2008-07-17

Assessing Multivariate Heritability through Nonparametric Methods

Benjamin Alan Carper
Brigham Young University - Provo

ASSESSING MULTIVARIATE HERITABILITY THROUGH NONPARAMETRIC METHODS

by

Benjamin A. Carper

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

August 2008

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Benjamin A. Carper

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	G. Bruce Schaalje, Chair
Date	William F. Christensen
Date	C. Shane Reese

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Benjamin A. Carper in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

G. Bruce Schaalje
Chair, Graduate Committee

Accepted for the Department

Scott D. Grimshaw
Graduate Coordinator

Accepted for the College

Thomas W. Sederberg
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

ASSESSING MULTIVARIATE HERITABILITY
THROUGH NONPARAMETRIC METHODS

Benjamin A. Carper
Department of Statistics
Master of Science

The similarities between generations of living subjects are often quantified by heritability. By distinguishing genotypic variation, or variation due to parental pairings, from phenotypic variation, or normal intraspecies variation, the heritability of traits can be estimated. Due to the multivariate nature of many traits, such as size and shape, computation of heritability can be difficult. Also, assessment of the variation of the heritability estimate is extremely difficult. This study uses nonparametric methods, namely the randomization test and the bootstrap, to obtain both a measure of the extremity of the observed heritability and an assessment of the uncertainty.

ACKNOWLEDGEMENTS

I'd like to thank my family, especially my parents, for all the support they have given me. I'd also like to thank the the faculty and staff of the Department of Statistics at Brigham Young University. A special thanks goes to my advisor, Dr. G. Bruce Schaalje for helping throughout the entirety of my project. A final acknowledgement goes to my fellow students, without whom I would never have been able to handle the pressure of my schoolwork and project.

CONTENTS

TABLES

Table

FIGURES

Figure

1. INTRODUCTION

Quantitative genetics uses statistical models to analyze the inheritance of traits (Klingenberg 2003). When the trait in question is *shape*, the influence of progenitors on the current generation can range from the visibly obvious to the minutely imperceptible, both of which are difficult to quantify. The statistical problem is separating the variability due to the parents' genetics from the random variation associated with an individual's environment. These two types of variation are denoted as genotypic for the parental effect and environmental for the species-environmental variation. The sum of genotypic and environmental variation is called phenotypic variation (Lande 1979). Even for simple traits and experimental designs, these variance components are often very hard to distinguish statistically. The problem becomes even harder when the trait in question is multivariate (such as shape). Recent developments in multivariate mixed methods have made this task easier. After adjusting subject measurements for size and converting the adjusted physical traits of the subjects to the most essential measures such as principal components, the separation of the phenotypic variation, denoted $\mathbf{P}$, from the genotypic variation, denoted $\mathbf{G}$, becomes somewhat easier.

While some progress has been made toward estimating these two types of variation, there are still many limitations on accuracy, for as the number of utilized principal components increases, the necessity of greater computational resources grows, requiring greater and greater time for computation and larger and larger amounts of computing power (Myers et al. 2006). Another difficulty is the lack of a reliable measure of uncertainty for the computed measures of variation and ultimately for heritability, a univariate summary of the proportion of overall generational variation attributable to genotypic influences. Due to the apparent difficulty of computing heritability estimates for even simple models, methods such as bootstrapping and other

permutational measures of uncertainty seem to be virtual impossibilities. In order for any real progress to be made toward the desired results of these studies of heritability, methods must be devised that can compute variance components for higher order models with much less time and computational power.

This study uses *R* (R Development Core Team 2008) code to call an efficient mixed model program designed specifically for genetic applications. This code is used to compute randomization tests and three forms of the bootstrap to obtain confidence intervals for multivariate heritability measurements, providing a range of possible values and a means of differentiating the heritabilities of different species.

These methods are applied to shape measurements for 109 western mosquito fish (*Gambusia affinis*) from a breeding study involving full siblings, half siblings, and feeding treatments. These shape data are shown in Figures 1.1 and 1.2. Relative warps are the principal components of the landmark-based shape measurements. The analysis is performed on the first ten relative warps, which account for approximately 90% of the shape variation.

Figure 1.1: Relative Warp by Warp Number(Diet=High). Each line represents a different fish

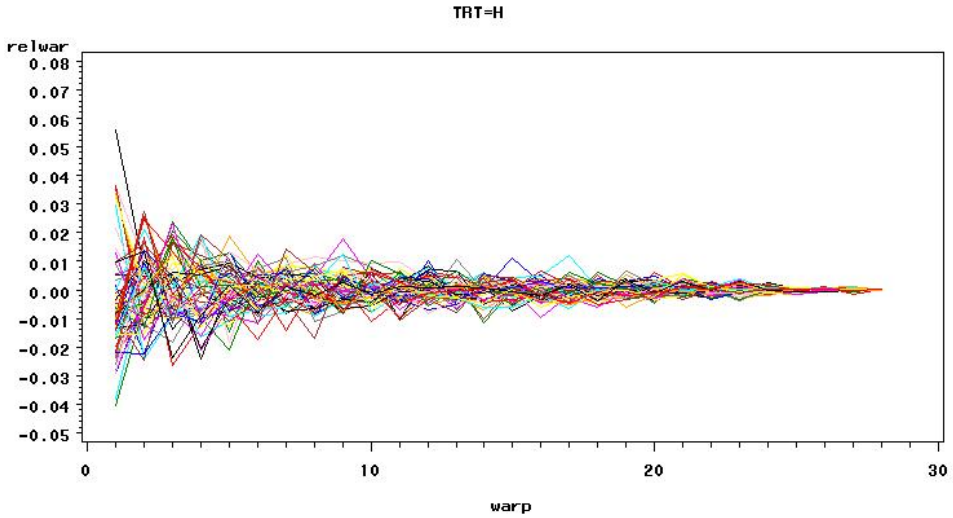
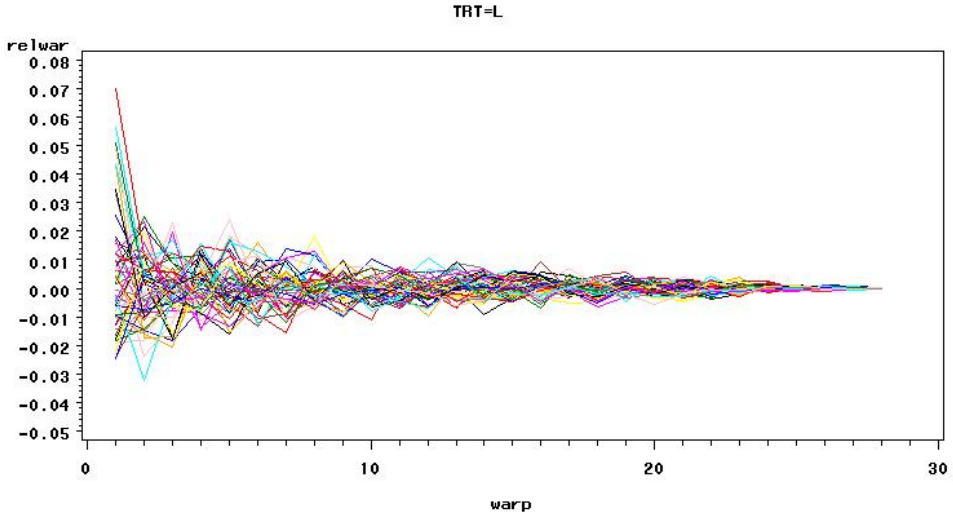


