Automatic Processing of Growth Data

BENI SOLOW, DR. ODONT.

INTRODUCTION

In the evaluation of physical growth and development, growth curves play an important part. The construction of such curves is difficult and time-consuming and presents theoretical as well as practical problems. A system for automatic processing of growth data with the aid of computers has therefore been developed at the Department of Orthodontics of the Royal Dental College in Copenhagen.

As is often the case when work which has previously been carried out manually is adapted for computers, the automation has given rise to a new appraisal of the principles and methods of analysis. The employment of computers has also made possible an expanded and improved statistical description of the data, particularly with regard to the evaluation of variability. Moreover, the availability of automatic plotting equipment has facilitated the actual drawing of the growth curves.

An outline of the system for automatic processing of growth data will be given below, after some aspects in the collection and analysis of the data have been considered. The first of these is concerned with the timing of examinations in growth studies; the others relate to the analysis of the data.

TIMING OF OBSERVATIONS

In the collection of longitudinal and mixed-longitudinal information for growth studies, the observations are usually made on the annual or semi-annual birthdays of the subjects. This procedure is natural in studies starting at birth, and has also been applied in studies commencing at later ages because it facilitates the subsequent analysis of the data. Close consideration reveals, however, a certain disadvantage in this procedure. The various phases of maturation, for instance, prepubescent minimum and pubescent maximum in growth rates, occur at different distances from the annual birthday in different individuals. When the annual observation is made for all subjects on the birthday, the information on the time interval between the annual birthday and these phases of maturation is eliminated. With sufficiently large samples, this disadvantage can be avoided by allowing the date of annual observation to occur at a random distance from the birthday in different individuals. This procedure has been followed in the Growth Study at the Department of Orthodontics since its inception in 1951.[2] The difference between the two methods is illustrated by Figures 1 and 2.

Figure 1 shows the age distribution of a hypothetical mixed-longitudinal sample in which the conventional method of observing all individuals on their birthdays has been followed. The composition of cross-sectional subsamples changes only at the birthdays (vertical lines). In the construction of velocity curves, minima, maxima, and points of inflection can therefore be plotted only to the nearest year.

An example of the distribution of a sample in which the date of observation is randomized with respect to birthday is given in Figure 2 which shows a mixed-longitudinal sample of seventy-six boys followed at the Department of

---

From the Department of Orthodontics, Royal Dental College, Copenhagen, Denmark.
Supported by USPHS research grant HD-154 from the National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA.

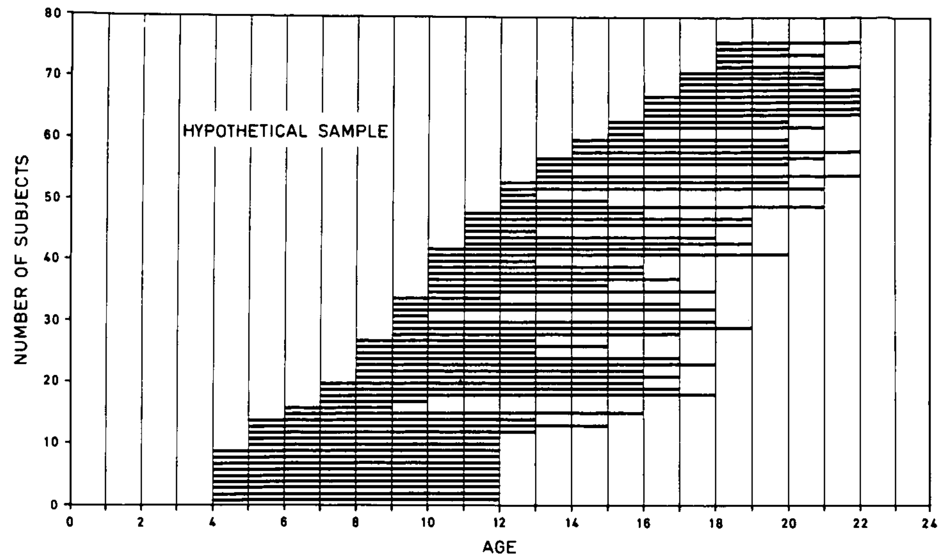

Fig. 1 Age distribution of a hypothetical mixed-longitudinal sample in which the subjects are observed on their birthdays (vertical lines). Horizontal lines indicate the period of observation for each subject.

