1
A 243	.	.	.	2	-1	-1	2	1
A 248	0	-2	-1	-1	-1	1	0	-2
A 251	1	0	2	-1	-2	1	0	0
A 255	1	2	2	2	-2	1	0	0
A 258	0	-2	-1	-2	-1	0	-1	-1
A 262	.	.	1	-1	.	-1	-1	-2
A 263A	.	2	2	1	2	.	0	2
A 264A	1	1	2	0	-1	0	-2	-1
A 264B	1	1	2	2	-2	2	-1	2
A 271A	.	0	2	1	-1	2	2	1
A 273A	1	-1	-2	-1	-2	-2	-2	-2
A 273B	.	2	2	1	0	2	-1	2
A 277A	1	0	-1	-2	-2	-2	0	1

A	280A	1	2	2	0	2	1	2	1
A	282A	.	.	2	2	-1	2	1	2
A	285	.	-2	-1	-1	-1	2	-1	-2
A	287A	0	-1	-1	-2	0	2	-1	-2
A	287B	.	-2	0	-1	-1	-2	-2	.
A	288A	1	1	0	2	-1	2	2	1
A	291B	1	-2	-2	-2	-2	-1	-1	-1
A	293	1	2	2	2	-2	2	0	2
A	295A	.	.	.	-2	-2	-2	-2	.
A	297	0	.	-2	-2	-2	-1	-1	-2
A	299A	.	0	2	2	0	2	.	2
A	300	.	0	2	2	2	2	.	0
A	301A	1	-2	-1	.	-1	2	1	1
A	311A	.	2	2	.	-2	2	2	1
A	311B	0	0	2	-2	-2	-2	-2	-3
A	315A	0	-2	-1	-2	-1	-2	-1	-2
A	318	0	-2	1	-1	-2	-2	0	-2
A	319A	.	0	2	.	-1	2	2	1
A	319B	1	2	2	.	-1	.	2	2
A	328	0	-2	1	-2	-2	-2	-1	-2
A	330A	0	-2	-2	.	-2	.	-2	-2
A	332	0	-2	0	-1	-2	0	0	-1
A	333	.	.	.	0	-2	1	1	-2
A	334	.	2	2	0	-1	2	-1	-1
A	335A	0	-2	-2	-2	-2	1	0	-1

A	341A	0	-2	-1	-2	-2	-1	-2	-1
A	341C	.	2	2	1	-2	1	-1	-1
A	349	0	0	-1	-1	-1	-1	-1	-1
A	355	.	0	1	-1	-2	0	-1	-1
A	356	1	2	2	2	-2	2	2	2
A	359A	1	0	0	2	-1	0	-2	-1
A	374	0	-2	-2	-1	-2	-2	-1	-2
A	377	0	-2	-1	-1	-2	0	-1	0
A	401A	1	0	2	-1	-1	2	-2	.
A	403	0	-2	-2	1	-2	2	0	1
A	417A	1	0	2	0	-2	-1	0	2
A	432A	1	2	2	2	0	1	1	1
A	432B	.	.	1	-2	-2	-1	-1	.
A	488A	1	1	2	.	1	.	-1	2
A	488B	.	.	.	0	1	-2	1	2
A	489	.	.	-1	-1	-2	-2	.	-2
A	490	0	-2	-2	-1	-1	-1	-1	-1
A	498	1	0	2	2	2	2	2	-2
A	505	1	2	2	1	2	2	2	2
A	508	.	2	1	.	1	-1	-1	0
A	511	.	-1	-1	1	-1	2	2	-2
A	512	.	2	2	2	2	2	1	2
A	514	1	1	2	2	2	1	1	1
A	515	.	-2	-2	-2	-2	-2	-2	-2
A	516A	.	-2	-2	-1	-1	-2	.	-1

A	516B	.	-2	-2	-1	-1	-2	0	-2
A	520	.	-2	-2	-2	-1	-1	-1	-1
A	521	0	-1	-2	-2	-2	-2	-1	-1
A	524	1	2	2	2	2	-1	1	1
A	525	0	-1	-1	-2	-1	-2	-2	-2
A	565	0	-2	-2	.	-2	-1	0	-1
A	684	0	1	-2	.	-2	-2	-1	.
A	700A	0	-2	-1	.	1	.	1	2

VITA

Samantha Marie Hens was born in Buffalo, New York on November 7, 1968. She attended elementary school in the Buffalo area and graduated from McKinley High School in June, 1986. That summer she began studies at The State University of New York at Buffalo and in May of 1990, received the degree of Bachelor of Arts in Anthropology with departmental honors. After a year of part time study, she entered into the graduate program in Anthropology at The University of Tennessee at Knoxville. She specializes in the subdiscipline of physical (biological) anthropology.

She plans to graduate with a Master of Arts degree in Anthropology with a minor in Statistics in December, 1994. Thereafter she will be enrolled in the PhD program at The University of Tennessee at Knoxville.