Figure 1.2: Relative Warp by Warp Number(Diet=Low). Each line represents a different fish



2. LITERATURE REVIEW

Heritability is defined as the proportion of the overall variation in a population that can be attributed to genetic variation among individuals. The genotypic variation can often be hard to distinguish from the overall phenotypic variation, which responds to environmental factors in addition to genetics. By measuring the trait of interest on a given generation of siblings and nonsiblings, the within-family and across-family variation can be used to obtain heritability estimates. In the case of shape measurements of mosquito fish, positions of various landmarks can be used to obtain covariance components for shapes of fish.

2.1 Univariate Heritability

The idea of heritability was originally developed in the univariate setting (Lande 1979). It is a numerical measure of the relationship between the mean response for a specific character or trait and the selection differential, or the difference in the means of selected and unselected individuals. Selection is the choosing of individuals as breeding stock based on desired physical traits. This relationship is defined in the Breeder's Equation (Lande 1979):

$$\Delta \bar{z}_b = (G_{bb}/\sigma_b^2) S_b = h_b^2 S_b, \tag{2.1}$$

where $\Delta \bar{z}_b$ is the change in character or trait z_b, G_{bb} is the additive genetic variance, σ_b^2 is the overall, phenotypic variance, and S_b is the selection differential of the trait (Lande 1979).

Estimation of these variance components is not easy. The phenotypic variance is the only component that can be measured directly. Genetic and environmental variance estimation requires a variety of sophisticated statistical techniques, including mixed linear models with maximum likelihood or restricted maximum likelihood

(REML) techniques. In addition, the asymptotic standard errors of these estimates are very inaccurate, especially with small sample sizes and even small amounts of imbalance. Obtaining large samples and balance is, more often than not, very difficult in animal breeding experiments.

In addition to these difficulties, when applied to a multivariate trait, there are assumptions required by the univariate method which often are not satisfied. The univariate approach assumes the traits measured on an individual are independent from one another. The univariate approach also requires larger designs, a requirement which is often difficult to satisfy, depending on the species in question. These limitations in application and inference led to the development of a multivariate generalization of the Breeder's equation and, consequently, heritability.

While the univariate method provides useful summary statistics for separate variables (Klingenberg 2003), it has less power than fully multivariate methods that simultaneously consider all the aspects of a trait.

2.2 Multivariate Heritability

Many traits are inherently multivariate. Lande (1979) proposed a multivariate generalization to the Breeder's equation. Defining $\mathbf{G}$ as the additive genotypic variance-covariance matrix, $\mathbf{P}$ as the total phenotypic variance-covariance matrix, and $\mathbf{s}$ as the vector of selection differentials, the Breeder's equation becomes

$$\boldsymbol{\Delta}\bar{\mathbf{z}} = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}. \tag{2.2}$$

This version of the Breeder's equation accounts for the multivariate nature of traits and heritability. The eigenvalues or other functions of $\mathbf{G}\mathbf{P}^{-1}$ can be used to summarize heritability of multivariate traits (Klingenberg and Leamy 2001). The dominant eigenvalue is often interpreted as the maximum heritability estimate. While both the genetic covariance matrix and the phenotypic covariance matrix are symmetric,

the matrix $\mathbf{GP^{-1}}$ may not necessarily be symmetric. Hence, caution must be used to obtain eigenvalues of this matrix. Myers et al. (2006) found the maximum heritability by scaling the maximum eigenvalue by the sum of the eigenvalues

Another measure for summarizing multivariate heritability was developed by Klingenberg and Monteiro (2005). In this method, heritability is taken as

$$h^2 = tr(\mathbf{G})/tr(\mathbf{P}), \tag{2.3}$$

the ratio of the sum of the diagonal elements of $\mathbf{G}$ and $\mathbf{P}$. Klingenberg and Monteiro (2005) note that this method does not consider the direction of the selection differential or the directionality of variation in $\mathbf{G}$ and $\mathbf{P}$. It ignores covariation among the traits.

The multivariate approach to heritability is currently widely used and accepted in the analysis of quantitative genetics; however, it has its drawbacks. It is very difficult to find the standard errors of eigenvalues. While imbalance is fairly easily accommodated in the multivariate setting, the inclusion of multiple traits greatly increases the number of variance and covariance components to be estimated, and consequently the computation time. The lack of reliable standard error estimates has led to the use of resampling techniques in connection with multivariate analysis. Myers et al. (2006) employed a permutation test to assess significance of heritabilities; however, due to the computational intensity and amount of time required for each permutation, they only used 999 permutations to obtain these assessments.

2.3 REML for Unbalanced Data

The multivariate method can more easily accommodate imbalance in study design due to the implementation of REML. REML has become the preferred method for estimating variance components of the mixed models used in quantitative genetics. REML uses all available information about relationships among individuals; for

example, parent-offspring and full-sibling and half-sibling relations (Klingenberg and Leamy 2001). Unrestricted REML estimates are asymptotically unbiased and have the same asymptotic distributional qualities as maximum likelihood.

Various programs have been developed for estimating the **G** and **P** matrices. Analyses have been performed using SAS Proc Mixed by considering the sires and dams, or parental pairings, as random effects; however, SAS is computationally inefficient when there are nested variance and covariance components. As more traits are considered, the time required to obtain the covariance matrices increases exponentially.