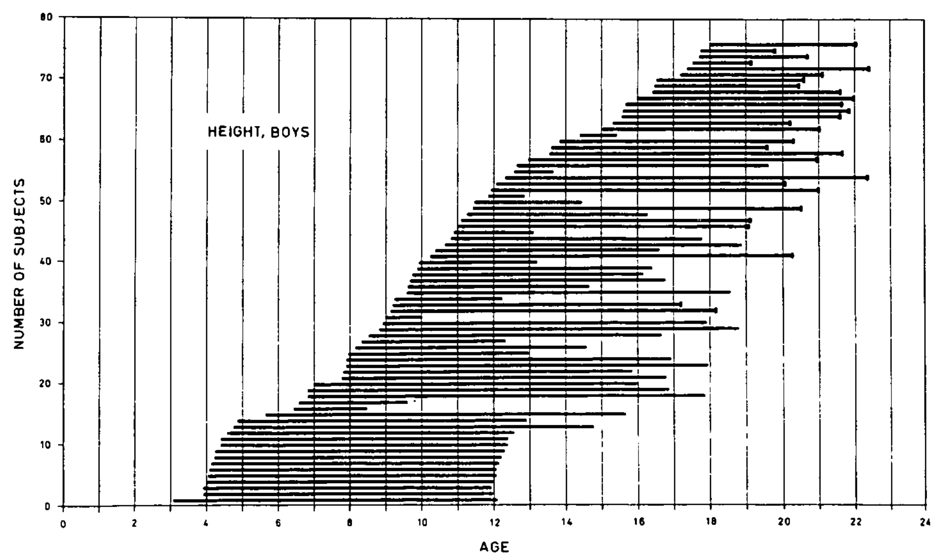

Fig. 2 Age distribution of a mixed-longitudinal sample of 76 Danish boys followed at the Orthodontic Department and analyzed for growth in height. Horizontal lines indicate. period of observation for each subject. Timing of observations is independent of birthdays (vertical lines).

Orthodontics. In spite of the interval between observations for each individual still being one year, it will here be possible by a suitable analysis to determine the time of fluctuations in growth rhythm to, say, the nearest month.

The disadvantage of employing a randomized date of observation is that midpoints and boundaries for the periods of observation vary from individual to individual. One subject is observed at the ages of 10 years 2 months, 11 years 2 months etc., and another at the ages of 10 years 7 months, 11 years 7 months, etc. Conventional statistical processing which requires identical class boundaries for all the subjects cannot, therefore, be carried out. This difficulty has been eliminated, however, by the adoption of a technique of multiple interpolations in the processing of the data.

Multiple Interpolations

Instead of using the measured values directly in the calculation of mean and variability, it was found useful in the processing of the data first to construct individual growth curves. These curves were not actually drawn, but were represented in numerical form in the computer. From the individual curves, values corresponding to each month of age could then be read. The further calculations of mean and variability were subsequently based upon such interpolated values.

While the direct utilization of measured values would yield one observation per year for each subject if annual examination were used, monthly interpolation on the individual curves yields twelve observations per year for the same subjects. The considerable increase in the number of observations thus obtained is of particular value in a sample where the date of observation is randomized with respect to birthday. If, in such a sample, monthly calculations were to be based only on the annually measured values, only one twelfth of the subjects forming part of the sample at any particular month would be exploited. When observations are obtained by monthly interpolation, observations at any particular month are available from all subjects forming part of the sample at that time.

It is an important consequence of this technique that, since at each month a twelfth of the interpolated observations change their values due to the randomized date of observation, the gradual change with age of the size and growth rate in the sample will be reflected in the changes of the monthly means and limits of variability, and an almost smooth curve will be produced although no smoothing is performed.

The technique of multiple interpolations may be used also with samples where observations have been made on birthdays. The improvement over conventional methods here depends on the type of individual curve used for interpolation. With zero-order step-curves no improvement is obtained, but with first- and higher-order curves a smoothing effect caused by the small class width is found.

Limits of Variability

The calculation of the curves indicating the limits of variability around the mean curve presents certain difficulties. A simple procedure for determining such curves consists of stating for each age group ± 1 or 2 standard deviations, or certain percentiles calculated on the standard deviation. This procedure requires that the observations of size or growth rate at each age are normally distributed about the mean. It has been demonstrated theoretically[16] as well as in practice,[7] however, that the observations are not always symmetrically distributed about the mean. Variability curves calculated on the basis of the standard deviation consequently do not always express the true variability of the observations.

A more exact statement of limits of variability demands determination of the true percentiles in the distribution at each age, and this requires knowledge of the shape of the distribution. Unfortunately, however, the shape of the distributions to be described at each age is unknown, being one of the factors to be investigated; furthermore, it changes continuously with age in periods of acceleration and deceleration of growth.

To overcome this difficulty one can, in large samples, arrange the observations at each age according to size and choose as percentile the $ixN/100$'th observation, where N is the number of observations and i the desired percentile. In small samples, however, the method will be rather susceptible to sampling variations.

