RAO'S POLYNOMIAL GROWTH CURVE MODEL FOR UNEQUAL-TIME INTERVALS: A MENU-DRIVEN GAUSS PROGRAM

EMET D. SCHNEIDERMAN[a], STEPHEN M. WILLIS[a], THOMAS R. TEN HAVE[b] and CHARLES J. KOWALSKI[c]

[a]Department of Oral and Maxillofacial Surgery, Baylor College of Dentistry, [b]Department of Biostatistics, The University of Michigan and [c]Department of Biologic and Materials Science, School of Dentistry, The University of Michigan (U.S.A.)

For lack of alternatives, longitudinal data are often analyzed with cross-sectional statistical methods, for instance, t-tests, ANOVA and ordinary least-squares regression. Appropriate statistical software has been generally unavailable to investigators using serial records to study growth and development or treatment effects. In an earlier paper (Schneiderman and Kowalski, Am. J. Phys. Anthropol., 67 (1985) 323–333.) we described a suitable method, Rao's polynomial growth curve model (Rao, Biometrika, 46 (1959) 49–58), and provided an SAS computer program for the analysis of a single sample of complete longitudinal data. This method included the computation of an average polynomial growth curve, its 95% confidence band, its coefficients and corresponding confidence intervals. The present paper extends this method to accommodate a sample with observations made at unequal time-intervals. Significant improvements in the accessibility, operation and user-friendliness of the program have been made, facilitated by recent advances in microcomputer technology. This stand-alone GAUSS program (no compiler necessary) runs on PC-compatibles and is available at a nominal cost. In this report we provide an overview of the statistical model, the general structure of the program, and give an example in which a developmental variable (human upper incisor angulation) is analyzed. Ease of installation and use, speed of execution and color graphic displays of growth curves and confidence bands, and most importantly, suitability to longitudinal data, make this method/program a potentially valuable tool for those interested in growth, development, and treatment effects in humans and other species. Some areas in which this method will have immediate applications are orthodontics, maxillofacial surgery and pediatrics.

Keywords: GAUSS; Polynomial growth curves; Longitudinal analysis; Unequal-time intervals; Generalized least-squares

Introduction

In an earlier paper [1] we described Rao's one-sample polynomial growth curve model [2], provided an SAS program (Statistical Analysis System, SAS Institute Inc., Box 8000, Cary, NC 27511, U.S.A.) for carrying out the associated computa-

---

tions, and illustrated both the method and the program on a longitudinal data set consisting of mandibular ramus height measurements. Given longitudinal measurements made at equally spaced time points, this program (1) determined the lowest degree polynomial to adequately fit a single-sample of longitudinal observations, (2) estimated the coefficients for this fitted average growth curve (AGC), and their confidence intervals, (3) obtained the confidence band for the AGC and (4) plotted individual growth curves, the AGC and its associated confidence bands. Subsequently, we described a closely related method of fitting polynomial growth curves in the one-sample situation [3], that due to Hills [4] and provided an interactive GAUSS program (GAUSS Matrix Programming Language, Aptech Systems, Inc., 26250-196th Place S.E., Kent, WA 98042, U.S.A.). This program determined the coefficients of the lowest degree polynomial adequate to fit the AGC and computed explicit measures of velocity and acceleration for both the sample as a whole and on an individual-specific basis.

User reaction to these programs indicated that while many wanted to obtain the output corresponding to Rao's approach [2], few had both access to and sufficient familiarity with SAS and the mainframe computing environment to do so. The earlier program required generous amounts of mainframe computer time (potentially expensive, depending upon the manner in which a given mainframe installation is funded). The plotting capabilities of SAS were also quite limited [1]. In marked contrast, the new software program, written in GAUSS, runs on a stand-alone basis on inexpensive personal computers (IBM-PC family of compatible computers). There is no need to purchase additional software (e.g. compilers or programming language) to run this program. No knowledge of the GAUSS system is required; only a modest understanding of MS-DOS, the de facto universal microcomputer operating system, is needed. Installation and execution of our program are straightforward. The color graphics, error checking routines, easy-to-use menus, increased speed and publication quality graphics (with a laser printer, if available) make it a potentially valuable tool for those interested in analyzing developmental data. These features enable investigators to perform a simple (from the user's perspective) yet appropriate analysis that addresses the unique properties of longitudinal data. Finally, the program is made available at a nominal cost (see Appendix).

