ELSEVIER International Journal of Bio-Medical Computing 36 (1994) 187–192

International Journal of Bio-Medical Computing

Implementation of exact and approximate randomization tests for polynomial growth curves

Emet D. Schneiderman*[a], Stephen M. Willis[a], Charles J. Kowalski[b], Ingrid Y. Guo[c]

[a]*Department of Oral and Maxillofacial Surgery and Pharmacology,* [c]*Department of Public Health Sciences, Baylor College of Dentistry, Dallas, TX 75266-0677, USA*

[b]*Department of Biologic and Materials Sciences and The Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, MI 48109, USA*

(Received 12 July 1993; accepted 19 August 1993)

Abstract

Two stand-alone, menu-driven PC programs, written in *GAUSS386i*, which compare groups of growth curves in a completely randomized design using either (a) exact or (b) approximate randomization tests, are described, illustrated, and made available to interested readers. The programs accomodate missing data in the context of studies planned to have common times of measurement, but where some of the measurement sequences are incomplete. The measurement whose growth is being monitored need not have a Gaussian distribution. We consider the hypothesis that the mean growth curves in G groups are the same; and either compute the exact P value (exact test), or estimate, and provide a confidence interval for, the P value (approximate test).

Keywords: Longitudinal studies; Repeated measurements; Polynomial growth curves; Randomization tests; PC programs

1. Introduction

A number of methods for estimating and comparing the average polynomial growth curves (AGCs) in several groups of individuals exist when subjects are measured at identical times, modeled with polynomials of the same degree, and when multivariate normality of the repeated measurements can be assumed. Among these are the methods by Potthoff and Roy [1], Rao [2,3] and Khatri [4], and we have written programs to carry out several of these analyses using SAS and/or GAUSS [5–10]. These methods and programs are able to provide considerable insight into growth and developmental processes whenever the conditions mentioned above are satisfied, but practical circumstances often preclude their application. The assumption of common times of measurement is especially troublesome: individuals invariably miss one or more appointments. Excluding such individuals from the analysis wastes information; estimating the missing values so that they may be

* Corresponding author.

SSDI 0020-7101(93)00932-8

included is difficult, and may introduce additional (strong) assumptions into the analysis which many researchers would rather avoid, although techniques for accomplishing this do exist [11].

Fortunately, there are procedures which do not make any of the above assumptions and consequently are of great potential value to those who must deal with all of the practical problems inherent in collecting and analysing longitudinal data sets. Zerbe and Walker [12] and Zerbe [13] developed randomization tests for the analysis of growth curve responses arising in the context of a completely randomized design, and this methodology has been extended to more complex experimental designs [14]. The general theory of randomization tests is described elsewhere [15–17]; here we consider only those tests dealing with longitudinal data collected in accordance with a completely randomized design. Our discussion parallels that given by Zerbe and Walker [12], but we maintain the notation established in a number of other papers [5–11]. We describe two menu-driven PC programs, written in *GAUSS386i*, implementing the exact and approximate forms of this procedure, and copies of the programs are made available to interested readers. Information concerning hardware requirements and obtaining copies are given in Appendix 1.

2. Randomization tests for longitudinal data

Suppose N subjects are randomly assigned to G groups, n_g to the gth group,

$$\sum_{g=1}^{G} n_g = N,$$

and that the planned times of measurement are

$$t_1, t_2, \ldots, t_T$$

Then, following the methods previously outlined by us [18–21], we compute the distances between the growth curves for all pairs of individuals, where the distance between the growth curves of the ith subject in the gth group and the kth subject the lth group over the time interval (a,b) is defined [12,13] by:

$$d(x_{ig},x_{kl}) = \left[\int_a^b [x_{ig}(t) - x_{kl}(t)]^2 \mathrm{d}t\right]^{1/2} \quad (1)$$

Our programs are limited to the case where the $x_{ig}(t)$ are polynomials. Determination of the appropriate degrees of these polynomials, and the structure of these parts of our programs were detailed by Schneiderman et al. [18,19]. The hypothesis that the mean growth curves in the G groups are equal over the interval (a,b), namely,

$$H_0\colon \mu_1(t) = \mu_2(t) = \ldots = \mu_G(t) \text{ for all } t \in (a,b) \quad (2)$$

may then be tested [12,13] using the statistic

$$Z = \sum_{g=1}^{G} \frac{1}{n_g} Z_{gg} \quad (3)$$

where

$$Z_{gg} = \sum_{i<k} \mathrm{d}^2(x_{ig}(t), x_{kg}(t)) \quad (4)$$

i.e. Z_{gg} is the sum of the squared distances between all possible pairs of individuals in group g.