A more universally applicable method consists in determining the percentiles mathematically through a maximum utilization of the knowledge of the general characteristics of the types of distribution under examination. Although the distributions are not always normal and their exact form not known, we know that they are unimodal, continuous, more or less skew curves which in cumulative form will have a sigmoid shape with an upper and lower asymptote. Small sections of such curves are described with a curve of the type $y=e^{ax+b}$, and determination of a given percentile can therefore be made by fitting a curve of this type to the cumulative distribution of observations for the age in question. Fitting can utilize, for instance, the least squares method, and must be repeated for each percentile determination. This method has been employed in the data processing to be described. It is less dependent on random fluctuations than the simple counting of the ordered observations, and has given good results with samples as small as twenty observations at each age.

Computer Processing

The computer program for the automatic processing of growth data was developed partly in FORTRAN and partly in FAP for an IBM 7074 computer. In the following an outline will be given of the various phases in the data processing.

The basic data for the computer processing consist of the serial information on the age at observation and the size of the variate being studied given for each subject in the sample. Increments are calculated automatically by the computer, but can also be entered directly instead of the information on size. This is of interest, for instance, when growth increments are measured directly from superimposed x-ray cephalometric films oriented after metallic implants in the jaws as described by Björk.[2]

When the data have been punched on cards and entered into the computer the following analysis takes place for *each subject*: The set of observations is given the numerical forms of a distance curve and a velocity curve. In the present version of the program a first-order curve is used for the distance curve, i.e., the points representing the observations are connected by straight lines. The velocity curve is constructed numerically as the corresponding zero-order curve, a step-curve where the level of each step indicates the average velocity in the period between the observations calculated as

$$\frac{\text{Increment x 12}}{\text{Interval between observations (in months)}}.$$

The subject's distance and velocity curves are then interpolated for each month of the whole period under in-

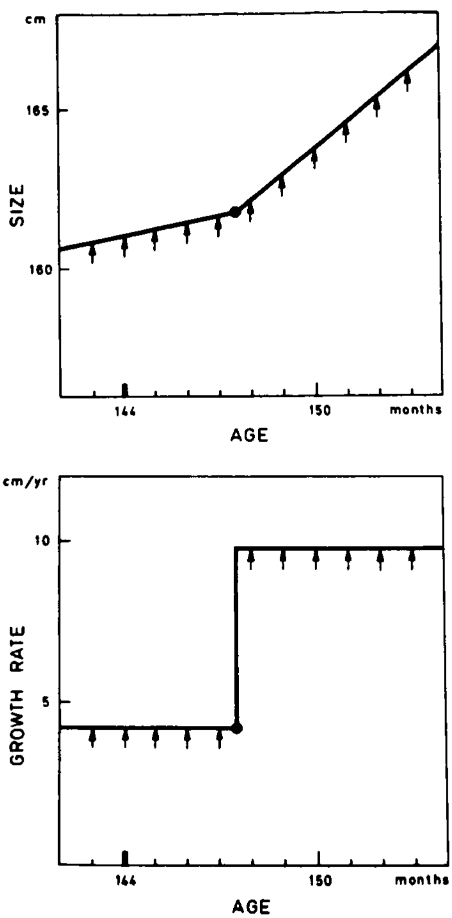

Fig. 3 Multiple interpolations on individual distance and velocity curves. Small segments of a distance and a velocity curve, including one observation point, are shown. The distance curve is obtained by connecting observation points with straight lines. The level of the velocity curve gives the corresponding average velocity of growth. Arrows illustrate monthly interpolation performed by the computer independent of timing of observations.

vestigation (Fig. 3). The subsequent computing of mean and variability curves is based on these interpolated values and not on the original, observed values.

The values of the curves read by the computer are stored in a table giving size or growth rate of the variate for each individual at each month. The columns represent the subjects and the rows, the age in months. Since in a mixed-longitudinal study each subject is observed only for a part of the total span of age studied, table entries outside this period of observation are marked as nonavailable.

When all the subjects have been processed, and all the columns of the table filled in, the computer proceeds to the *statistical analysis* of the data. This analysis is carried out for one row at a time, each row containing a distribution describing the size or growth rate of the group at a specific month. The statistical description of such a distribution falls into three parts: 1) First, the conventional measures of mean and variability are determined. 2) Then certain parameters describing the shape of the distribution are calculated according to Solow.[17] Skewness is expressed by $\sqrt{b1}$ and kurtosis by $b2$ and a (Pearson,[12] Geary,[8,9] and Rao[14]). Tables of 5% and 1% significance limits for these parameters are given by Pearson and Hartley.[13] 3) Finally, selected percentiles are calculated. The distribution at the month in question is first represented cumulatively. The abscissa of the observation closest to a desired percentile is then selected, and one of the curves $y=e^{ax+b}$ (percentiles below the modal point of the distribution) or $y=1-e^{-(ax+b)}$ (percentiles above the modal point of the distribution) is fitted by the least squares method to all the points having the preceding, the same, and the following abscissae. The percentile is read from this curve (Fig. 4). The procedure is repeated for each percentile.