A few programs specific to the study of quantitative genetics and heritability have been developed to reduce the computation time required by the multivariate approach. Programs such as VCE5 (Covač and Groeneveld 2003) and WOMBAT (Meyer 2008) employ REML methods to compute variance components. VCE5 uses the method of analytical gradients (Klingenberg and Leamy 2001) and has a marked improvement in computation time over SAS while producing similar estimates.

2.4 Shape Analysis and Heritability

Heritability is often of interest in the analysis of size and shape. While size is an important attribute, shape is much more difficult to quantify and analyze. Many different methods have been used to measure shape differences. The univariate approach to shape analysis uses a univariate transformation of shape data called the Procrustes distance (Klingenberg 2003), which assumes there is no directional variation in the landmarks. When shape heritability is considered, this assumption is grossly unrealistic. Landmarks measured on individuals typically have more than just a distance or length component — there is usually a directional component.

The multivariate approach to shape analysis makes no assumptions about the independence or homogeneity of landmarks. As an example, Klingenberg and Leamy

(2001) performed shape analysis on the mandibles of a set of laboratory mouse strains. The x and y coordinates of eleven landmarks were recorded for over one thousand individuals. Due to the large number of measurements on each subject, the data were transformed into principal components to reduce the dimensionality of the measurements. They noted that this reduction resulted in little, if any, loss of estimability.

A similar procedure was performed by Myers et al. (2006) on the plastron shape of slider turtles. The plastron shape was quantified using similar landmark-based morphometric methods, but the measurements were not transformed to principal components. Over 1300 hatchling slider turtles were used in the study. Myers et al. (2006) were able to implement the program VCE5 to obtain estimates of heritability of shape for two different locations. They also used a randomization test to assess the statistical significance of the observed heritability; however, as noted before, they used relatively few permutations to perform this test of significance.

2.5 Resampling Methods

Nonparametric methods require minimal assumptions about the form of the distribution of the population (Higgins 2004). While there are many different classes of nonparametric statistics, of particular interest to this study are permutation and resampling methods. Due to the inability to gather more data, these methods can be of great use to this study because of their simplicity, ease of implementation, and desirable statistical properties.

Permutation-based methods such as the randomization test provide a way to assess statistical significance when distributional and asymptotic assumptions are not met. These tests involve permuting the responses onto the existing treatments or, as in the case of this study, onto the existing sire-dam pairs. For each of many permutations the sample statistic is calculated and recorded to create a reference distribution. The reference distribution is used to find the relative quantile position of the observed

test statistic. The p-value is calculated as the proportion of observations from the reference distribution that are as extreme or more extreme than the observed test statistic. The randomization test is primarily used to perform tests of significance.

Another nonparametric method of interest to this study is the bootstrap. Bootstrapping is used to estimate the sampling distribution of an estimator or test statistic. The basic idea behind bootstrapping is to sample observations from the observed data *with replacement* to create a new data set. By repeating this, an empirical sampling distribution of the statistic of interest can be developed. These methods are particularly useful in situations such as heritability where the distributional qualities are not very well known since the bootstrap requires no distributional assumptions. Bootstrapping is most often used to obtain robust estimates of standard errors and confidence intervals.

This study also makes use of a multistage bootstrap, which has the same aims as the basic bootstrap. The basic idea and usage of this estimator is to sample from the data while maintaining the design structure. In this study, this is performed by sampling from sires, then sampling from the dams respective to those sampled sires, and likewise for the offspring of the sampled sire-dam pairs. The goal behind this bootstrap is to preserve the designed family structure that was present in the original data. The unit of the bootstrap is particularly important in heritability studies. The basic bootstrap can be carried out with offspring, dams, or sires as the bootstrap unit. All of the previously mentioned nonparametric methods are applied in this study using mosquito fish landmark data.

3. DATA AND METHODS

3.1 Data Description

The data used in this study consists of the 28 principal components of the landmark measurements of 109 mosquito fish, although this study only considers 10 principal components. In addition to these measurements, for each fish there is a record of the parental pairing from which the fish is offspring. There were 8 sires, or males, and 21 dams, or females, in the parental generation. Dams were nested in sires, with an average of about three dams paired with each sire. There are no measurements of landmarks for the parental generation, and there is no record of gender of the individual fish. Each fish was randomly assigned one of two diets.

The study design was a mixture of a full-sibling design and a half-sibling design, meaning one sire was mated with multiple females, while females were mated with unique sires. This created full siblings where fish shared the same two parents and half-siblings where fish only shared a common sire. The data were obtained and analyzed as a precursor to the June Sucker Recovery Project, where efforts to identify heritabilities for different subspecies may help to classify them as endangered and thus eligible for recovery funding.

3.2 Methods

There are difficulties in properly estimating heritability in the multivariate setting. While point estimates are attainable for the data, standard errors and significance tests of most estimates are difficult to obtain, especially with smaller samples and imbalance. Various methods have been used to assess the statistical significance of the estimated heritability, but due to the computational intensity, confidence in-

tervals have rarely been reported.

This study employs nonparametric methods in assessing significance as well as obtaining confidence intervals. These nonparametric methods require none of the distributional assumptions that often plague other methods and are much less sensitive to influential observations.

3.2.1 Randomization Test

The first nonparametric method employed in this study assesses the statistical significance of the heritability estimate for the mosquito fish data. We randomly sampled 109 observations from the data without replacement and then assigned the responses to the existing diet and sire-dam combinations. The entire vector of 10 shape principal components for a given subject was kept together and randomly assigned to an existing parental pairing 1000 times. For each permutation, we recalculated and recorded both measures of multivariate heritability. The resulting sets of heritability estimates were considered to represent empirical null distributions from which p-values are calculated.

3.2.2 Basic Bootstrap

In order to obtain an estimate of the standard error of heritability as well as a confidence interval for heritability, we used the bootstrap. The idea behind the bootstrap is similar to that of the randomization test in that by repeatedly sampling from the original data, an empirical, assumption-free sampling distribution of the parameter of interest is attainable. The basic bootstrap consisted of sampling the principal component measurements as well as the accompanying diet and sire-dam pairing of each fish from the original data. Unlike the randomization test, the samples were taken *with replacement* from the 109 individuals, meaning that a given individual could appear in the resultant data set multiple times, or not at all.

This resampling was repeated 1000 times. Heritability was recalculated for each data set. A confidence interval was estimated from the final distribution of heritabilities by picking out the relevant percentiles of the empirical measurements.