As has been noted earlier, methods and computer programs designed explicitly for the analysis of longitudinal data have remained inaccessible to most investigators [1,3,5], the major commercial statistical packages have consistently neglected this area. The purpose of the present paper is to illustrate and make available a user-friendly program for this purpose. A human dentofacial variable is used for illustration. We refer the reader to our earlier work [1] for the details of Rao's method.

Materials and Methods

The method we present here is applicable to investigations where the passage of time (i.e. ordering of observations) is a relevant factor. Subjects need not be growing to be analyzed with this approach. What is important is that the phenomenon of interest occurs over time. For instance, this program can be applied to problems con-

cerning short- and long-term treatment effects with different surgical, medical or orthopedic therapies. It is also suitable for longitudinally following the progression of disease processes or behaviors, such as bruxing. In addition to the dental areas of orthodontics and maxillofacial surgery, this method is also appropriate for questions encountered in pediatrics and obstetrics.

For an example, we apply the method to a longitudinal human craniofacial data set derived from the University School Growth Study at The University of Michigan [6]. In particular, we evaluate developmental changes in upper incisor angulation (Atlas Variable No. 53 from that study: the angle between the sella-nasion line and the long axis of the central incisor) in 22 boys between the ages of 6 and 14 derived from that study. The raw data are not presented here to conserve space.

In short, the steps involved in the statistical method are as follows. A step-up procedure is used to determine the polynomial that adequately fits the data using the fewest number of parameters (lowest order). This involves fitting increasingly higher order polynomial equations to the observations, beginning with the linear or first order. Each successive equation is evaluated for adequacy using a goodness-of-fit criterion. When this is met, the estimates of the parameters defining the AGC are then used for subsequent computations. In particular, a class of simultaneous confidence intervals for the AGC are calculated. These express the precision of the estimate of the AGC.

One factor makes this statistical model different from ordinary polynomial regression, and for that matter, all cross-sectional analyses. The variance-covariance matrix of the original observations is used in estimating the vector of parameters. In ordinary polynomial regression, the parameters are modelled simply as a function of the observations themselves and the times of observation (design matrix); the covariances among variables (i.e. dependencies among means for the T time-points) are implicitly zero. In contrast, the interrelationship among observations, within and between individuals is explicitly incorporated into the longitudinal statistical model. We thus compute the best linear unbiased estimators of the parameters (polynomial regression coefficients) defining the AGC using generalized least-squares (Eqn. 3 in Ref. 1). At each step of the procedure, a scalar quantity G is computed; this reflects the goodness-of-fit of that particular equation to the data. The value G is adjusted for sample size, number of parameters and time points, yielding an F-statistic that is then evaluated for significance; if less than the critical value of the F-statistic for the appropriate degrees of freedom and chosen significance level (e.g. 0.05), then that equation is designated to be an adequate fit. If the equation is not an adequate fit, the next higher order polynomial is used, and a new G is calculated. The method involves stepping up through the equation having one less parameter (P) than time-points (T).

Structure of program

The analysis begins with a longitudinal data set, multiple serial observations made on the same subjects. The method requires that the data be complete, i.e. all time-points represented by all subjects (N). Time-points (times of measurement) can be equally or unequally spaced. Three or more time-points are required. Also, there must be considerably fewer time-points than subjects. The method is *multivariate* in

the sense that the measurements made at the *multiple* time-points are considered simultaneously, however, only a *single* anatomical, physiological or behavioral measurement (variable) is considered in an analysis. The only assumption concerning the sample is that it be derived from the multivariate normal distribution. No assumptions are required concerning the structure of the variance-covariance matrix (Σ). The implications of the structure of Σ, a key element in multivariate methods, is discussed elsewhere [7].

The computer file containing the data to be analyzed should be organized such that each row is a subject and each column is a time-point. Thus, we use an $N \times T$ matrix $\mathbf{X}$ with elements x_{ij} corresponding to the value of the measurement under consideration for individual i ($i = 1,2,...,N$) at time j ($j = 1,2,...,T$). This data set can be in ASCII (American Standard Code for Information Interchange) format, as one might typically generate using a word processor or a cephalometric analysis program, for example Ceph-Master Software (Trilobyte Software, 15894 Northville Rd., Plymouth, MI 48170, U.S.A.). Data can also be in GAUSS format, which requires additional steps to create, but is simpler to use, once created. The program release notes, supplied with the software, give more detailed information on the latter format.