When the above calculations have been performed for each row in the table, the statistical description of the curves can be printed (Fig. 5). The description of the growth of a variable during twenty-five years takes up a few pages of output which can be assembled directly as a booklet.

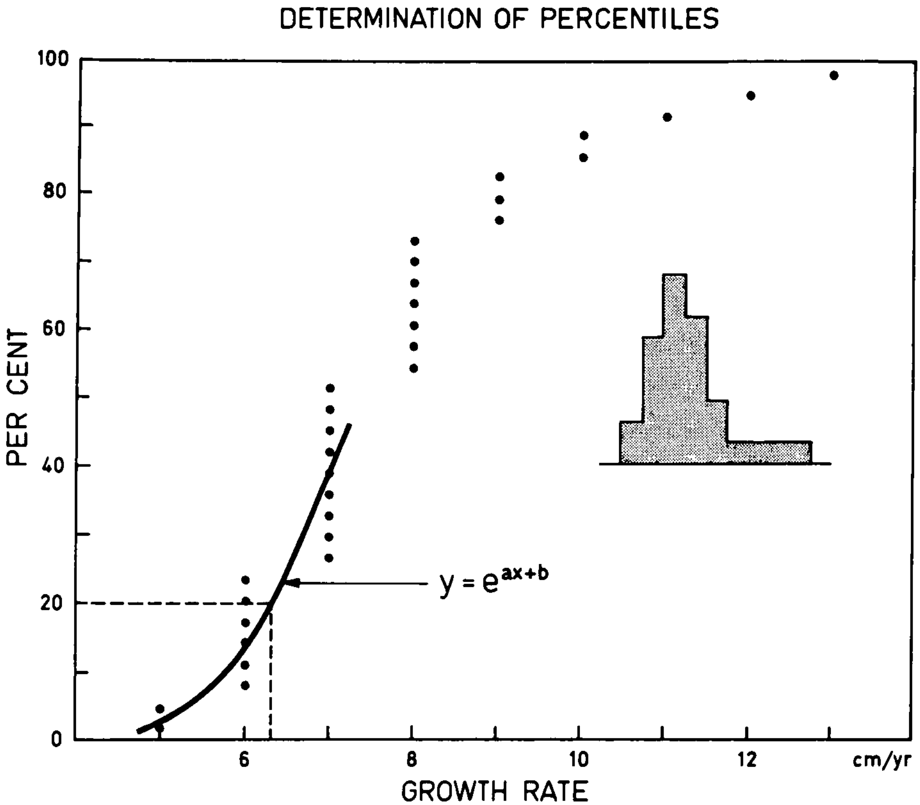

Fig. 4 Calculation of percentiles. Shaded figure indicates distribution of growth rate at a specific month. Points depict the same distribution cumulatively. X-value of the observation closest to the desired percentile is selected. The curve $y = e^{(ax+b)}$ is fitted by the least squares method to all points having the previous, the same and the next X-value. The percentile is read from this curve.

AGE YR	AGE MO	N	M	S.E.	S.D.	RANGE MIN.	RANGE MAX.	PERCENTILES LOWER	PERCENTILES UPPER	SKEWNESS SQRT(B1)	KURTOSIS B2	KURTOSIS A
13	0	36	7.05	0.36	2.14	3.5	12.5	4.5	9.9	0.496	2.70	0.8305
	1	36	7.07	0.35	2.13	3.5	12.5	4.5	9.9	0.511	2.73	0.8306
	2	35	7.17	0.35	2.09	3.5	12.5	4.8	9.9	0.513	2.81	0.8288
	3	34	7.20	0.37	2.19	3.5	12.5	4.7	9.9	0.440	2.45	0.8565
	4	34	7.55	0.37	2.15	3.5	12.5	5.0	10.1	0.223	2.33	0.8507
	5	34	7.55	0.37	2.15	3.5	12.5	5.0	10.1	0.223	2.33	0.8507
	6	34	7.54	0.37	2.18	3.5	12.5	5.0	10.1	0.256	2.24	0.8602
	7	35	7.64	0.37	2.19	3.5	12.5	5.0	10.3	0.180	2.17	0.8618
	8	36	7.77	0.35	2.11	3.5	12.5	5.1	10.3	0.077	2.30	0.8562
	9	36	7.83	0.35	2.13	3.5	12.5	5.1	10.3	0.003	2.24	0.8641
	10	36	7.90	0.34	2.02	3.5	12.5	5.1	10.1	-0.217	2.46	0.8467
	11	37	7.75	0.36	2.18	3.4	12.5	4.9	9.9	-0.342	2.47	0.8542

Fig. 5 Part of a page of computer output giving monthly statistical description of growth rate in height for boys age 13 years 0 months to 13 years 11 months. The unit is cm/year, the sample is the same as in Fig. 2. N = size of subsamples, M = arithmetical mean, *S.E.* = standard error of the mean, *S.D.* = standard deviation. The percentiles given are the 10 and the 90.