3.2.3 Multistage Bootstrap

Due to the sire-dam structure of the data, a second, modified bootstrap technique was also used to obtain confidence intervals. Termed a *multistage bootstrap*, this method makes use of the family groupings. The basic bootstrap by nature produces bootstrap data sets with 109 observations but varying numbers of sires, dams per sire, and offspring per dam. A comparison to the multistage bootstrap can help to determine the importance of the family structure in the bootstrap analysis of the mosquito fish data.

The procedure for the multistage bootstrap consisted of performing a basic bootstrap of the eight unique sires followed by a basic bootstrap within each of their respective dams. Once the sire-dam pairs were thus sampled, a basic bootstrap was performed on the offspring belonging to the sampled sire-dam pair. To ensure that the same number of subjects (109) was obtained for each bootstrap data set, the original numerical structure present in the data was used as the basis for the number of bootstrap samples at each step. As an illustration, consider the first sire from the original data set. Suppose that this sire was mated with three dams with offspring counts of eight, six, and six, respectively. Then the dams corresponding to the first sampled sire were resampled three times, and from each of these three dams (not necessarily unique) the respective children were resampled eight times, six times, and six times, respectively. This sampling structure maintained the original size and structure of the mosquito fish data.

Once again, for each data set, the heritability was recalculated and recorded as part of an empirical distribution of heritability. Just as with the basic bootstrap,

confidence intervals were calculated using the percentile method.

3.3 Programming

There is a complication in the mixed models software. Different programs take dramatically different amounts of processing time. For example, SAS proc MIXED took anywhere from 15 to 20 minutes to estimate the genotypic and phenotypic variance-covariance matrices for three traits, but VCE-5 only took one minute to obtain similar results. Due to these complications, this study utilized the VCE-5 (Covač and Groeneveld 2003) program to estimate the relevant covariance matrices. The randomization, bootstrapping, and postprocessing to obtain heritability estimates were performed in *R* (R Development Core Team 2008).

For the different pieces of this project the sampling procedures were developed in *R* (R Development Core Team 2008). Each section contains a looping structure where for each of 1000 iterations, the sampling procedure is performed on the original data. Each sample was then written to a file that could be read by the executable VCE-5. This executable was called from *R* using the system command. After each VCE-5 (Covač and Groeneveld 2003) analysis, the resulting genotypic and error variance-covariance matrices was then read into *R* for postprocessing. The first measure of heritability for each iteration was found by calculating the eigenvalues of the matrix $\mathbf{GP^{-1}}$. Maximum heritability was found as the maximum eigenvalue. The second measure of heritability was found by calculating the trace of the $\mathbf{G}$ and $\mathbf{P}$ matrices, and then computing the ratio of the traces.

3.3.1 Randomization Test

In order to create the files necessary to perform the randomization test, the original data files were read into *R* (R Development Core Team 2008). The data files were separated into the file containing the principal components and diet, and a file

containing the pedigree for each fish. The samples were created by permuting the file of the principal components but leaving the pedigree file as it was. This dispersed the siblings into random families. This data was then written to a file to be used in VCE-5 (Covač and Groeneveld 2003). The input parameter file for VCE-5 was the same for each iteration. The executable was then called in *R* using the system command and the postprocessing was performed for each iteration, with the heritability estimates recorded each time.

3.3.2 Basic Bootstrap

The postprocessing and input parameter file for the basic bootstrap were similar to the randomization test. The creation of the input data files was different. After the data and pedigree files were read into *R*, the bootstrap sample was created by sampling with replacement the entire data vector and corresponding pedigree information for the sampled fish. In this manner, while siblings were separated, the sire and dam of each fish was preserved from the original data. For each iteration, 109 fish were sampled and written to the data file while their sires and dams were written to a pedigree file. Confidence intervals were obtained by finding the appropriate percentiles of the resultant list of heritabilities.

3.3.3 Multistage Bootstrap

Creating the input files for the multistage bootstrap was a bit more complicated. The goal of maintaining the original design structure required that the number of dams mated with each sire and the number of offspring from each dam be used as the resampling structure. These numbers were obtained from the original data by counting the number of unique dams to each sire and doing the same for the offspring of each dam. For efficiency this counting was done before the looping of data sampling began. For the data files, eight sires were sampled with replacement from the list of

sires. For each sampled dam, the number of dams sampled was determined by the number of the sampled sire. For instance, if the sire under consideration was the third sire to be sampled, then the number of his dams that were sampled was determined by the third number in the pre-loop list of dam numbers. A similar process was performed for the offspring of each dam. In this manner, the original design structure was preserved and a sample of 109 fish for each iteration was assured. The data for each fish were then written to the data file and the pedigree information written to the pedigree file. The postprocessing was then performed in the same manner as the permutation test. Confidence intervals were obtained in a manner similar to that employed in the basic bootstrap.

3.4 Multivariate Heritability Measures

Both the eigenvalue-based and trace-based heritability were calculated for each method. The trace-based method was relatively straightforward; however, the dominant eigenvalue method was not as clear. Following Myers et al. (2006) and personal communication with Dean Adams, a coauthor on the paper, this measure was calculated using the singular-value decomposition of $\mathbf{GP^{-1}}$. The square of the largest singular value was divided by the sum of the squares of all the singular values.

4. RESULTS

The benefits of the nonparametric methods far outweigh the potential problems. One great advantage is the lack of assumptions for nonparametric techniques. While care must be taken in the sampling procedure to ensure proper inference, the lack of distributional assumptions makes the bootstrap a very useful tool, even for very complex design structures. Another great benefit is the relative ease of implementation of the various sampling methods. Resampling and bootstrapping is very simple with current statistical software. All inferences are made on an observable empirical distribution of relevant estimates. This is especially beneficial on smaller samples where the asymptotic properties of REML are not quite in effect. These methods are less sensitive to distributional violations as well, given that there are no assumptions to violate.

The ease of and intuition behind nonparametric methods makes them especially useful as possible solutions to the problems with heritability estimation; however, there are some drawbacks to these methods. When sample sizes are large and family sizes balanced, regular parametric methods may be more powerful in detecting statistical significance. Also, nonparametric methods suffer from the curse of dimensionality, in that the more principal components considered simultaneously, the longer the time and the greater the computational resources needed to obtain estimates. However, the waiting time and computational intensity for these methods can actually be much less than for the traditional mixed models approach.