The exact P value for the test may then be obtained by evaluating Z for each of the

$$R = \frac{N!}{\prod_{g=1}^{G} n_g!} \quad (5)$$

possible random assignments of N individuals to G groups (fixed n_g). If M of these are less than or equal to Z_0, the observed value of Z from the original assignment, the P value is

$$P = M/R \quad (6)$$

Since R may be quite large in practice and the evaluation of every possible Z prohibative, a random sample of assignments may be taken. If in r assignments, m values of Z are less than or equal to Z_0, the P value (Eq. 6) is estimated by

$$\hat{P} = m/r \tag{7}$$

If, as in our program, the sampling of assignments is done with replacement (i.e. if the same assignment may be made more than once), m has a binomial distribution with parameters m and P, and so confidence intervals for P are easily constructed [22], or read from tables [23], or charts [24]. The interpretation of this confidence interval, however, deserves some comment. If, say, the 95% confidence limits for P include values all less than 0.05, then 95% is the 'confidence' that examination of all R assignments would have resulted in the same decision (to reject H_0 at the 5% level of significance). More generally, as long as the confidence interval does not include our prespecified level of significance (α), we would be 95% confident that that examination of all R assignments would have resulted in the same decision. Should the interval contain α, the decision based on all R assignments might differ from the one reached on the basis of the sample of r assignments. The choice of r, then, should be large enough to ensure that the width of the resulting confidence interval is sufficiently narrow, but not so large as to require excessive computing time.

In any event, we provide $(1 - \alpha) \times 100\%$ confidence intervals for P of the form (P_L, P_U), where [22]

$$P_L = \frac{m}{m + (r - m + 1)F_L}$$

$$P_U = \frac{(m + 1)F_U}{r - m + (m + 1)F_U}$$

In the above,

$$F_L = F[1 - \alpha/2;\ 2(r - m + 1),\ 2m]$$

and

$$F_U = F[1 - \alpha/2;\ 2(m + 1),\ 2(r - m)]$$

representing the $(1 - \alpha/2)$th percentile of the F distribution with the indicated numbers of degrees of freedom.

3. The programs

The user is asked to prepare either an ASCII or GAUSS data set containing the values of the measurements for a study in which the times of measurement were planned to be the same for all individuals, but some data may be missing. Periods ('.') are used to represent missing data. One column in the data set should be reserved for the group indicator variable. To illustrate, consider a study with planned times of measurement $t = 1, 2, 3, 4, 5$ and an individual from group g (= $1, 2, \ldots, G$) with observations 20, 30, and 45 at times $t = 1$, 2, and 4. If the group indicator variable is put in the first column, the corresponding row in the data set would be

g 20 30 . 45 .

Note that the (common) times of measurement are not part of the data set; they are entered (once) when running either program.

The program performing the exact randomization test is invoked by the command *gsruni zrte*, and the program for the approximate test by *gsruni zrta*. In both cases the user provides the menu-driven program with information concerning the name and location of the data file, and the times of measurement. The total number of assignments of subjects to groups (R, the number of unique permutations) is printed. If the user is running the exact form of the test, we compute the distances (Eq. 1) and provide the exact P value as in Eq. 6. If the approximate form, the user specifies the number, r, of assignments to be sampled. Given r, the program treats the obtained result as the first data permutation and randomly permutes the data an additional $r - 1$ times. The output includes the estimated P value and the corresponding confidence interval at the user-specified level of confidence (e.g. 0.95). The user then has the option of continuing the sampling process should he or she wish to obtain tighter confidence limits. One might, for example, initially choose r = 1000 and add to this if the resulting confidence interval was equivocal.

4. Examples

We consider two examples, the first employing the exact test, the second the approximate test. Consider first the data set shown in Table 1, consisting of mandibular ramus height measurements (in mm) on $N = 12$ young male rhesus monkeys taken at the $T = 5$ time points $t = 1(1)5$. This data set was previously used by us for illustrative purposes [5,11]; here we have added a group indicator variable and discarded several observations. The number of ways to assign 12 subjects to two groups of 6 each is 924, and we choose the exact form of the test. The time required to complete the procedure will, of course, depend on the configuration of the machine used; but using a 486-based PC running at 33 MHz the elapsed time was approximately 30 s, and the resulting P value was $P = 0.197$. To give some idea of the time needed to complete analyses based on samples of assignments, for the above example, 200 took less than 10 s; 500 less than 20 s; and 700 approximately 25 s.