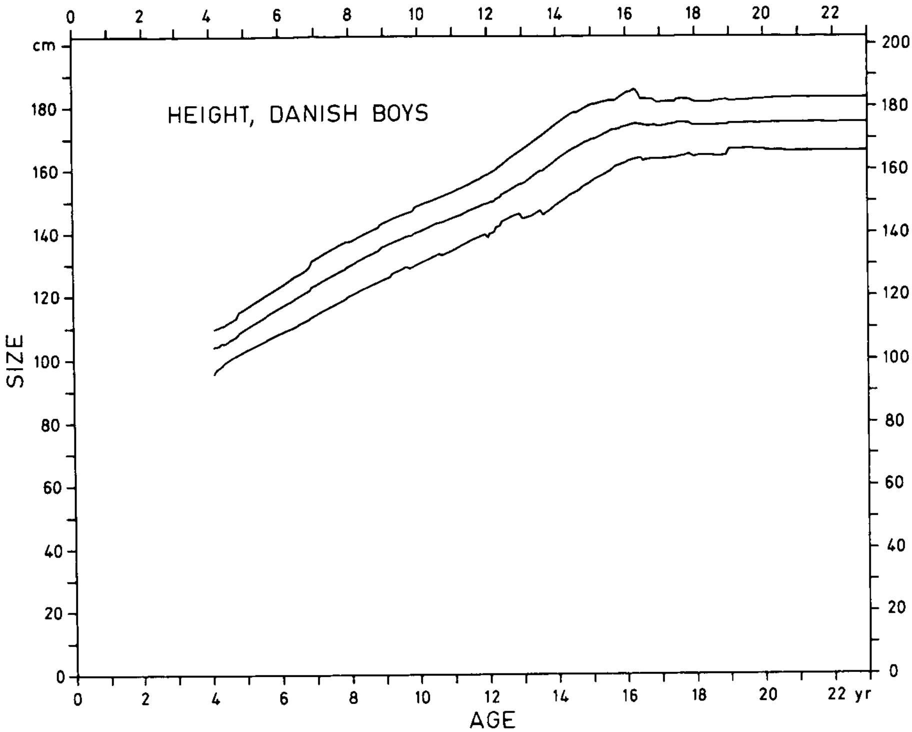

Fig. 6 Distance curves produced by automatic processing of data on height in the sample shown in Fig. 2. Mean, 10 and 90 percentiles.

Growth Curves

The amount of detail in the statistical description provides an excellent basis for a further mathematical or statistical analysis of details in the shape of the curve, but makes manual tracing of the growth curves difficult. Most university computers are, however, equipped with plotting equipment, automatic pens, the movement of which in a coordinate system is directed by the computer. By suitable programming the plotter can be employed to present the statistical description of the growth graphically. There is naturally a free choice of which of the computed parameters to plot, and we have decided to let the computer trace a mean curve and two variability curves; for example, the arithmetic mean and the 10% and 90% percentiles. If "growth channels" are desired, it is quite easy to program the computer to draw four or six variability curves. Distance and velocity curves obtained by using this system are shown in Figures 6 and 7.

Discussion

The construction of growth curves raises several problems of a theoretical nature. The relation between individual curves which describe the growth of a dimension in the individual subjects, and group curves which characterize the growth in a group of subjects is complicated. For both types of curves, moreover, there are difficulties in transforming the irregular curves representing the raw data into smooth curves that may be used for further analysis.

Individual curves are essential for the detailed study of the various phases of maturation, for instance, in the analysis of the timing and magnitude of pre-