The program is then invoked with a single command: *gaussrun gca7*. Soon thereafter, the program menu appears. The user is then prompted for the location of the data files, which can be on a different drive (floppy or hard disk) or directory from the program itself. The indicated directory is searched and the names of those data files in either GAUSS or ASCII format are displayed. The user then highlights the file of choice using the cursor arrow keys and selects the file with the return key. Using this menu, the user supplies various simple pieces of information concerning the structure of data, and the manner in which it is to be analyzed. These options, implemented in the form of questions, include:

(1) Are the observations (measurements) made at equal time points? That is, are the time intervals between each data collection point t the same? If not, the user is prompted for each of the time-points. Fractional time-points (i.e. 1.3, 12.65) are allowed. If the time-points are equal, the user is given the option of starting at 1 (default) or any other starting time. The interval between data points can also be specified by the investigator, using whatever units are desired.

(2) What level of significance (α) does the user wish to use? Usually, α is initially set at 0.05. After each run of the program, the user may supply a different α, if desired, and rerun the program.

A further feature of the menu is an error checking routine which detects keyboard entry errors, rejects the erroneous entry, and prompts for a correct response. This helps prevent annoying program crashes. After the menu is completed the computer starts the numerical portion of the program. The numerical output is shown on the screen and is also written to a disk file called *gca7.out* which may then be printed for a hardcopy of the program output. This output file can be readily modified, highlighted etc., using a word processor. Finally, the graphics part of the program generates full color plots of the data and confidence bands. An example of the program output is given below.

Results

Although the program generates internally all the intermediate statistics presented in our earlier SAS program [1], we have rewritten the GAUSS program to deliver the edited output in a concise, tabular format which is easier to read. The output starts with a header which includes the date and the time the program was run, and proceeds with the analysis. What follows is an abridged version of the numeric and graphic output generated by the program.

Program output

```
DATA FILE :    UIEM .ASC

DATA SET NUMBER: 1

DATA FORMAT IS ASCII

NUMBER OF SUBJECTS: 22

TIME POINTS: 9

TIME POINTS ARE EQUALLY SPACED.


HIGHEST POTENTIAL EQUATION IS THE     7th_deg


SAMPLE MEANS FOR EACH TIME POINT:

    92.708      96.531      99.716      103.82      103.8     103.64    103.07

    102.45      102.3


N =        22

T =         9

      TEST FOR ADEQUACY OF FIT OF A     LINEAR  EQUATION
    _____


            F-statistic :     18.188

            Probability : 2.8601E-006


    _____


    LINEAR  equation of this data is inadequate.

    QUADRAD  equation will now be examined.

    TEST FOR ADEQUACY OF FIT OF A     QUADRAD  EQUATION
    _____
```

```
                F-statistic :    2.789072

                Probability : 0.04723166
```

QUADRAD equation of this data is inadequate.

CUBIC equation will now be examined.

TEST FOR ADEQUACY OF FIT OF A   CUBIC EQUATION

```
           F-statistic :    2.170228

           Probability :   0.1059535
```

CUBIC specification is OK.

COEFFICIENTS FOR AVERAGE GROWTH CURVE

TAU

```
        1 .    1.472
        2 .    26.29
        3 .   -2.227
        4 .    0.06196
```

AVERAGE GROWTH CURVE

Time Point	AGC	Length of Half Band	95% Confidence Band		
6.00	92.41	6.931	85.48	–	99.34
7.00	97.61	5.33	92.28	–	102.9
8.00	101	5.225	95.74	–	106.2
9.00	102.8	5.275	97.56	–	108.1

10.00	103.6	5.423	98.17	–	109
11.00	103.6	5.727	97.89	–	109.3
12.00	103.3	5.891	97.39	–	109.2
13.00	102.9	5.652	97.3	–	108.6
14.00	103	5.949	97.04	–	108.9

The graphic output, which is displayed as a high resolution color image on an EGA or VGA monitor, is presented in Figs. 1–3. The graphs were produced directly by the program on a laser printer, and merged with the numerical output using a word processor – two very simple steps.