Table 1
Data set used in the first example

Group	Time				
	1	2	3	4	5
1	25.2	29.0	33.6	35.2	35.8
1	27.3	32.1	37.0	.	.
1	26.3	30.7	36.1	.	.
1	26.0	.	39.0	42.3	44.4
1	25.5	29.5	34.4	38.3	37.9
1	28.2	32.5	36.3	42.3	43.8
2	25.4	33.4	38.0	.	.
2	27.2	34.8	37.2	44.0	44.0
2	26.0	34.5	38.0	43.5	43.8
2	28.5	33.8	38.0	39.2	42.0
2	27.0	.	36.0	41.7	43.8
2	26.0	.	40.2	42.5	43.8

Our second example is based on the data previously considered by Ten Have and coworkers [9,25], consisting of three samples of children living in Guatemala which were studied in depth by Bogin et al. [26]. The children comprising these samples differ in socioeconomic status (SES) and ethnicity: one is of high SES Ladino children (G_1); the second is of low SES Ladino children (G_2); and the third is of low SES Mayan children (G_3). There are 20 individuals in each group and we analyse their statural growth, this being measured $T = 6$ times at ages 7, 8, 9, 10, 11 and 12 years. There are no missing data. We use this data set so as to be able to compare the results with those obtained via the Potthoff–Roy analysis [9], but the user is reminded that the program does accomodate data sets with missing values.

The size of this data set is such that only the approximate randomization analysis is feasible. There are $60!/(20! \times 20! \times 20!) = 5.7783121 \times 10^{26}$ possible assignments of 60 children to $G = 3$ groups with 20 in each group, and it is neither practical nor necessary to perform the exact randomization test (see below). Following Ten Have et al. [9], we fit polynomials of degree $D = 2$ to each individual, and specify a sample of $r = 1000$ random assignments of 20 children to each of the groups. In this case, the time required to complete the analysis was approximately 5 min; the estimated P value (Eq. 7) was $\hat{P} = 0.001$ and the corresponding 95% confidence interval was (0.000,0.006). This indicates that the overall form of the curves are different among the three groups. The Potthoff–Roy analysis [9] also showed differences between the groups ($P = 0.0001$ for coincidence irrespective of the choice of the arbitrary matrix).

The reader will have noted that the time required to complete either the exact or approximate randomization test depends on the machine employed, and on the structure of the data set. It is difficult to provide estimates covering all of the possibilities but, in both programs, the user is kept abreast of progress with a running display of:

```
TOTAL NUMBER OF PERMUTATIONS SELECTED:
CURRENTLY WORKING ON PERMUTATION #
PERCENT CALCULATION REMAINING:
```

and may abort the program if his or her estimated time to completion exceeds expectation by simultaneously pressing the control and break keys.

5. Discussion

Randomization tests are statistical tests in which the data are repeatedly divided, a test statistic is computed for each division, and the P value for the test equals the proportion of the data divisions with as small (or large, depending on context) a test statistic as the value determined from the original data. When all possible divisions or, in our case, assignments, are evaluated, we speak of systematic data permutation, and exact randomization tests. When only a sample of assignments is evaluated, we use the terms random data permutation, and approximate randomization tests. The exact test will be preferred, whenever feasible. Often, however, the user will have to choose the approximate procedure. There are two primary reasons for this, and these are considered in turn below.

First, exact tests are feasible only for very small data sets. With moderate sample sizes, there may be so many possible assignments that it would not be practical to consider all data divisions, even with modern computers. There are, for example, over 5 trillion ways to assign 30 subjects to 3 treatments with 10 subjects per treatment (Ref. 15, p. 20). In an application somewhat different than ours (Ref. 16, p. 14), it was noted that even if 1000 of a total of 16! possible permutations could be generated and evaluated every second, it would take more than 6 centuries to exhaust the list.

Second, random data permutation methods may be effective with as few as 1000 data permutations (Ref. 15, p. 43). It is also true that approximate tests are valid in the sense that if the level of significance of the exact test is α, the probability of a type I error using the approximate test will be no more than α. While random permutation methods are less powerful than systematic methods based on all possible assignments, increasing the number of samples increases the power, and there is often but little loss in power when the approximate test is used. It has been shown (Ref. 15, p. 45), for example, that if $P = 0.01$ using the exact test, then with probability 0.99, $P < 0.018$ for the approximate test based on 1000 samples. Similarly, under the same specifications, if the exact P is <0.05, the approximate P is ≤ 0.066.

Finally, we note that we have presented only a brief sketch of Zerbe and Walker's procedure. As shown by Zerbe and Walker [12], it can be developed in analogy with the simple one-way analysis of variance, where the total sum of squared distances (from the overall mean curve) is partitioned into sums of squared distances between and within groups. In this formulation, Z, as given by Rao [3], corresponds to the within sum of squares. Since the total sum of squares is constant for all assignments, small values of Z correspond to large values of the usual F statistic, and point to rejection of the null hypothesis.

Acknowledgment

This work was supported by DE08730 from the National Institute of Dental Research.