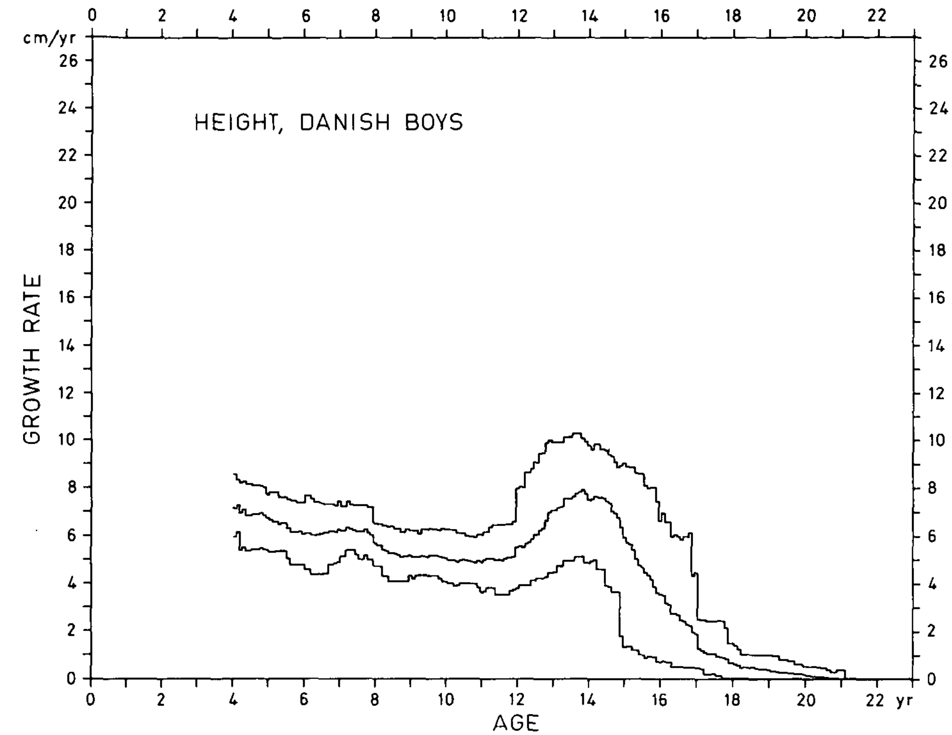

Fig. 7 Velocity curves produced by automatic processing of data on height in the sample shown in Fig. 2. Mean, 10 and 90 percentiles.

pubescent minimum and pubescent maximum in growth as well as the timing of the termination of growth.[2,3] The detailed shape of the individual curves is difficult to determine because the error of measurement sets a lower limit for the length of the interval between observations. In order to obtain a more efficient description of the individual curves, attempts have been made to fit different mathematical curves to the observed data.[4,5,6,10] The fitted curves are effective in describing marked fluctuations in growth, but suffer from the drawback that the number of minima, maxima, and points of inflection must be determined in advance.

Group curves which show the mean and variability in growth are of interest as norms when growth of the individual subject is assessed clinically. The problems involved in the construction of such curves have been discussed by Merrell,[11] Shuttleworth,[15] Tanner,[18] and Tanner, Whitehouse and Takaishi.[19] Construction of group curves may be based on a chronological scale or a maturation scale by superimposing on maximum pubescent growth. The two types of group curves have different properties, and both types are of interest when the growth of the individual subject is to be appraised. The mean and variability curves determined by the method described above are those based on a chronological scale.

As previously mentioned, *interpolation* on the individual curves was used in the calculation of group curves. The values obtained by this interpolation obviously depend upon the method used in the construction of individual curves from the annual observations. A wide range of curve-fitting procedures is available for obtaining such curves. Unfortunately, however, any choice of a particular higher-order curve implies some form of predetermination of the number of minima, maxima, and points of inflection to be found in the fitted curve. In a subsequent analysis of the group curves it would be difficult, therefore, to decide whether particular details in the shapes of the curves were due to properties exhibited by the sample or merely reflected the assumptions made in the curve-fitting procedure.

To avoid this disadvantage it was decided not to fit higher-order individual curves to the observed data at the present stage of the project. Instead, the simplest possible types of individual curves were used, namely, a first-order curve for size and the corresponding zero-order step-curve for growth rate (Fig. 3). No assumptions thus were made concerning the presence of minima, maxima, and points of inflection in the individual curves, and any such features found in the group curves consequently can be considered to reflect properties of the sample. The method, moreover, has the advantage that it may be used also with variables for which the shapes of the growth curves are more or less unknown, for instance, angular dimensions used in the study of craniofacial growth, dimensions of the dental arch, etc.

If, on the other hand, higher-order curve-fitting is considered desirable in a particular project, the design of the system is such that a curve-fitting procedure may readily be inserted.

Although no *smoothing* of the observed or processed data has been carried out, and despite the simple type of individual curves used, it will be apparent from Figure 7 that the computed and plotted curves appear almost continuous. This is due to the combination of two factors, the very small class width used in the interpolation, and the method of sampling where the date of observation in different individuals lies at a random distance from the annual birthday. The effect of the latter factor can be seen by selecting the annual observations from longitudinal material measured on annual and semiannual birthdays[20] and subjecting these data to the same automatic processing as has been described above (Fig. 8). The mixed-longitudinal material used in Figure 6 produces a considerably more detailed curve than the longitudinal material used in Figure 8, thus indicating that in studies based upon annual observations a considerable amount of information is lost when the observations are made on the subjects' birthdays.

The smoothness of the curves which have been based on data sampled at a randomized distance from the birthday facilitates the construction of charts for clinical purposes. The necessary further smoothing can be made free-hand with very small bias. Figure 9 shows a chart produced by manual smoothing of the computer output. The curves show the growth rate in height for Danish boys and are based on the mixed-longitudinal sample shown in Figure 2. This chart can be used in the clinic as an aid in evaluating the growth of the individual subject in relation to chronological age.

Besides the theoretical advantages of the method of randomizing the date of observation with respect to birthday, this method has the practical advantage that the randomizing procedure may consist in observing the subjects on, for instance, the date of their first visit to the growth clinic. The visits therefore can be planned to fit the most convenient working periods of the institution.

