Comment

Paul D. Sampson

Dr. Bookstein has been working since the late 1970's to reorient the field of morphometrics to better recognize and exploit the fundamental geometric structure of morphometric data. I am very pleased that the methodology he espouses is being presented here to the audience of a major statistical journal for assessment and critique, and I would therefore like to thank Dr. Bookstein and the editors of this journal for presenting this interesting paper. I hope that this paper and the questions it raises will spur more statisticians to research in the field of morphometrics, which was featured in 1984 at sessions of the Western Regional Meeting of the Biometric Society in Logan, Utah, and the XIIth International Biometric Conference in Tokyo. To lend further perspective on this work within the field of morphometrics I would like to recommend Bookstein's own review paper (Bookstein, 1982a) and the recently published second edition of *Multivariate Morphometrics* by Reyment et al. (1984; see especially Chapter 12 on The Analysis of Size and Shape).

I hope that Dr. Bookstein will not be too disappointed in my admitting that I have heard of people who have found some of his writings on the subject of morphometrics difficult to follow. I would like to reassure any readers who may have skipped to this discussion before finishing the paper that this work provides, in my opinion, a very clear exposition of Bookstein's approach to the analysis of morphometric data. Some parts, such as the sections on deformation and factors, may be difficult at first reading because of their novelty, but they are well worth rereading. The connection between the deformational description of shape changes/differences and the selection of size and shape variables (Sections 1 and 4) is the key theme of this work and is, I believe, very important.

In my discussion I will first comment and elaborate on the interpretation of the size and shape variables Bookstein proposes in Sections 2–5. I will then raise a number of practical issues (including assumptions, nonparametric tests, simultaneous inference, and statistical computing) faced in our own recent experience applying Bookstein's methodology to a study at the University of Washington of minor facial dysmorphology associated with the fetal alcohol syndrome (or, more accurately, "probable fetal alcohol effects").

Paul D. Sampson is Assistant Professor, Department of Statistics, GN-22, University of Washington, Seattle, Washington 98195.

INTERPRETING THE SIZE VARIABLE S AND "OPTIMAL" SHAPE VARIABLES

As a prelude I would like to remark briefly on Bookstein's mathematical notation. Bookstein provides rigorous support for the definition of a space of size variables as the space spanned (redundantly) by the set of $K(K-1)/2$ distances among K landmarks, and for the definition of a space of shape variables as the space spanned by the complex affine shape ratios for any $K-2$ triangles used to triangulate the K landmarks. However, the differential notation used in this exposition may be unfamiliar to some statisticans (as it was to this statistician) who more commonly represent random displacements in multivariate data by ε_i's rather than differentials dz_i in the complex plane. The "first order" operations are perhaps more often expressed as "delta method" or "propagation of error" approximations based on first order Taylor series expansions. It may be helpful for some readers to recognize that, for example, Lemma 1 can easily be verified by the delta method applied to Z_i's considered as real vectors rather than complex numbers. The corresponding two-dimensional diagram, expressed in terms of either the complex plane or the real Euclidean plane, is also helpful.

Bookstein makes a convincing argument for the central role that the size variable S should play in studies of allometry and trends in size. It will be instructive to relate S explicitly to the unique (log) size variable $\sum a_i V_i$ orthogonal to variation in shape which he also cites. However, we should note first that stochastic independence of this size variable and shape, referred to in connection with Mosimann (1970), holds only when the V_i follow a multivariate normal distribution, i.e., when the original size variables are log normally distributed (cf. Sampson and Siegel, 1984, 1985).

It is important to recognize that the null hypothesis motivating the use of S is a statement of no pattern in shape variation at all. One may observe systematic variation in shape, reflected in noncircular scatters for the shape coordinates $Q(Z_1, Z_2, Z_3)$, so that the null hypothesis is not true, and yet find that S is still uncorrelated with shape. Acknowledgment of this is suggested in the discussion of the example at the end of Section 3. Other interpretable size measures are necessarily correlated with shape, and therefore we must be careful about general claims of isometry without qualification. On the other hand, Bookstein notes

in his concluding remarks, point (2), that observed size shape covariances may not in fact represent "allometry." It would seem therefore that issues of allometry/isometry will remain difficult to establish unambiguously even with such geometric multivariate data.

