DER REACTION

Regression Models for Repeated Measurements

Sylvan Wallenstein

ion of Biostatistics, School of Public Health, 600 West 168th Street, New York, New York 10032, U.S.A.

Response to a Query, Aitkin (*Biometrics* **37,** 831–832, December 1981) uses d ANOVA methodology to analyze data from a trial in which rectal temperatures easured on a group of ten subjects, under four combinations of ambient tempera-ith six equally-spaced observations taken for each temperature. I feel that the tions needed for the 'standard' ANOVA are almost certainly not satisfied for this, parable data sets. To simplify the discussion, I assume that the ambient tempera-ere applied in random order. Under this condition, the tests for temperature and for the linear components of the time effect and interaction, do have the stated F tion if modified slightly, whereas the other tests do not have an F distribution H_0.

suggests a partition of the residual term to investigate the possibility that the ries between the subjects. The partition should be done even if the slope is not of since it permits analysis of a more general model which includes Subject × Time ject × Temperature interactions (Neter and Wasserman, 1974, p. 732). In addi-indicated below, it allows evaluation of the temperature effect, even if serial on due to time is present.

ugh the necessary and sufficient condition for validity of the model does not qual variances and covariances, it does imply that the covariances are equal if the are equal (Huynh and Feldt, 1970). Furthermore, Huynh (1978) states that 'it difficult to conceptualize a situation which would give rise to a Huynh–Feldt suggest that for all *practical* purposes the required assumption should be that of iances and equal covariances. This assumption is probably not satisfied for the of time in this data set, since the correlation between the 20- and 40-minute in rectal temperature is probably not the same as between 20 minutes and two ince the ambient temperatures were applied in random order, it would be le to assume that the correlations between temperatures are equal and thus MS ature)/MS (Subject × Temperature) has an F distribution (see also Mendoza, r and Cain, 1976). Since the test for trend in the time effect is based on a linear ion of the within-individual observations, MS {Time(L)}/MS {Subject × Time(L)} distribution.

(1978) cites various approaches for giving approximate critical values for the F ses where the Huynh–Feldt assumptions do not hold. If one can postulate that correlation between observations u time units apart is ρ^u, conservative critical e obtained by multiplying the nominal degrees of freedom by $5(t+1)/(2t^2+7)$ ein and Fleiss, 1979).

Huynh, H. and Feldt, L. S. (1970). Conditions under which mean square ratios in re measurement designs have exact *F*-distributions. *Journal of the American Statistical A* **65,** 1582–1589.

Huynh, H. (1978). Some approximate tests for repeat measurements designs. *Psychome* 161–175.

Mendoza, J. L., Toothaker, L. E. and Cain, B. R. (1976). Necessary and sufficient conditio ratios in the $L \times J \times K$ factorial design with two repeated factors. *Journal of American S Association* **71,** 992–993.

Neter, J. and Wasserman, W. (1974). *Applied Linear Statistical Models.* Homewood, Illinois:

Wallenstein, S. and Fleiss, J. L. (1979). Repeated measurements analysis of varianc the correlations have a certain pattern. *Psychometrika* **44,** 229–233.

F. Yates

Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, Engla

Murray Aitkin's Response to a Query on the use of regression in the analy experiments with repeated measurements (*Biometrics* **37**, 831–832, December requires some comment.

As a specific example the Query instanced 'an experiment in which rectal temper were measured on a group of 10 subjects at 20-minute intervals over exposure peri two hours, at four different ambient temperatures'. Aitkin states that experiments kind are called 'split-plot' experiments and refers to Cochran and Cox (1953) detailed discussion of their analysis. This is incorrect terminology. ment different treatments are applied to the subplots constituting a whole plo experiment is properly randomized these treatments are *randomly* assigned subplots. In an experiment with repeated measurements ('successive' might be a term) the measurements have a temporal sequence, with the consequence that me ments on the same subject separated by a small time interval will in general be highly correlated than those more widely separated. Cochran and Cox do not disc analysis of experiments with repeated measurements.