4.1 Heritability from the Original Data

Heritability was calculated for the original data so that the bootstrap confidence intervals and randomization tests had a basis for comparison. The eigenvalue-based method yielded an estimate of heritability of 0.6293 and the trace-based method

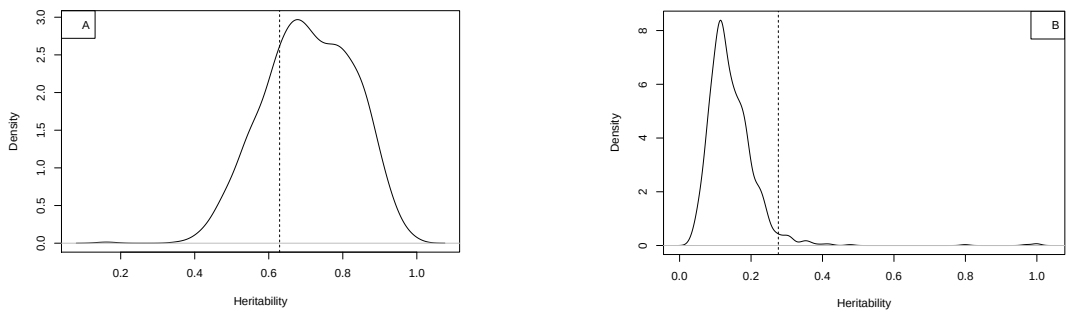


Figure 4.1: Reference distributions of eigenvalue-based (A) and trace-based (B) heritability estimates for randomization test. The vertical lines indicate the heritability estimate from the original data.

yielded an estimate of 0.2764. These estimates did not agree with each other, but the trace-based method ignores covariation while the eigenvalue method does not. A similar discrepancy was obtained by Myers et al. (2006).

4.2 Randomization Test

The randomization test was used to assess the significance of the heritability estimate. The resulting reference distributions illustrate the range of heritability estimates when the family group has no effect on the estimate of heritability. Figure 4.1 shows the reference distributions for the randomization test for both methods. From the graph of the estimates of heritability based on the eigenvalues of $\mathbf{GP}^{-1}$, the observed value is not statistically significant. The graph of the trace-based heritabilities indicates that the observed value is statistically significant. The randomization tests yielded p-values of 0.734 and 0.031 for the eigenvalue method and trace method, respectively. The disparity in conclusions as well as the difference in the estimates themselves indicates that the two methods are not equivalent.

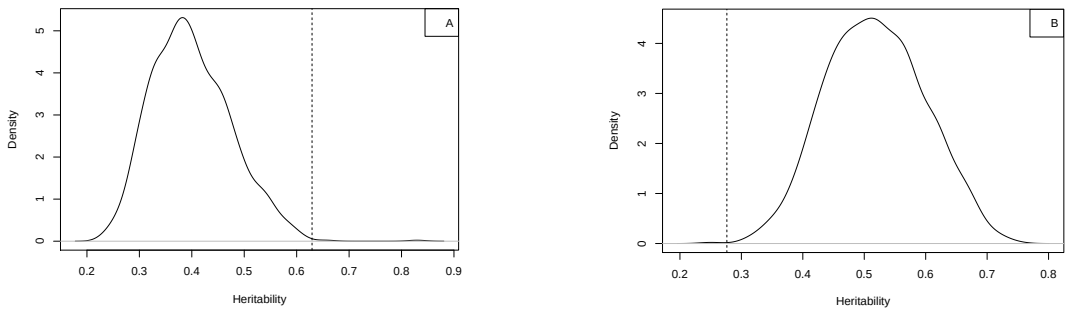


Figure 4.2: Reference distributions of eigenvalue-based (A) and trace-based (B) heritability estimates for basic bootstrap. The vertical lines indicate the heritability estimate from the original data.

4.3 Basic Bootstrap

As can be seen in Figure 4.2, the bootstrap heritability estimates for both the eigenvalue and trace methods occurred in the tails of the bootstrap sampling distributions. Another interesting observation of the bootstrap distributions is the fact that the two observed estimates occurred in opposite ends of their respective sampling distributions. This fact is reflected in their bootstrap confidence intervals.

Table 4.1: Confidence intervals for heritability using basic bootstrap

Method	Lower Confidence Limit	Upper Confidence Limit
Eigenvalue	0.2756453	0.5574426
Trace	0.3719606	0.6742037

The 95% confidence intervals, found using percentiles, are shown in Table 4.1. Surprisingly, neither confidence interval included the observed value. For a discussion on this point see Chapter 5. Although bias-corrected intervals were considered, the noninclusion of the point estimate in either interval renders any bias correction unimportant. Also notice that the distribution for the eigenvalue method was not symmetric. The trace distribution was relatively symmetric, though it was not centered around the observed value.

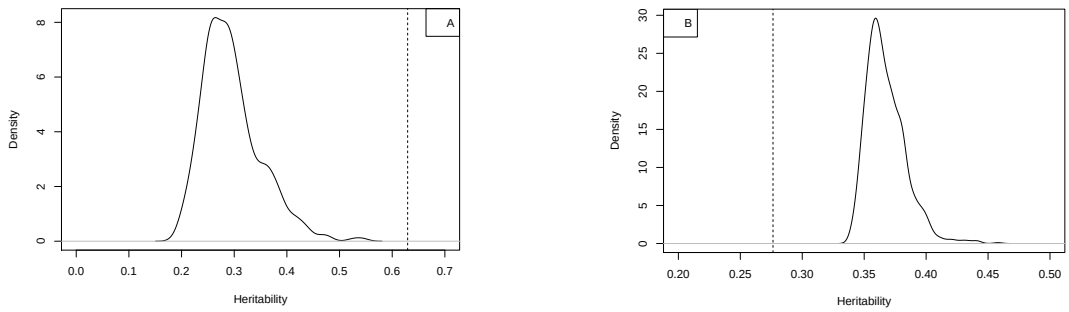


Figure 4.3: Reference distributions of eigenvalue-based (A) and trace-based (B) heritability estimates for multistage bootstrap. The vertical lines indicate the heritability estimate from the original data.

4.4 Multistage Bootstrap

Confidence intervals were also found for both estimation methods using the multistage bootstrap distributions. Figure 4.3 shows the observed distributions of the multistage bootstrap. Similar behavior as in the basic bootstrap was shown here. The observed estimates were not contained in the confidence intervals. Their positions relative to the sampling distribution were even further into the tails. Also of note was the relative peakedness of the distribution as compared to the basic bootstrap distributions. This led to the narrower confidence intervals (Table 4.2). Also, both distributions were skewed right.