We can shed some further light on these issues by returning to the log size variable $\sum a_j V_i$. Bookstein et al. (1985, Section 2.2) provide a number of justifications for the common practice of analyzing log transformed distances. These same arguments lead them to use "length elements" (which is consistent with the current work's focus on differentials) instead of long measured distances wherever possible. If we shift our attention to logs of the interlandmark distances we find, using the same notation and first order approximations, that

$$d \log |Z_i - Z_j| \approx d|Z_i - Z_j|/|W_i - W_j|,$$

and so a weighted average of log distances, or the log of a weighted geometric mean,

$$S^* = k \sum_{i<j} |W_i - W_j|^2 \log |Z_i - Z_j|,$$

yields, for small variations of the Z_i about the means W_i, the differential

$$dS^* \approx k \sum_{i<j} |W_i - W_j| \cdot d|Z_i - Z_j|,$$

i.e., S^* is also statistically equivalent to Bookstein's size variable S. Therefore, under the null model of independent circular normal scatters about unobserved landmark means, S^* is also approximately uncorrelated with the space of shape variables. The first order equivalence of S, $S^{1/2}$, and now S^* suggests that a more detailed investigation of the adequacy of first order approximations is necessary if we are to interpret and choose among these size variables in practice.

It is instructive to demonstrate the orthogonality of S^* and "shape" by a direct argument (which Bookstein notes as unnecessary in view of his Theorem 1). We take the uncorrelated circular disturbances dz_i at the landmarks and propagate these into approximate covariances among the $d \log |Z_i - Z_j|$. (One could also assume that these log distances follow a multivariate normal distribution.) Then, using the results in Sampson and Siegel (1984, 1985) or Bookstein et al. (1985, Section 2.2), this covariance matrix specifies S^* as the unique size measure orthogonal to the space of shape variables defined by (logs of) ratios of distances. Since the vector of weights $\{|W_i - W_j|^2, i < j\}$ is simply derived from the vector of row sums of the inverse of the covariance matrix of $\{\log |Z_i - Z_j|, i < j\}$, we can readily specify other covariance matrices with the same row sums, and hence other models for the dispersion about the landmark means, for which S^* (or S) will still be approximately uncorrelated with the space of shape variables. Further understanding of the observed pattern of shape variation must then be sought. See Bookstein's discussion of size allometry in Section 4.

In Section 5 Bookstein defines the "optimal" shape variable for a group contrast, where optimal is explicitly defined to mean the greatest *net* proportional change between the two groups. Although perhaps not often of practical importance, we can evaluate this optimal shape variable case by case using the construction indicated in Figures 10 and 11. It is worth noting that, in this context of multivariate two (or many) sample problems, we also have the usual "statistically optimal" shape variable: that derived from Fisher's linear discriminant function. Fisher's linear discriminant function is the linear combination of the (shape) coordinates $Q(Z_1, Z_2, Z_3)$ given by the coefficient vector $\Sigma^{-1}(Q_1 - Q_2)$, where Σ denotes the (pooled) covariance matrix of the scatter of shape coordinates (as in Figure 11) and Q_1 and Q_2 represent the group mean vectors (Q and $Q + dQ$ in Figure 10). This linear combination specifies the projection of the Q coordinates on a gradient direction (the discriminant axis) which will be the same as dQ $(= Q_1 - Q_2)$ only in the case that Σ is a multiple of the identify, i.e., when the assumption of circular scatters at the landmarks is satisfied. By the construction of Figure 10 we can in general determine the simple ratio of orthogonally measured distances that most clearly describes this "Fisher's linear discriminant shape variable" expressed as a linear combination of the shape coordinates of Q.

The definition and geometric depiction of the Fisher's linear discriminant shape variable may reveal itself as relatively more important in the general case of multiple landmarks where correlations among the shape coordinates of different triangles are perhaps more likely to arise (in violation of the null assumption of no systematic variation in shape). Theorem 2 of Section 5 shows us how the size variables (distances) showing the maximum or minimum mean ratio of difference between two configurations can be easily determined. However, these distances are unlikely to be determined from orthogonal distances and we should probably not settle for their ratio as *the* optimal shape variable. Indeed, trying to reduce a complex deformation relating two configurations of landmarks to one or two simple size or shape variables is generally inadvisable in view of the highly multivariate and nonlinear nature of the deformations that must be expected for configurations of many landmarks. (See Bookstein's discussion of "visualizing mean changes by biorthogonal grids.")

