In. J. Janssen, F. Marco torchino, & J. M. Proth (Eds.)

Data analysis. The Ins and Outs of Solving Real Problems

New York: Plenum, 1987

SECTION 1.2

1987

Multivariate and longitudinal data on growing children.

INTRODUCTION TO THE PROPOSED SOLUTIONS

pp 7 - 11

Pieter M. Kroonenberg

Department of Education, University of Leiden

In this introduction an attempt is made to give an overview of the analyses made of the multivariate growth curves. The detailed description of the data is contained in Sempé *, and will not be repeated here. In short, the data consist of 12 yearly measurements on 8 morphological variables for 30 normal French girls.

All methods used to analyse these data are descriptive, and all authors but one (Mineo) use in one way or another linear combinations of the variables, years and/or individuals to analyse the data. None of the papers deals explicitly with multivariate time series or standard multivariate statistical techniques based on multivariate normal distribution theory. No author uses the analysis of covariance structures approach to longitudinal data (see, for instance, Jöreskog & Sörbom, 1977; Goldstein, 1979; Swaminathan, 1984; and their references).

In his paper Mineo searches for clusters of girls with specific characteristics. To describe these characteristics he primarily presents the separate means and the rate of growth per cluster for

* References without a date refer to papers included in this volume.

each of the variables. Furthermore, logistic curves are used to fit the means of the variables for each cluster. It should be noted that none of the authors either link their contributions to the basic literature on physical anthropology (see e.g. Borow et al, 1984) and/or the more commonly used methodology in that field (for an overview of generally used functions to fit growth curves and other related procedures with respect to growth curves, see Goldstein, 1979).

The approach taken by Pontier & Pernin is in a sense somewhat different from most other papers using some kind of linear combinations as their search for indices rests on a motivation exterior to these particular data, and is derived from the subject matter itself. In particular, they look for indices which accentuate either the relative state of growth at a particular age compared to all other ages and relatively independent of the individuals, or accentuate the relative growth of individuals compared to all other individuals and relatively independent of age.

With these indices they intend to assist the relative assessment of new individuals compared to the present (training) sample by using these two indices, and eventually other ones produced by the method (see their section 1.5). In each of the other methods, however, such assessment is also possible by using the group information, and performing some form of multiple regression, or treating the new individual as an 'individu supplémentaire'. For instance, in STATIS (Lavit & Pernin) this would entail using $X_k X_k' DY\Lambda^{-1}$ (section 1.3), where X_k are the data for the new subject k. Similarly in SPECTRAMAP (Lewi & Calomme) and in TUCKALS (Kroonenberg; using $T'X_k U$) such an appraisal could be carried out, be it not as compact as in LONGI (Pontier & Pernin).

Setting the Mineo and Pontier & Pernin study aside for the moment, the other three studies may be ordered by the simultaneity

of analysis of the entire data set. Lewi & Calomme start with analysing the averages per variable and time point over individuals, and after having found a satisfactory representation, they continue by adding more detail of the subjects. In particular, they examine individuals at specific time points and investigate specific individuals over time. This approach identifies growth patterns in length and circumference measurements. After eliminating the above mentioned averages, Lavit & Pernin first perform a global analysis concentrating on the developments of the variability over time, and the relationships between the measurements over the years. Instead of simple averaging to find what they call a compromise solution for the individuals, they use a special kind of 'optimal' averaging of the matrices of scalar-products between individuals at each occasion. This compromise solution is the basis for interpreting the trajectories or evolutions over time for both variables and individuals. After eliminating the same averages, the Kroonenberg study attempts one single analysis of the entire data set to find simultaneously optimal component (or compromise) spaces for both variables and individuals, and their relationships at each age. From the basic parameters of the solution various quantities are then derived to investigate the patterns in more detail. In a sense the increase in simultaneity has to be bought by an increase in complexity of interpretation, which is not necessarily a good thing. It should also be mentioned that Lewi & Calomme use the centred average solution as a reference point and discuss the individual characteristics by portraying them together with the (condensed) average solution, while both Lavit & Pernin and Kroonenberg only display the deviations from the averages. This leads to far greater visual similarity between the figures of the latter two studies, compared to the former.

Two further aspects of the analyses deserve special attention and consideration. The first is a global one. Only the curve fitting of Mineo takes explicitly into account the most salient factor of the design, namely time. In their discussion and conclu-

sion Pontier & Pernin see the non-inclusion of time as an asset rather than a liability of their technique, and their conclusions and arguments could be directly extended to the other analyses. One may, however, question their point of view as an analysis which does not include a design variable should be inherently less powerful than one that does. Whether techniques exist that deal adequately with large amounts of multivariate longitudinal data and explicitly make use of the sequential information is not clear from the present collection of papers.

The second aspect which is crucial in the analyses is the way the data are handled before the analysis proper. Pontier & Pernin base one index on per age centred and standardized variables, using $\bar{x}_{.jk}$ and $s_{.jk}$, and the other index on variables centred and scaled over all individual-age combinations using $\bar{x}_{.j.}$ and $s_{.j.}$. Lewi & Calomme indicate that their analyses use 'logarithmic expression, row- and columnwise centring of the data and global standardization' before the analysis proper. This results in an analysis of ratios such as between the various measurements of length and circumference. Lavit & Pernin indicate that the data were standardized to mean zero and unit variance at each age, which means that they removed $\bar{x}_{.jk}$ and $s_{.jk}$. Finally, Kroonenberg uses yet another standardization, namely the data are transformed by first subtracting $\bar{x}_{.jk}$, and then scaled by $s_{.j.}$. Without going into further discussion of this issue (see some remarks by Pontier & Pernin, section 1.5, and Kroonenberg), it is clear that the standardization used is of vital importance, and is far more complex for three-way data than for two-way data. Standardization critically influences what is exactly analysed by a technique and how the numerical results may be interpreted. The proper standardization should, therefore, be carefully considered.

With respect to the results it is evident that in a general sense the analyses agree, and come to similar general conclusions,

showing that the data are well-structured. If a choice needs to be made, the details provided by an analysis and the ease of interpretation are crucial, while also the simplicity of the analysis itself plays a rôle. From the latter point of view, considering the methodological rather than the substantive or practical orientation of the contributions, one may wonder with Sempé in his final comments at the Symposium if an auxologist, or even more your country physician, might not prefer just the average curves per variable of the group under discussion, and plot the position of a girl brought in for consultation for each variable separately, as was done in Figures 4 and 5 (lefthand panels) of Lavit & Pernin, and leave the nice technicalities for us methodologists to ponder about.

References

J.Borow, R.Harespie, A.Sand, C.Sussanne, H.Hebbelink (Eds). (1984). Human growth and development. New York, Plenum.

H.Goldstein (1979). The design and analysis of longitudinal studies. London, Academic Press.

K.G.Jöreskog & D.Sörbom (1977). Statistical models and methods for the analysis of longitudinal data. In D.J.Aigner & A.S.Goldberger (Eds.), Latent variables in socioeconomic models. Amsterdam, North Holland.

H.Swaminathan (1984). Factor analysis of longitudinal data. In H.G.Law, C.W.Snyder Jr, J.A.Hattie & R.P.McDonald (Eds.), Research methods for multimode data analysis. New York, Praeger.