Table 4.2: Confidence intervals for heritability using multistage bootstrap

Method	Lower Confidence Limit	Upper Confidence Limit
Eigenvalue	0.2115204	0.4288450
Trace	0.3452012	0.4037008

The intervals again did not include the observed heritability value for either method. Bias-corrected intervals were again not computed as in the case of the basic bootstrap. The estimate for the trace method was closer to its bootstrap distribution, but was not included in the distribution. The difference between the eigenvalue

heritability and the sampling distribution was more extreme. The location of the distribution indicates that the true heritability value was closer to the range of .30 to .35.

5. DISCUSSION AND CONCLUSIONS

5.1 Discussion of Results

There was a significant lack of performance for both of the bootstrap methods. Not only were estimates far outside the confidence intervals and data range, but the different estimation methods behaved differently. This seems to be an indication that this application of the bootstrap is inadequate, or at the very least in need of modification. The distribution for all methods appeared to be shifted heavily away from the observed estimates. Christensen and Sain (2002) discuss a similar problem with eigenvalue-based methods. When using a multivariate block bootstrap, they noticed that there was an increasing bias in the smallest eigenvalues as block size increased. They noted these biases on the log scale, which indicates that the bias on the original scale of the eigenvalues could be severe. They note that this can be explained by the increasing amount of linear dependencies that exist among bootstrap replicates as the block sizes, or in the case of this study, family sizes, increase. In this study, the eigenvalues of the bootstrap samples may also have been very biased since biasing in either $\mathbf{G}$ or $\mathbf{P}$ can severely affect the eigenvalues of $\mathbf{GP}^{-1}$. Also, the eigenvalues calculated by different programs and different functions within programs are not necessarily the same. For example, the eigen function in R automatically scales the eigenvalues to be between zero and one. SAS and other programs do not, though the estimates from SAS can be made to match R. Thus, estimation methods that use either the maximum eigenvalue or the maximum scaled eigenvalues must be carefully monitored such that the desired values are obtained.

Methods based on the trace might also be affected by problems within the eigenvalues, since the trace of any square matrix is equal to the sum of the eigenvalues. Distributions of the bootstrap sampling methods were similar in location. Also, the

multistage bootstrap provided a much tighter sampling distribution, indicating that this bootstrap may have more value than the basic bootstrap.

There may be modifications to the bootstrap that maintain the data structure and also account for the bias: for example, a bootstrap sample that samples sires, or, equivalently, families. This way, the internal family structures are retained in the reference distribution. The sampling would not be much more complicated than the multistage bootstrap used in this study, but care must be taken when specifying the pedigree information. Sampled sires and their accompanying dams and offspring must be numbered differently when they are resampled. The VCE-5 program used in these analyses requires that families be unique; otherwise they are grouped together.

Another interesting feature discovered in the analysis of these relative warps concerned the two randomization tests. The randomization distribution of the eigenvalue heritability indicated that even in situations of low heritability, or where family groupings do not determine shape, there are possibilities of obtaining very large heritability values. This must be due to the scaling of the eigenvalues.

The behavior of the eigenvalue-based heritability varied widely as the number of relative warps was increased. The lack of an exact monotonic relationship between heritability and the number of relative warps indicates that this scaled eigenvalue may not be the best for characterizing multivariate shape heritability. The eigenvalue-based heritability decreased as the number of relative warps was increased, a pattern contrary to what was believed should happen. This may also have been due to the scaling of the eigenvalues. The scaling of the eigenvalues obtained from the singular-value decomposition is not equal to other methods used in other programs or other functions native to *R*. The behavior of the trace-based heritability was much more regular, with a nearly monotonic increasing relationship between the estimate and the number of relative warps (5.1); however, trace method estimates were much lower than eigenvalue method estimates, and the implicit independence assumption among

Table 5.1: Point estimates for increasing number of relative warps

Number of Warps	Eigenvalue Method	Trace Method
2	0.9524	0.1864
4	0.7428	0.2050
6	0.7247	0.2586
8	0.5965	0.2880
10	0.6293	0.2764

relative warps is not entirely true.

Problems were encountered in the estimated genotypic and phenotypic variance-covariance matrices. These matrices were not always positive definite as they should have been. This had an adverse effect on the eigenvalues of $\mathbf{GP^{-1}}$ and, consequently, the estimates of heritability for both methods. Both the behavior of these estimation methods and the difficulties encountered in the covariance matrices indicate that the current methods for estimating multivariate heritability are not very reliable for relative warps. With nonpositive definite matrices, some of the eigenvalues were negative, causing the scaling factor, the sum of the eigenvalues, to decrease.

5.2 Conclusions and Future Work

The results of the study give a strong indication that the current method for characterizing multivariate heritability is possibly flawed, at least in the use of the eigenvalue measure of heritability. The analysis of shape heritability can be thought of as related to multivariate analysis of variance, or MANOVA. If we consider the genotypic covariance matrix as the within-family variation and the error covariance matrix as the across-family variation, other multivariate test statistics such as Wilks' lambda are possibly more effective in evaluating heritability based on the relative warps. Another benefit of statistics such as Wilks' lambda is that the distributions of the test statistics are known or very well approximated. Wilks' lambda can be expressed in terms of the eigenvalues of $\mathbf{G}$ or $\mathbf{P}$.

Great care must be taken in the choosing of methods for obtaining eigenvalues. Inferences and conclusions based on scaled eigenvalues may be dramatically different from those based on unscaled eigenvalues. These problems can be avoided by using functions in *R*, such as the *eigen* function. The eigenvalues of $\mathbf{GP^{-1}}$, which is not necessarily symmetric, can be obtained by finding the eigenvalues of the symmetric matrix $\mathbf{P^{-1/2}GP^{-1/2}}$.

Another difficulty encountered in this study was the estimation of the **G** and **P** matrices and their nonpositive definiteness. One solution to this problem is to try to use methods that approximate these matrices with the closest positive definite matrix. For a discussion of some of these methods see Amemiya (1985).

Many multivariate methods already exist that could possibly be adapted to testing the significance of the sire-dam structure of heritability studies. The current eigenvalue method is similar to Roy's largest root, a MANOVA test statistic based on the largest eigenvalue; however, Roy's test is more powerful than other methods only if the observation vectors are collinear, that is, not spread out in several dimensions (Rencher 2002). Wilks' lambda utilizes all of the eigenvalues of **G** and $\mathbf{P}^{-1}$ matrices. An adaptation of this statistic to heritability measurement might more easily accommodate the multidimensionality of shape data. This statistic has well known approximations to its distribution, and is more powerful in situations when the observations vectors are not collinear (Rencher 2002), as often occurs in heritability studies. There are also measures of association based on Wilks' lambda that might be used as measures of heritability (Rencher 2002).