Rather than focus on a single optimal variable we can, using any given triangulation of K landmarks,

determine Fisher's linear discriminant function defined as a linear combination of the $2(K - 2)$ total shape coordinates. This combination may be broken down into its $(K - 2)$ bivariate components, each of which defines a linear combination and shape variable for one of the constituent triangles. These $(K - 2)$ shape variables may then be considered as jointly optimal for disciminating between the groups. Of course, all $(K - 2)$ of the shape variables may not contribute significantly to the discrimination, and the specific results obtained will be expressed in terms of the given triangulation. However, the choice of a triangulation will not affect the statistical discrimination (up to the usual first order approximations). One might consider extending the question of optimal size variables to one of optimal triangulations for displaying group differences, at least in so far as it relates to the problem Bookstein poses in Section 7 on the selection of interesting landmarks.

Finally, for multiple population discrimination problems we can define sets of shape variables in the same manner using multiple canonical axes (rather than Fisher's single discriminant axis) to provide geometric depictions of the primary dimensions of shape differences.

SOME PRACTICAL ISSUES

Researchers in fisheries, zoology, and medical sciences at the University of Washington have begun applying Bookstein's methodology for analyzing landmark data. Here I will raise some practical issues that we have addressed in a study of facial photographs used to diagnose children at 7 years of age for fetal alcohol effects. This particular study is one part of an ongoing longitudinal study of the effects of maternal alcohol consumption during pregnancy on the health of the offspring. Background can be found in Streissguth et al. (1981). A complete report on the current analysis is in preparation.

For this study we digitized 23 landmarks judged to be homologous and identifiable on standard full face and lateral photographs of 21 seven-year-old children gestationally exposed to "known," high quantities of ethanol (at least 2 ounces of average absolute alcohol consumption per day during early pregnancy by mother's self-report; 10 children were exposed to at least 4 ounces of average absolute alcohol per day), and 21 children with negligible ethanol exposure. These "control" children were group-matched with the alcohol-exposed children on the basis of marginal distributions of race, sex, maternal education, nicotine use, and certain other drugs used. Figure 1 provides a sketch of the locations of landmarks on the lateral photograph. In comparison with the cephalometric data analyzed by Bookstein, the relative locations of these

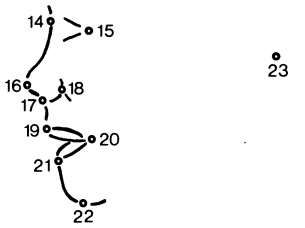

FIG. 1. *Locations of* 10 *landmarks located on lateral photographs.*

landmarks are generally less well defined and they may be disturbed by variations in head position and facial expression. A small number of photographs were missing or judged unusable because of head position. For purposes of this discussion I will cite only a few examples from analyses using the lateral photographs to contrast the 8 most heavily exposed children with 28 "control" and "moderately" exposed children. (Children exposed to 2 to 3 ounces of alcohol per day could not be distinguished from the control group.)

Little prior experience (with children of age 7) was available to suggest facial shape features that might distinguish the alcohol-exposed from the control group; therefore, our exploratory analysis involved a nearly exhaustive search for triangles which differed significantly in shape between the two groups. Building on these analyses, triangles were combined to assess the significance of shape differences in quadralaterals, and in a five-sided polygon. Specific issues we explored were: the adequacy of first order approximations under which results are invariant with respect to choice of baseline for triangle analyses (the effects of first order approximations on allometry studies and the choice of a size variable, S, $S^{1/2}$, or S^*, has not been explored as we have not yet computed absolute scales for the photographs); the assumption of normally distributed scatters and a nonparametric alternative to Hotelling's T^2 test; and concerns about simultaneous inference in the face of our analyses of numerous triangles.

To assess the effects of choice of baseline we need more information than is provided by Figure 6 of Bookstein. As an example providing some experience with these effects, Figure 2 shows plots of the shape coordinates $Q(Z_{17}, Z_{23}, Z_{19})$, $Q(Z_{17}, Z_{19}, Z_{23})$, and $Q(Z_{19}, Z_{23}, Z_{17})$, for the triangle with vertices at the top of the philtrum (17), the upper lip (19), and the center of the ear (23). Although the pattern in these scatters is somewhat irregular, there is some (weak) suggestion of noncircularity (i.e., systematic variation in shape) within both groups. Corresponding Hotelling T^2 values and p values computed assuming the scatters represent bivariate normal samples with a common covariance structure are also given. For this triangle the distance between points 17 and 19 (the length of the philtrum) is not very large relative to the scale of