In the analysis of agronomic experiments it has long been recognised that measurements of two or more different quantities are made plot by plot, separate an of variance of the plot-by-plot values of each quantity give valid estimates of erro tests of significance for the treatment means for each quantity, but that if a funct these quantities is also of interest, separate values of the function must be calcula each plot to provide a basis for a further valid analysis. Experiments on sugar bee simple example. In these a sample of roots is taken from each plot to determine th by-plot values of the sugar percentage. These, multiplied by the plot-by-plot yi roots, give estimates of the plot-by-plot yields of sugar; it is these values that must for an analysis of variance relating to the effects of treatments on the sugar yiel

The same principle applies to repeated measurements. If, for example, the effec treatments on the rates of change are of interest, linear regressions can be fitte measurements on each experimental plot. An analysis of variance of the resultant regression coefficients can then be made. The subject is well discussed by and Walters (1976), with particular reference to agronomic experiments harve number of successive occasions, such as those on fruit and vegetables.

The present experiment is analogous to an ordinary randomized block experim four treatments (temperatures) in ten blocks (subjects) with repeated measureme

-plot experiment. The only difference is that in an agronomic experiment there plots in each block which are treated simultaneously, whereas in the present nt the *same* subject is subjected *successively* to the four temperatures. If the ures are applied in random order to each subject individually an ordinary zed block analysis of any function of the repeated measurements is valid.

mplify the experiment, all the subjects are placed simultaneously in a single large ure chamber, this randomization requirement cannot be met. The same analysis priate, but any residual effects of previous treatments will be confounded with the fects of the current treatment, instead of being included in experimental error. If a interval is allowed to elapse between treatment periods such residual effects are be unimportant; in other types of experiment, in which they are of importance, r designs, in which changes of treatment are incorporated in the design, with or repeated measurements on the same treatment, are available. The term 'repeated ements' must not be used to denote a crossover design.

present experiment the six values for any one subject exposed to a particular temperature can be fitted exactly by a fifth-degree polynomial. This can be nted by its mean and five orthogonal polynomials of degree one to five; that of one is the best-fitting linear regression, that of degree two the improvement in fit ding a quadratic term, etc. A separate analysis of variance (or a single analysis d in sections) can then be made on each of these functions. Because they are nal linear functions, the total sum of squares for the six analyses will equal the squares of deviations for all 240 observations, though orthogonality is not ry for separate analyses to be valid.

able below gives the degrees of freedom for these analyses. To save space, and for bility with Aitkin's table, the analyses for quadratic and higher terms are lumped . Also for comparability I have accepted Aitkin's supposition that only six values served in each test, though a wise experimenter would, I am sure, have included tional observation at the start of the test.

	Means over all times	Regressions on time: Linear	Regressions on time: Nonlinear
Mean	[cfm]	1	4
Subjects	9	9*	36*
Ambient temperatures	3	3	12
Subjects × Temperatures	27*	27*	108*

ems marked with an asterisk in my table are all included in the 'Residual within item (207 df) in Aitkin's table, and by implication the items not so marked can d for significance against this residual term. In fact, as explained above, items can properly tested against the 27 df for Subjects × Temperatures in the same column column being split into four columns for quadratic, cubic, etc. effects). Further- the subjects react differently to the different ambient temperatures, and if linear ons express these reactions 'adequately', the Subjects × Temperatures mean squares the first two columns of my table, and the Subjects mean squares of the second are likely to be substantially larger than the corresponding mean squares for the mn. This may well result in the values of the mean squares for the 4 df and 12 df nonlinear components of the means and temperatures not attaining significance, even at a low level, when tested against Aitkin's residual, thus giving a false impression the adequacy of the linear regressions. More importantly, the significance of the aver effects of the different ambient temperatures will be greatly exaggerated, and corresponding standard errors will be similarly underestimated.

Aitkin does point out that to investigate the possibility that the slope varies betw subjects his residual term could, 'if necessary', be subdivided into the asterisked term my table, but does not indicate to the reader that the Subjects linear term (9 df) should tested against the Subjects×Temperatures linear term (27 df), *not* the residual (108 df). It is of course true that the Subjects×Temperatures terms may contain gen repeatable subject differences as well as experimental error: subjects may react diff ently to different temperatures because of physiological differences, such as those to age or sex. An indication of such differences might be obtained by extracting linear component of Subjects×Temperature, or subdivision by, or covariance on, and sex.