The bootstrap methods used in this study failed to accurately reflect the maximum heritability. For both heritability measures, heritability was higher for resampled data than for the original data. The resampling with replacement that is performed by bootstrapping has the effect of lessening the variation within families, an effect that can only increase heritability. This inflation of heritability is an artificial

effect caused by the basic bootstrap methods.

Determining the strength of the relationship between parents and offspring is an essential part of the study of genetics. In determining the heritabilities of shape and other traits, efforts can be made toward identifying and assisting endangered species. The overall heritability of shape is an inherently multivariate concept that requires complex multivariate methods to approximate, analyze, and assess. While methods exist, current methods contain drawbacks ranging from lack of estimates of error to instability in the estimates themselves. This study presented an approach to the standard error estimation problem through the use of nonparametric methods. Further adaptations and modifications of these methods can yield better, less biased methods for determining not only the significance of heritability in a given species but also a confidence interval on the magnitude of heritability.

BIBLIOGRAPHY

Amemiya, Y. (1985), "What should be done when an estimated between-group covariance matrix is not nonnegative definite?" *The American Statistician*, 39, 112–117.

Christensen, W. F. and Sain, S. R. (2002), "Accounting for Dependence in a Flexible Multivariate Receptor Model," *Technometrics*, 44, 328–337.

Covač, M. and Groeneveld, E. (2003), *VCE-5*.

Higgins, J. J. (2004), *Introduction to Modern Nonparametric Statistics*, Thomson Learning, Inc.

Klingenberg, C. P. (2003), "Quantitative Genetics of Geometric Shape: Heritability and the Pitfalls of the Univariate Approach," *Evolution*, 57, 191–195.

Klingenberg, C. P. and Leamy, L. J. (2001), "Quantitative Genetics of Geometric Shape in the Mouse Mandible," *Evolution*, 55, 2342–2352.

Klingenberg, C. P. and Monteiro, L. R. (2005), "Distances and Directions in Multidimensional Shape Spaces: Implications for Morphometric Applications," *Systematic Biology*, 54, 678–688.

Lande, R. (1979), "Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain:Body Size Allometry," *Evolution*, 33, 402–416.

Lande, R. and Arnold, S. J. (1983), "The Measurement of Selection on Correlated Characters," *Evolution*, 37, 1210–1226.

Meyer, K. (2008), *WOMBAT: A Program for Mixed Model Analyses by Restricted Maximum Likelihood*.

Myers, E. M., Janzen, F. J., Adams, D. C., and Tucker, J. K. (2006), "Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain:Body Size Allometry," *Evolution*, 60, 563–572.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rencher, A. C. (2002), *Methods of Multivariate Analysis, (2nd ed.)*, John Wiley & Sons, Inc., 2nd ed.

A. *R* FOR DIFFERENT RESAMPLING PROCEDURES

A.1 *R* code for Randomization Test

```
### read in data and pedigree files
proj.dat<-read.table("C:/VCEstuff/examples/test/data/fishdat.dat",header=T)
proj.ped<-read.table("C:/VCEstuff/examples/test/data/fishped.ped",header=T)

### load library with write.matrix command
library(MASS)

### initialize output vectors
tr<-NULL
h2_random<-NULL

### truncate pedigree to only offspring variable vectors
file.ped<-proj.ped[1:109,]

### begin sampling loop
for (k in 1:1000) {

### create sampling index for randomization iteration (without
### replacement)
ranseq<-sample(1:109,109,replace=F)

### request data corresponding to sampling index
datai<-proj.dat[ranseq,]

### create variables necessary for input files of VCE-5
animal<-TIER<-c(1:109)
DUM<-datai[,32]

### create data file for VCE
vcedat<-cbind(datai[,1:30],TIER,DUM)

### write data file to input text file
writedata<-format(vcedat,digits=10,width=14)
write.matrix(writedata,"C:/VCEstuff/examples/test/data/Random/fishran.dat",sep="")
```

```
### call VCE-5 based on above created input file and input parameter
### file pp10
system(paste('"C:/VCEstuff/examples/test/temp/Random/vce.exe"',
    '"C:/VCEstuff/examples/test/master_pfiles/Random/pp10.txt"',sep="
    "),show.output.on.console=F,wait=T)

### read VCE output to obtain G and E matrices
a<-read.table('C:/VCEstuff/examples/test/temp/Random/cov.txt',skip=3,nrow=10)
geno<-as.matrix(a)

a<-read.table('C:/VCEstuff/examples/test/temp/Random/cov.txt',skip=16,nrow=10)
error<-as.matrix(a)

### calculate phenotypic covariance matrix
pheno<-geno+error

### calculated eigenvalue based heritability using singular value
### decomposition
h2<-geno%*%solve(pheno)
b<-svd(h2)
eigvals<-b$d*b$d
maxeig<-max(eigvals)
h2_random[k]<-maxeig/sum(eigvals)

### calculate trace based heritability
tr[k]<-sum(diag(geno))/sum(diag(pheno))
}
```

A.2 *R* code for Basic Bootstrap

```
### read in data and pedigree files
proj.dat<-read.table("C:/VCEstuff/examples/test/data/fishdat.dat",header=T)
proj.ped<-read.table("C:/VCEstuff/examples/test/data/fishped.ped",header=T)

### load library with write.matrix command
library(MASS)

### initialize output vectors
h2_boot<-NULL
tr_boot<-NULL

### truncate pedigree file to offspring variable vectors only
file.ped<-proj.ped[1:109,]

### begin bootstrap sampling loop
for (k in 1:1000) {

### create sampling index
ranseq<-sample(1:109,109,replace=T)

### request data and pedigree information corresponding to sampling
### index
datai<-proj.dat[ranseq,]
pedi<-file.ped[ranseq,]

### create variables necessary for VCE-5 input files
animal<-TIER<-c(1:109)
DUM<-datai[,32]

### create data and pedigree files for VCE-5
vcedat<-cbind(datai[,1:30],TIER,DUM)
vceped<-rbind(cbind(animal,pedi[,2:5]),proj.ped[110:138,])

###write files to input text files for VCE-5
writedata<-format(vcedat,digits=10,width=14)
write.matrix(writedata,"C:/VCEstuff/examples/test/data/Boot/fishboot.dat",sep="")

writeped<-format(vceped,digits=3,width=10)
write.matrix(writeped,"C:/VCEstuff/examples/test/data/Boot/pedboot.ped",sep="")
```

```
### call VCE based on input parameter file pp02
system(paste('"C:/VCEstuff/examples/test/temp/Boot/vce.exe"',
    '"C:/VCEstuff/examples/test/master_pfiles/Boot/pp02.txt"',sep="
    "),show.output.on.console=F,wait=T)

### read VCE output of the G and E matrices
a<-read.table('C:/VCEstuff/examples/test/temp/Boot/cov.txt',skip=3,nrow=10)
geno<-as.matrix(a)

a<-read.table('C:/VCEstuff/examples/test/temp/Boot/cov.txt',skip=16,nrow=10)
error<-as.matrix(a)

### calculate phenotypic covariance matrix
pheno<-geno+error

### calculate eigenvalue based heritability using singular value
### decomposition
h2<-geno%*%solve(pheno)
b<-svd(h2)
eigvals<-b$d*b$d
maxeig<-max(eigvals)
h2_boot[k]<-maxeig/sum(eigvals)

### calculate trace based heritability
tr_boot[k]<-sum(diag(geno))/sum(diag(pheno)) }
```