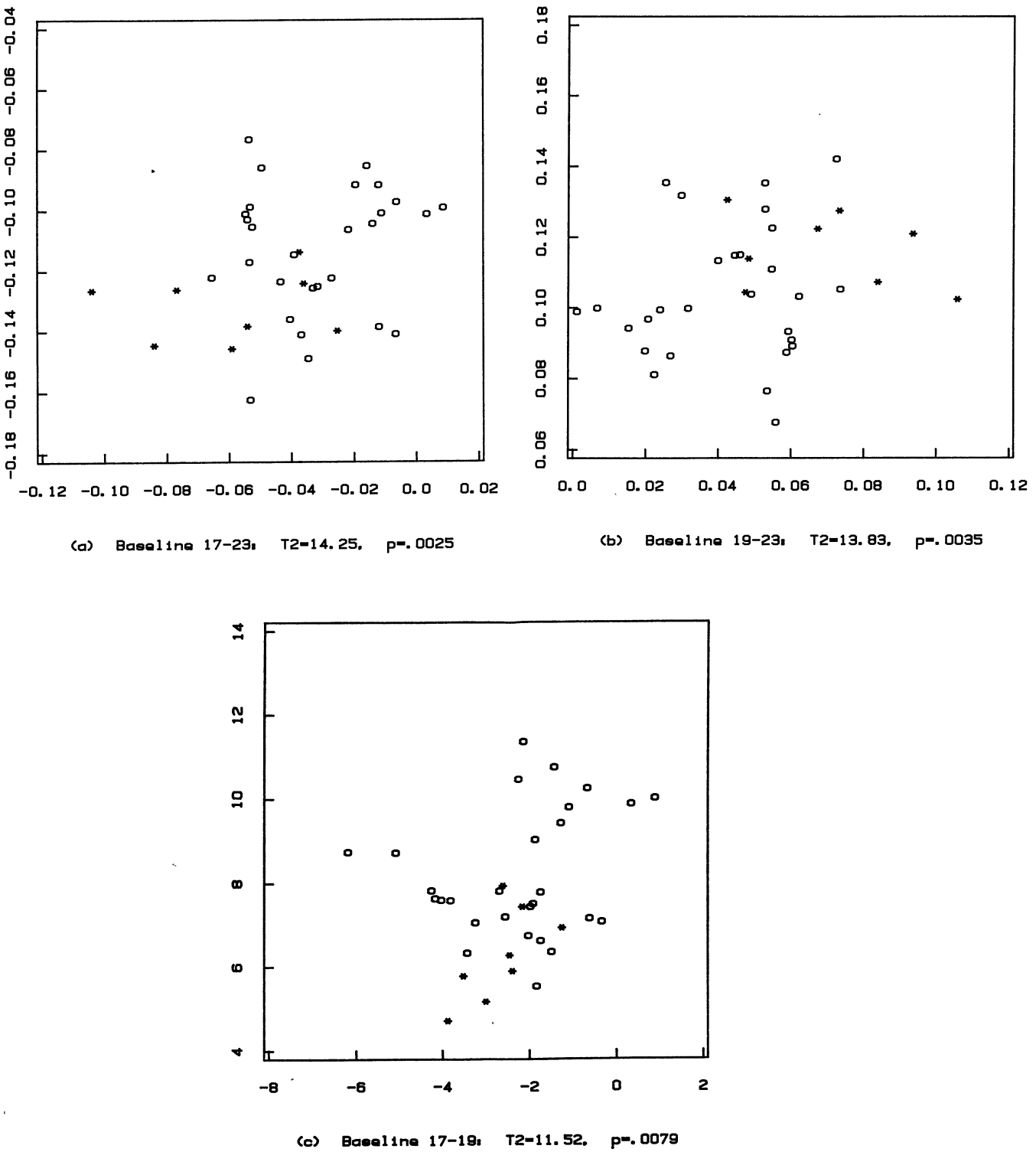

FIG. 2. *Triangle* 17-19-23 *shape coordinates derived from different baseline choices (points set to coordinates* (0, 0) *and* (1, 0)) *with Hotelling's two sample* T^2 *statistics and p values. High alcohol exposure cases are marked with a "*"; "moderate" and low alcohol cases are marked with a "0."*

the scatter. Nonetheless, in this particular case our inferences are not highly dependent on choice of baseline, although the p value increases by a factor of two to three when the short side 17–19 is made the baseline. For the slightly larger triangle 17-23-14 our results are much more nearly invariant under change of baseline; T^2 values are 5.66, 5.64, and 5.49.