One of the most commonly encountered problems in the analysis of growth data has been how to "adjust"

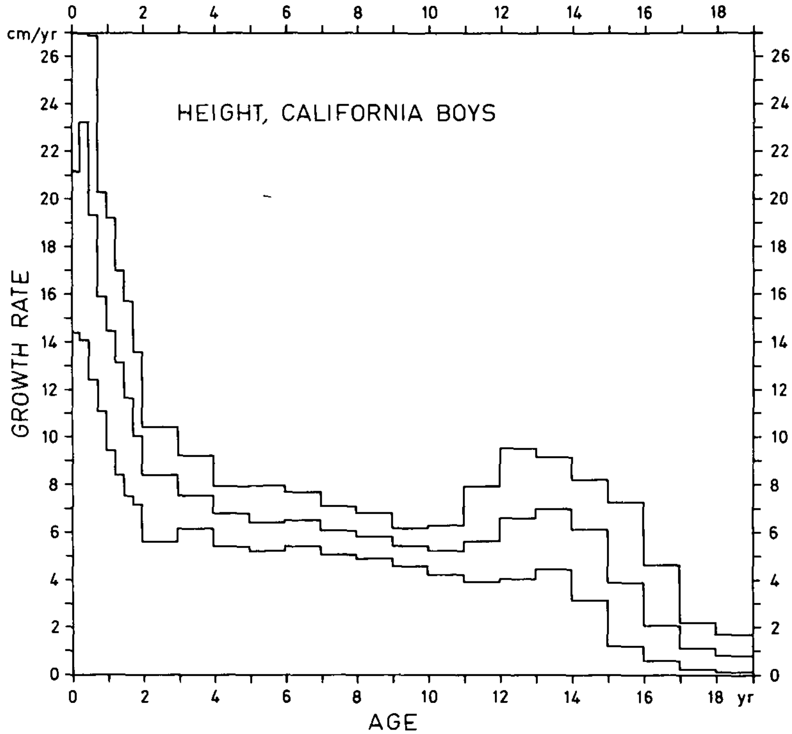

Fig. 8 Velocity curves produced by automatic processing of data on height obtained from Tuddenham & Snyder (1954). The sample consists of 66 California boys, all followed from birth to 18 years and measured on their birthdays. Mean, 10 and 90 percentiles.

the data to compensate for the slight variations in the intervals between observations which arise when the subjects sometimes cannot be measured on the appointed date. In distance curves the procedure has often been used to determine the size of the variate at the desired date by means of interpolation, and to use the "adjusted" value in the further calculations. In the velocity curves similar "adjustments" of increments and observation intervals were used.

In the program described, the need for such "adjustments" of the data has been entirely eliminated by means of the technique of multiple interpolations. No "adjustments" of observation dates are called for, because the timing of the interpolations is independent of the dates of observations, and no "adjustments" of increments are needed in the velocity curve, because the length of the observation interval has been taken into consideration in the formula for calculating the velocity. The elimination of these difficulties constitutes a considerable advantage in relation to conventional techniques of handling data for growth studies.

In conclusion it may be said that the system described for automatic processing of growth data has several theoretical as well as practical advantages as compared with conventional techniques. It may be used with all types of growth data, and practical experience with the system has shown it to be particularly suited for handling of data from mixed-longitudinal studies of