Appendix 1

Computer implementation

A full set of PC programs for longitudinal data analysis, including these programs, can be obtained on high density 5.25-inch or 3.5-inch diskettes (please request type) by sending $25 to defray the cost of handling and licensing fees. These progams require an 80386- or 80486-based personal computer (PC) running the MS-DOS operating system (version 5.0 or higher is recommended, although versions as low as 3.3 will suffice). 80386 computers must also be equipped with an 80387 math coprocessor. At least 4 Mb of memory is required, and must be available to *GAUSS386i*, i.e. not in use by memory resident programs such as Windows. EGA or VGA graphic capabilities are required to display the color graphics; VGA or SVGA is suggested to display optimally the graphic results. Runtime modules are supplied with the programs so that no additional software (i.e. compiler or interpreter) is required to run these programs. One can create and edit ASCII data sets for use by these programs using the full screen editor supplied with MS-DOS version 5.0. The programs are written and compiled using *GAUSS-386i*, version 3.0, require no additional installation or modification, and are run with a single command. When requesting the programs, address inquiries to the corresponding author and make checks payable to Baylor College of Dentistry.

References

1 Potthoff RE and Roy SN: A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, 51 (1964) 313–326.

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

3 Rao CR: The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika*, 52 (1965) 447–458.

4 Khatri CG: A note on a MANOVA model applied to problems in growth curves. *Ann Inst Stat Math*, 18 (1966) 75–86.

5 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.

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

7 Ten Have TR, Kowalski CJ and Schneiderman ED: A PC 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, Willis SM, Ten Have TR and Kowalski CJ: Rao's polynomial growth curve model for unequal-time intervals: A menu-driven GAUSS program, *Int J Biomed Comput*, 29 (1991) 235–244.

9 Ten Have TR, Kowalski CJ, Schneiderman ED and Willis SM: A PC program for performing multigroup longitudinal comparisons using the Potthoff–Roy analysis and orthogonal polynomials, *Int J Biomed Comput*, 30 (1992) 103–112.

10 Ten Have TR, Kowalski CJ, Schneiderman ED and Willis SM: Two SAS programs for performing multigroup longitudinal analyses, *Am J Phys Anthropol*, 88 (1992) 251–254.

11 Schneiderman ED, Kowalski CJ and Willis SM: Regression imputation of missing values in longitudinal data sets, *Int J Biomed Comput*, 32 (1993) 121–133.

12 Zerbe GO and Walker SH: A randomization test for comparison of groups of growth curves with different polynomial design matrices, *Biometrics*, 33 (1977) 653–657.

13 Zerbe GO: Randomization analysis of the completely randomized design extended to growth and response curves, *J Am Stat Assoc*, 74 (1979) 215–221.

14 Foutz RV, Jensen DR and Anderson GW: Multiple comparisons in the randomization analysis of designed experiments with growth curve responses, *Biometrics*, 41 (1985) 29–37.

15 Edgington ES: *Randomization Tests*, 2nd Edn., Marcel Dekker, New York, 1987.

16 Noreen EW: *Computer Intensive Methods for Testing Hypotheses*, Wiley, New York, 1989.

17 Manly BFJ: *Randomization and Monte Carlo Method in Biology*, Chapman & Hall, London, 1991.

18 Schneiderman ED, Willis SM and Kowalski CJ: PC program for estimating polynomial growth, velocity, and acceleration curves when subjects may have differing times of measurement, *Int J Biomed Comput*, 33 (1993) 249–265.

19 Schneiderman ED, Willis SM and Kowalski CJ: Clustering on the basis of longitudinal data, *Comput Biol Med*, 23 (1993) 399–406.

20 Schneiderman ED, Willis SM, Kowalski CJ and Guo IY: A PC program for diagnosing abnormal growth, growth velocity and acceleration from longitudinal observations, *Int J Biomed Comput*, 35 (1994) 247–254.

21 Schneiderman ED, Willis SM, and Kowalski CJ: A PC program for classification into one of several groups on the basis of longitudinal data, *Int J Biomed Comput*, in press.

22 Brownlee KA: *Statistical Theory and Methodology in Science and Engineering*, Wiley, New York, 1960, p. 118.

23 Ingelfinger JA, Mosteller F, Thibodeau LA and Ware JA: *Biostatistics in Clinical Medicine*, Macmillan, New York, 1983, p. 291.

24 Remington RD and Schork MA: *Statistics with Applications to the Biological and Health Sciences*, Prentice-Hall, Englewood Cliffs NE, 1970, p. 392.

25 Schneiderman ED, Kowalski CJ and Ten Have TR: A GAUSS program for computing an index of tracking from longitudinal observations, *Am J Hum Biol*, 2 (1990) 475–490.

26 Bogin B, Sullivan T, Hauspie R and Mac Vean RB: Longitudinal growth in height, weight, and bone age of Guatemalan Ladino and Indian schoolchildren, *Am J Hum Biol*, 1 (1989) 103–113.