Although an analysis of the above type, using orthogonal polynomials, is conceptu and computationally simple, it may not provide the best approach in specific problems. the present example a preliminary plot of the results (which should always be made) well indicate that changes in rectal temperatures over time are of the sigmoid typ asymptotic to an initial level (which may differ slightly from subject to subject), and asymptotic to a changed level dependent on the change in ambient temperature and the subject's physiological reactions. If this is so, the main quantities of interest probably be (i) the changes in level and (ii) the speed with which the new level is attained.

Evidence of sigmoid-type curves will indeed be provided by the existence of signific quadratic and cubic components when orthogonal polynomials are fitted as above, but differences of the maxima and minima of third degree polynomials are unlikely to satisfactory estimates of the changes in level. If a suitable computer program such as Rothamsted Maximum Likelihood Program (MLP) is available the ideal solution might to fit a Gompertz curve (which does not assume equal rates of convergence to the asymptotes) to each sequence of observations. An alternative and very simple proced is to estimate the change of level for each subject from the differences of the mean of last few observations (e.g. the last two) from the initial observation. An indicatio differences in rates of change between subjects will then be obtained from the ratios of means for the subjects over all times to the corresponding changes of level. When the sets of quantities have been calculated a separate analysis of variance can be made each set.

It may be objected that this latter procedure does not extract the full amount information available in the data. The information which is extracted however, is re intelligible, and the particular contrasts selected can be varied in the light of the form regularity of the observed curves. That a wise choice is capable of recovering most of available information is illustrated at an elementary level by the estimation of a regre coefficient from the difference of the means of the first and last two or three observa of a short series of equally-spaced observations. If the regression is truly linear and errors are independent, the loss of information for series of seven, nine and 11 obser tions and means of two observations are 11%, 18% and 26%; for means of three they 14%, 10% and 13%. Means of two are therefore a good choice for seven observ and means of three for nine and 11 observations.

At the end of his paper Aitkin makes reference to various involved multiva methods of analysis that have appeared in the literature. I will not comment on these except to say that I much doubt their relevance to real problems. His last parag reads 'Other complications, like serial correlation between plots in a block, have to ...ted by general maximum likelihood methods, though other simple approaches have suggested' is particularly revealing. Correlations of many kinds, both between ...sive observations on the same units and between simultaneous observations on ...ent units, are usually to be expected. An incorrect model will also introduce spurious ...tions between residuals from basically independent observations, as for example the true functions are second-degree polynominals but linear regression lines are ... It is, I believe, failure to appreciate these facts that is the main cause of Aitkin's ...ive analysis of variance.

REFERENCES

...n, W. G. and Cox, G. M. (1957). *Experimental Designs*, 2nd ed. New York: Wiley.
..., J. G. and Walters, D. E. (1976). Analysing data with repeated observations on each experimental unit. *Journal of Agricultural Science* **87**, 423–432.

...PONSE

...ief response to a general question necessarily omitted qualifying details. Readers ...ave asked for more information will welcome the detailed comments by Wallenstein ...ates. A clear and careful exposition of the analysis of repeated measures designs is ...by Bock (1975, Chapter 7), who gives also the general multivariate analysis ...ary when the covariance matrix of the repeated measures does not satisfy the ...ion for the validity of the standard analysis of variance. The necessary and sufficient ...ion is sphericity of the *differences* of pairs of the repeated measures: sphericity of ...peated measures themselves is sufficient, but not necessary. The general multivariate ...sis is valid, though not efficient, for particular patterns of covariance structure like ...correlation.

...k also gives the analysis for individual polynomial contrasts of the repeated ...ures.

...computing, the statistical package GENSTAT is particularly useful for split-plot ...s, as very simple declarations of the block, plot and treatment structure are ...ent to provide complete ANOVA breakdowns for generally balanced designs. For a ...l correlation structure in repeated-measures designs, multivariate analysis of var... programs are required. These are available in most statistical packages.

REFERENCE

...R. D. (1975). *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw–...ill.

Murray Aitkin
Centre for Applied Statistics,
Cartmel College,
University of Lancaster,
Lancaster LA1 4YL, England.