A.3 *R* code for Multistage Bootstrap

```
### read in data and pedigree files
proj.dat<-read.table("C:/VCEstuff/examples/test/data/fishdat.dat",header=T)
proj.ped<-read.table("C:/VCEstuff/examples/test/data/fishped.ped",header=F)
names(proj.ped)<-c("animal","dam","sire","birth_order","type")

### load library with write.matrix function
library(MASS)

### create vector of unique sires
sire<-unique(proj.ped$sire[1:109])

### create vector of the number of unique dams per sire
dam.counts<-NULL
for(i in 1:length(sire)){
    dam.sire<-unique(proj.ped$dam[proj.ped$sire==sire[i]])
    dam.counts[i]<-length(dam.sire)
    }

### create vector of the number of offspring per dam
dams<-unique(proj.ped$dam[1:109])
ani.counts<-NULL
for(i in 1:length(dams)){
    ani.counts[i]<-sum(proj.ped$dam[1:109]==dams[i])
    }

### initialize output vectors
h2_multiboot<-NULL
tr_multiboot<-NULL

### begin bootstrap loop
for (k in 1:1000){

### sample from sires required number
ss<-sample(sire,length(sire),replace=T)

### sample dams of each sire based on number from dam.counts
j<-1
ds<-NULL
parent<-NULL
```

```
for (i in 1:length(ss)){
    dam<-unique(proj.ped$dam[proj.ped$sire==ss[i]])
    ds[j:(j+dam.counts[i]-1)]<-sample(dam,dam.counts[i],replace=T)
    parent<-cbind(parent,rbind(ds[j:(j+dam.counts[i]-1)],ss[i]))
    j<-j+dam.counts[i]
    }

### create dam sire pair data set
parent<-as.data.frame(t(parent))
names(parent)<-c("dam","sire")

### sample offspring from each sampled dam based on numbers of
### ani.counts
fam<-NULL
j<-1
for(i in 1:length(parent$dam)){
    anis<-proj.ped$animal[proj.ped$dam==parent$dam[i]]
    animal[j:(j+ani.counts[i]-1)]<-sample(anis,ani.counts[i],replace=T)
    fam<-cbind(fam,rbind(animal[j:(j+ani.counts[i]-1)],parent$sire[i],parent$dam[i]))
    j<-j+ani.counts[i]
    }

ped<-as.data.frame(t(fam)) names(ped)<-c("animal","sire","dam")

### request offspring data based on animal index
datai<-proj.dat[animal,]
animal<-TIER<-c(1:109)
DUM<-datai[,32]

### create data and pedigree files for VCE processing
vcedat<-cbind(datai[,1:30],TIER,DUM)
vceped<-rbind(cbind(animal,ped[,2:3],proj.ped[1:109,4:5]),proj.ped[110:138,])

### write data and pedigree files to input text files for VCE
### processing
writedata<-format(vcedat,digits=10,width=14)
write.matrix(writedata,"C:/VCEstuff/examples/test/data/Multiboot/fishboot.dat",sep="")

writeped<-format(vceped,digits=3,width=10)
write.matrix(writeped,"C:/VCEstuff/examples/test/data/Multiboot/pedboot.ped",sep="")
```

```
### call VCE based on input parameter file pp03
system(paste('"C:/VCEstuff/examples/test/temp/Multiboot/vce.exe"',
    '"C:/VCEstuff/examples/test/master_pfiles/Multiboot/pp03.txt"',sep="
    "),show.output.on.console=F,wait=T)

### read VCE output for G and E matrices
a<-read.table('C:/VCEstuff/examples/test/temp/Multiboot/cov.txt',skip=3,nrow=10)
geno<-as.matrix(a)

a<-read.table('C:/VCEstuff/examples/test/temp/Multiboot/cov.txt',skip=16,nrow=10)
error<-as.matrix(a)

### calculate phenotypic covariance matrix
pheno<-geno+error

### calculate eigenvalue based heritability using singular value
### decomposition
h2<-geno%*%solve(pheno)
b<-svd(h2)
eigvals<-b$d*b$d
maxeig<-max(eigvals)
h2_multiboot[k]<-maxeig/sum(eigvals)

### calculate trace based heritability
tr_multiboot[k]<-sum(diag(geno))/sum(diag(pheno)) }
```

```
comment  Fish heritability
   pp10      nt=1   same   no             109            138              6+2              no
            |      |      |              |              |                |                |           |
   pfile    |nt    |model |missing       |data file     |ped file        |fixed           |random     |

   ntrait:               1         same model:          YES
   missing data?:       NO         data file:           109
   animal?:            YES         ped:                 138
   fixed effects:      6:2         random effects:       NO
   maternal effects?:   NO

system

DATA
   datfile='C:/VCEstuff/examples/test/data/Random/fishran.dat' format='(28f14.10,10f14.0)'
   dep=w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 w11 w12 w13 w14 w15 w16 w17 w18 w19 w20 w21 w22 w23 w24 u1 u2 u3 u4
   indep= blk ntrt tier;
   pedfile='C:/VCEstuff/examples/test/data/Random/pedran.ped' format='(5i10)' link=tier;
covariance
  tier;
model
   w1 w2 w3 w4 w5 w6 w7 w8 w9 w10 = ntrt + tier;
output
   covfile = 'C:/VCEstuff/examples/test/temp/Random/cov.txt' form='full';
end
```