As a check on the normal theory p values computed using Hotelling's T^2 we carried out a number of permutation tests. For comparisons of 8 heavily exposed

children with 28 "moderate" and "control" children we repeatedly assigned 8 cases at random to a group labeled "exposed," the remaining 28 cases to the "control" group, and computed the corresponding T^2 value. For example, for the observed sample groups we obtain $T^2 = 13.0$ (with $7 + 27 = 34$ d.f. for error) and associated p value of .0047 when analyzing the triangle with vertices 19, 23, and 14. Computation of 400 permutation samples yielded only two T^2 statistics more extreme than 13.0, and thus a permutation p value of $3/401 = .0075$ which is quite consistent with the normal theory value.

Our interest in permutation tests was raised as much by an unexpected aspect of our experimental design as it was by normal theory assumptions. Although the "exposed" and "control" groups of size 21 were balanced by race and sex, we found ourselves focussing on only the 10 most highly exposed cases. Of these cases, 7 were nonwhite and 5 were nonwhite females. Our exposed-control comparisons were therefore clearly confounded with race and sex. Small sample sizes prohibited individual modeling of race and sex effects, so to obtain more interpretable significance levels for the comparison of interest we computed random permutation groupings subject to the constraints that the race, or race and sex, proportions matched those in the observed samples. For example, conditioning on the observed race proportions we computed 250 permutations and a p value of $8/251 = .032$ for triangle 19-23-14; conditioning on both the observed race and sex proportions in the exposed group we obtained a p value of $13/201 = .065$ from 200 permutation samples.

On the basis of these and other tests we conclude that our data is reasonably consistent with Bookstein's normal model. We also generally confirm apparently significant differences after adjusting for confounding effects via permutation tests.

Our search for significant shape differences in triangles selected from among 23 landmarks defined in frontal and lateral views raises obvious questions about simultaneous inference. The methodology Bookstein introduces opens up the systematic geometric and statistical analysis of large sets of landmarks. In Bookstein's craniofacial example attention is focussed (on a priori grounds) on a small number of landmarks, and primarily one specific quadrilateral. In our application a larger number of landmarks were eligible for consideration and prior information was inadequate to restrict attention to a small number of triangles of quadrilaterals. We now face the prospect of analyzing much richer (higher dimensional) data sets, but with corresponding concerns for simultaneous inference in the course of investigation. For descriptive purposes the extra information contained in large data sets is quite valuable, especially as it is used in computing biorthogonal grids. In contrast with other types of highly multivariate data, summaries of morphometric multivariate analyses are guaranteed to have clear two-dimensional representations. Bookstein notes that most models for deformation analysis have too few parameters, but with the current treatment of landmark data we will often increase the dimensionality of the parameter space to a point that challenges us to interpret tests of group contrasts. The research problem that Bookstein raises about the selection of landmarks (or triangles) contributing significantly to a group difference (Section 7) is based in this simultaneous inference problem.

Formally, any particular triangulation of K landmarks provides $2(K-2)$ shape variables to be considered, and an overall T^2 test can be carried out on $p = 2(K-2)$ and $n = (N_1 + N_2 - 2)$ degrees of freedom for samples of sizes N_1 and N_2. Percentage points of the T^2 distribution are given by multiples of F percentiles, $T^2_{p,n} = [pn/(n - p + 1)]F_{p,(n-p+1)}$, and simultaneous inferences for all triangle comparisons could be carried out by referring individual bivariate $T^2_{2,n}$ statistics to $T^2_{p,n}$ percentiles in accord with the union-intersection test principle. However, for our application, we would have $p = 16$ for the 10 landmarks considered in the lateral photograph alone, and $p = 22$ for the 13 landmarks in the frontal! I suggest that the extreme conservativeness of formal simultaneous inference in this case suggests a logical flaw.

Even if certain triangular regions of a figure of interest are (statistically) stable in shape across a group comparison, it will generally make sense to draw a continuous, differentiable extension over the entire figure of interest of the effects of a significant difference in shape observed in some subregion. Biorthogonal grids provide this extension. Formal simultaneous inference may prohibit a conclusion of significant differences in some region of a figure, but the biological connectivity of triangular regions will often justify examining this extension. Indeed, this is perhaps a setting where knowledge of the proximity of triangular regions provides prior information that should not be ignored by formal simultaneous inferences that recognize estimated statistical covariances but not explicit "physical covariances" or connectivity. In this setting the biologist should not pay a penalty for measuring too many landmarks/variables in studying a shape change or deformation. The association that Bookstein draws between deformations and factor models leads me to believe that a proper resolution to this issue can be expressed in terms of Herman Wold's "soft modeling" approach to the analysis of factor models (cf. Bookstein, 1986).