Discussion and Conclusions

This method and program provide explicit and reliable information concerning the measurement being analyzed. An average growth curve, represented by a polynomial regression equation is generated. The lowest order equation (i.e. having fewest terms) is determined. Thus, the program will first attempt to fit a linear equa-
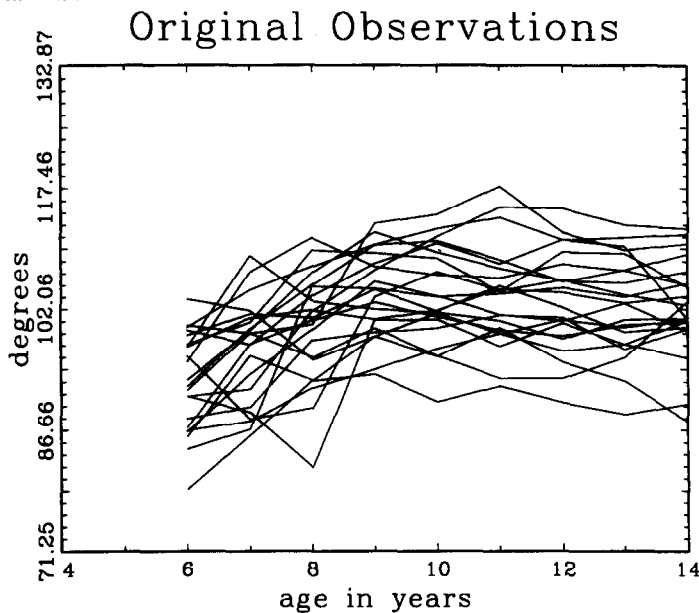

Fig. 1. Upper incisor angulation relative to the sella-nasion line in boys. This is variable No. 53 from Ref. 6. Unfitted growth curves for 22 subjects having complete longitudinal data.
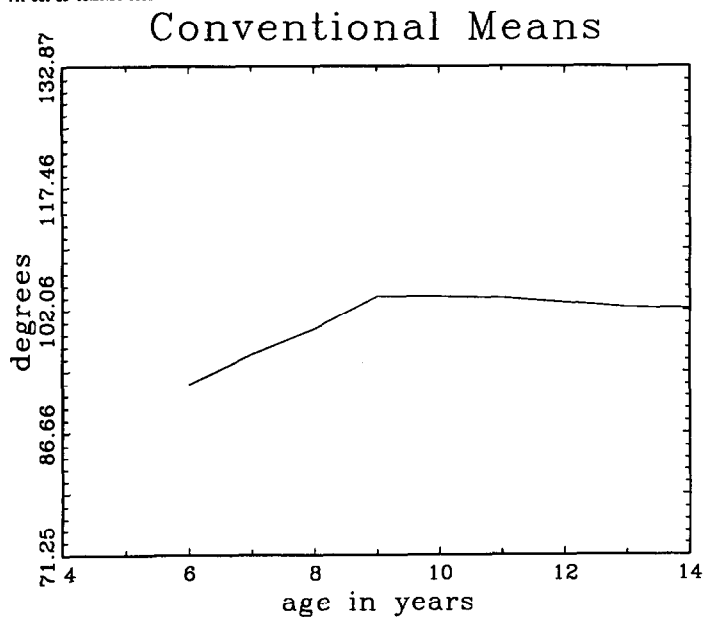
GAUSS Fri Oct 05 15:29:54 1990

Conventional Means

Fig. 2. Conventionally calculated means for each of the nine time points at which upper incisor angle was measured.

GAUSS Fri Oct 05 15:32:22 1990
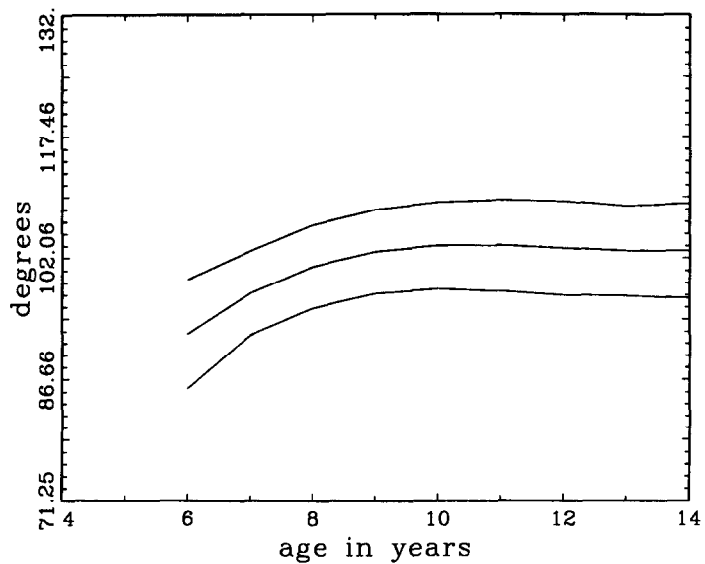
Average Curve & 95% Confid. Band

Fig. 3. Average growth curve and 95% confidence band for upper incisor angle estimated using Rao's polynomial growth curve model.

tion. If, as in the case of our example, neither the linear or quadratic equation meet the goodness-of-fit criteria, the program will go on to fit a cubic equation. Thus, a sigmoidal curve as defined by a cubic equation (i.e. two inflections) is suitable. The curve is expressed by its coefficients $r_1, r_2, ..., r_p$, or alternately by the average values at any of the T (or other) time-points.