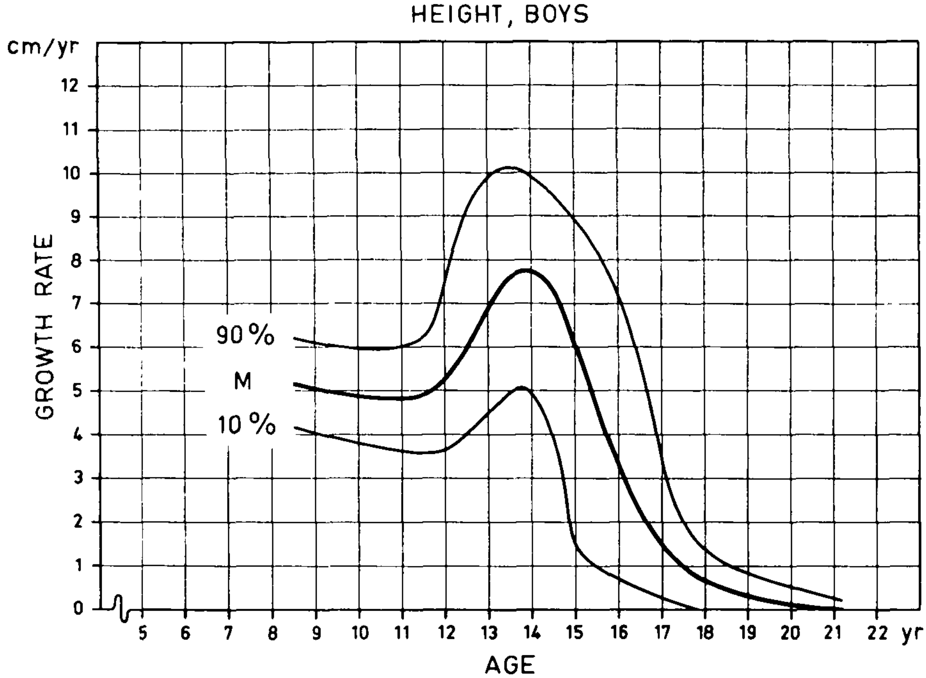

Fig. 9 Growth rate in height for Danish boys. Chart obtained by manual smoothing of computer output for the sample shown in Fig. 2.

craniofacial growth where a relatively limited number of subjects is followed with annual observations.

Summary

A method for automatic processing of growth data by means of computers and automatic plotting equipment is described.

As input, serial information on the ages at observation and the corresponding sizes of the variate being studied is given for each subject in a group. The output presents the size or growth rate of the group and consists of tables giving mean, variability, percentiles, and form of distribution for each month over the period studied, as well as graphs presenting mean and selected percentiles.

It is demonstrated that sampling on birthdays may be undesirable because much information is thereby eliminated from the data. The advantage is demonstrated of a technique of sampling in which the annual observation occurs at different time intervals from the annual birthday in different subjects although constant intervals between observations are maintained.

A method for calculation of percentiles is described which is suitable for use with growth data because it does not require normality of the distribution.

Jagtvej 160
2100 Copenhagen, Denmark

Acknowledgments

The author wishes to express his appreciation to Peter Kierkegaard-Hansen, cand. polyt. et stat., and Inger Koch for their assistance and cooperation in the planning and development of the computer program.

REFERENCES

1. Bayer, L. M. & Bayley N.: *Growth Diagnosis*. Univ. Chicago Press, Chicago, 1959.
2. Björk, A.: Variations in the growth pattern of the human mandible: Longitudinal radiographic study by the implant method. *J. Dent. Res.*, 42:400-411, 1963.
3. ————: Sutural growth of the upper face studied by the implant method. *Acta odont. scand.*, 24:109-127, 1966.
4. Count, E. W.: A quantitative analysis of growth in certain human skull dimensions. *Hum. Biol.*, 14:143-165, 1942.
5. Deming, J.: Application of the Gompertz curve to the observed pattern of growth in length of 48 individual boys and girls during the adolescent cycle of growth. *Hum. Biol.*, 29:83-122, 1957.
6. Deming, J. & Washburn, A. H.: Application of the Jenss curve to the observed pattern of growth during the first eight years of life in forty boys and forty girls. *Hum. Biol.*, 35: 484-506, 1963.
7. Garn, S. M., Rohmann C. G. & Robinow, M.: Increments in hand-wrist ossification. *Amer. J. Phys. Anthrop.*, 19:45-53, 1961.
8. Geary, R. C.: The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, 27:310-332, 1935.
9. ————: Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biometrika*, 28:295-305, 1936.
10. Jenss, R. M. & Bayley, N.: A mathematical method for studying the growth of a child. *Hum. Biol.*, 9:556-563, 1937.
11. Merrel, M.: The relationship of individual growth to average growth. *Hum. Biol.*, 3:37-70, 1931.
12. Pearson, E. S.: A further development of tests of normality. *Biometrika*, 22:239-249, 1931.
13. Pearson, E. S. & Hartley, H. O. (ed.): *Biometrika Tables for Statisticians*, Vol. 1. Cambridge Univ. Press, 1954.
14. Rao, C. R.: *Advanced Statistical Methods in Biometric Research.* Wiley & Sons, New York, 1952.
15. Shuttleworth, F. K.: Sexual Maturation and the Physical Growth of Girls Age Six to Nineteen. *Monogr. Soc. Res. Child Develop.*, 2: No. 5, 1937.
16. Sollberger, A.; Frequency distributions arising from biological movement, *Acta morph neerl.-scand.*, 3: 179-191, 1960.
17. Solow, B.: The pattern of craniofacial associations. A morphological and methodological correlation and factor analysis study on young male adults. *Acta odont. scand.*, 24: Suppl. 46, 1966, p. 36.
18. Tanner, J. M.: Some notes on the reporting of growth data. *Hum. Biol.*, 13:93-159, 1951.
19. Tanner, J. M., Whitehouse H. & Takaishi, M.: Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children 1965 - I & II. *Arch. Dis. childh.*, 41:454-471, 613-635, 1966.
20. Tuddenham, R. D. & Snyder, M. M.: Physical Growth of California Boys and Girls from Birth to Eighteen Years. *Univ. Calif. Publ. in Child Develop.*, 1:183-364, 1954.