My final remarks concern two aspects of statistical computing for landmark data. Analyses suggested by Bookstein involve only simple manipulations of

coordinate data and standard multivariate statistical analyses which are available in most major statistical packages. At the University of Washington we have found it especially convenient to carry out our analyses using an interactive statistical programming language such as "S" (Becker and Chambers, 1984), or "ISP" (Dunlap, 1985), both of which have facilities for user-defined special purpose macros. In this interactive macro environment we easily extract shape coordinates (using the simple expressions of complex arithmetic given by Bookstein) for arbitrary sets of landmarks, compute the usual statistical analyses, and generate various graphical displays of the results.

One of Bookstein's most important contributions to the field of morphometrics was the method of biorthogonal grids which he introduced in 1978. We are finding biorthogonal grids very useful for graphically synthesizing the findings from the discrete analyses of multiple triangles (as Bookstein describes in Section 6). However, to our knowledge no one but Bookstein himself at the University of Michigan has ever had software to generate a biorthogonal grid. This is probably due to the complexity of the algorithms originally described. We have recently implemented (with Bookstein), in the "S" environment, new and simpler algorithms for the computation of biorthogonal grids. The computed homology which maps and smoothly interpolates one set of landmarks onto another is derived from easily programmed "thin-plate" spline interpolators (Meinguet, 1979). This algorithm does not constrain the mapping to be linear on a specified boundary as does Bookstein's original algorithm. Our algorithm for drawing out the biorthogonal grids, the integral curves of the symmetric tensor field (Figure 15b), is based on a widely available differential equation solver. A report describing this new biorthogonal grid software and applications is in preparation.

ACKNOWLEDGMENTS

I would like to thank Dr. Sterling Clarren and Deborah Donnell, colleagues on the Pregnancy and Health Study at the University of Washington, for their contributions to the analysis cited above. This research was supported in part by Grant AA01455-10 from the National Institute of Alcoholism and Alcohol Abuse.

ADDITIONAL REFERENCES

BECKER, R. A. and CHAMBERS, J. M. (1984). *S: An Interactive Environment for Data Analysis and Graphics.* Wadsworth, Belmont, Calif.

DUNLAP, W. W. (1985). ISP User's Guide. Associated Students of the Univ. of Washington Lecture Notes, Seattle.

MEINGUET, J. (1979). Multivariate interpolation at arbitrary points made simple. *J. Appl. Math. Phys.* **30** 292–304.

SAMPSON, P. D. and SIEGEL, A. F. (1984). The measure of 'size' independent of 'shape' for multivariate lognormal populations: definition and applications. *Proceedings of the XIIth International Biometric Conference, Tokyo.*

SAMPSON, P. D. and SIEGEL, A. F. (1985). The measure of 'size' independent of 'shape' for multivariate lognormal populations. *J. Amer. Statist. Assoc.* **80** 910–914.

STREISSGUTH, A. P., MARTIN, D. C., MARTIN, J. C. and BARR, H. M. (1981). The Seattle longitudinal prospective study on alcohol and pregnancy. *Neurobehav. Toxicol. Teratol.* **3** 223–233.

Comment

Colin Goodall

Fred Bookstein's energy, enthusiasm, leadership, and innovative thinking about morphometrics are highly valued, greatly appreciated, and a spur to further work. The present paper is a major advance in multivariate morphometrics, and contains some of the few substantive results available. The linear spaces for size and shape statistics are derived, however, at the cost of restrictive assumptions, including a simple error structure (the null model), almost uniform deformation (negligible curvature), and small errors (linearized, normally distributed, statistics). This discussion looks at a broader approach, and, while lacking the detail and rigor of Bookstein's paper, suggests that statistical machinery, centered on function estimation, is mostly available.

The author has convincingly demonstrated how to move back and forth between deformations and multivariate statistics. These statistics are based directly on linear combinations of landmarks. I prefer to emphasize a two-stage procedure, in which estimation of the biological process, namely the deformation (strain) tensor field varying in space and time, is primary. Only at the second stage statistics that summarize (are functionals of) the deformation tensor field are used in multivariate comparisons. As Bookstein

Colin Goodall is Assistant Professor, Engineering Statistics and Management Science Department, Princeton University, E-Quad, Princeton, New Jersey 08544.