A set of simultaneous 95% confidence intervals are computed for the average growth curve. Together these define a confidence band or region around the AGC. The explicit interpretation is that we have 95% confidence that the true population growth curve falls within the band. For example, we have 95% confidence that the true curve for upper incisor angulation is between 95.74 and 106.2 degrees at 8 years of age (t_3).

Many of the popular commercial statistical packages now offer sophisticated procedures such as repeated measures ANOVA and ordinary polynomial regression, which would appear to be suitable for the analysis of longitudinal data. However, these procedures should not be mistaken for true longitudinal methods; they make explicit assumptions about the correlation/covariance structure of the data that can not be expected to hold true with most longitudinal data sets. Repeated measures ANOVA assumes uniformity of covariance, i.e. the requirement that all correlations between measurements be equal. Methods based upon ordinary least-squares (OLS) require that all the correlations between pairs of measurements be zero. With serial measurements made on the same individuals it is highly unlikely that these conditions will be satisfied. The exact operational differences between Rao's approach [2] and these more familiar methods are addressed in a recent paper [8]. Included in this paper is the quantification of error introduced by the inappropriate use of cross-sectional methods. As has been shown using simulated [9] and real data [8], the error is of sufficient magnitude to warrant concern. The method we present here is a straightforward tool for analyzing longitudinal data correctly, thereby eliminating this error.

Acknowledgments

Appendix

Obtaining the program
This software can be obtained on 5.25 in. 360K floppy disk by sending US$10 to defray the cost of handling and licensing fees. The program will run on an IBM-PC/XT/AT or 80386 compatible computer. The computer must be equipped with a numerical coprocessor from the 8087 family and 640K of memory. The computer must be configured so that at least 430K of memory is available, i.e. not tied up with memory resident programs such as Windows. For an operating system, MS-DOS, version 3.3 or higher is required. EGA or VGA graphics capability is required to display the color graphics. No additional software is required (other than what one would normally use to enter a data set); run-time modules are supplied with the pro-

gram so that no compiler or interpreter is necessary. The program, written in GAUSS, version 2.0, revision 20, requires no additional installation or modification, and is run with a single command. When requesting the program, address inquiries to the first author, and make checks payable to Baylor College of Dentistry.

References

1   Schneiderman ED and Kowalski CJ: Implementation of Rao's one-sample polynomial growth curve model using SAS, *Am J Phys Anthropol,* 67 (1985) 323–333.

2   Rao CR: Some problems involving linear hypotheses in multivariate analysis, *Biometrika,* 46 (1959) 49–58.

3   Schneiderman ED and Kowalski CJ: Implementation of Hills' growth curve analysis for unequal-time intervals using GAUSS, *Am J Hum Biol,* 1 (1989) 31–42.

4   Hills M: A note on the analysis of growth curves, *Biometrics,* 24 (1968) 189–196.

5   Palta M and Cook T: Some considerations in the analysis of rates of change in longitudinal studies, *Stat Med,* 6 (1987) 599–611.

6   Riolo ML, Moyers RE, McNamara JA and Hunter WS: *An Atlas of Craniofacial Growth, Monograph No. 2, Craniofacial Growth Series,* Center for Human Growth and Development, Ann Arbor, University of Michigan, 1974.

7   Ten Have TR, Kowalski CJ and Schneiderman ED: A GAUSS program for analyzing one-sample longitudinal data sets which satisfy the two-stage polynomial growth curve model, *Am J Hum Biol,* 3 (1991) 269–279.

8   Schneiderman ED, King AH and Kowalski CJ: Error involved in using cross-sectional statistics on longitudinal craniofacial data, *J Dent. Res.,* 70(S1) (1991) 334.

9   Hoel PG: Methods for comparing growth curve types, *Biometrics,* 20 (1964) 859–872.