ANALYSIS OF REPEATED MEASURES DATA

Ramon C. Littell

Recommended Citation

ANALYSIS OF REPEATED MEASURES DATA

RAMON C. LITTELL
Department of Statistics
Institute of Food and Agricultural Sciences
University of Florida
Gainesville, FL 32611

ABSTRACT

Data with repeated measures occur frequently in agricultural research. This paper is a brief overview of statistical methods for repeated measures data. Statistical analysis of repeated measures data requires special attention due to the correlation structure, which may render standard analysis of variance techniques invalid. For balanced data, multivariate analysis of variance methods can be employed and adjustments can be applied to univariate methods, as means of accounting for the correlation structure. But these analysis of variance methods do not apply readily with unbalanced data, and they overlook the regression on time. Regression curves for treatment groups can be obtained by fitting a curve to each experimental unit, and then averaging the coefficients over the units. Treatment groups can be compared by applying univariate and multivariate methods to the group means of the coefficients. This approach does not require knowledge of the correlation structure of the repeated measures, and an approximate version of it can be applied with unbalanced data.

Key words: repeated measures, analysis of variance, regression, random coefficient model

1. INTRODUCTION

The expression "repeated measures data" refers to multiple measurements on the same unit. In the general sense, the term "unit" could refer to an experimental unit in a designed experiment, a sampling unit in a sample survey, or a subject in a retrospective study. In most situations, the repeated measures are over time, but they could be over space as well. The term "point" will be used to refer to a point in time or a point in space.

Data with repeated measures occur in virtually all fields of agricultural research. Perhaps the most frequent occurrence is in growth measurements of plants or animals over time. Other examples of repeated measures over time include crop yields from multiple harvests on experimental plots, daily milk yields from individual cows, and weekly livestock prices at individual auction markets. Examples of repeated measures in space are moisture determinations at several depths in soil core samples, measurements of a pollutant at numerous points on a line transect, and amounts of spray deposition at several sites within citrus trees.

For the sake of convention and uniformity, we shall use terminology for repeated measures in time. Also, we shall use the term "unit" to refer to the sampling unit or experimental unit, whichever the case may be.

Studies with repeated measures data have traditionally been analyzed as "split plot in time" experiments. This means that a split plot analysis of variance is performed with the "units" on which the repeated measures are taken treated as main-plot units and the "units" at particular points in time treated as a sub-plot units. This approach has many shortcomings. The repeated measures must be taken at the same time points on each unit, and there must be a certain correlation structure between the repeated measures on the same unit in order for analysis of variance F statistics to be valid. Also, trends over time are commonly ignored when the "split plot in time" approach is used. In recent years, many alternative statistical methods have been developed.

This paper provides a brief survey of several methods for analysis of repeated measures data. It draws from recent work of several authors and the facilities of certain computer program

packages. Section 2 discusses analysis of variance of "balanced" repeated measures, in which all units have complete data at a specified set of points. Methods for this case are well developed and several program packages provide adequate computing facility for these methods. Disagreement exists, however, on the merits of various methods, and when one is preferred to another. Section 3 discusses a regression approach to repeated measures data in the balanced case, treating time as a regression variable. Section 4 discusses analysis of "unbalanced" repeated measures data, in which case different units have data at different points. Methods of analysis for this situation are much more complicated, and generally require some degree of approximation.

2. ANALYSIS OF VARIANCE FOR BALANCED REPEATED MEASURES DATA

Repeated measures data will be called "balanced" if every unit has complete data at the same time points. For balanced repeated measures data, the traditional method is the split-plot in time analysis of variance. This is also called the "univariate" repeated measures analysis of variance.

An example of balanced repeated measures data is given in Freund, Littell and Spector (1986). These data came from a large study of the effects of nutrition and exercise on physical strengths of geriatric citizens. Concern here is with the exercise aspects of the study, which involved three weight training programs to which subjects were randomly assigned. Each subject's strength, as determined by the amount of weight the subject could lift, was measured every other day for two weeks. The first program was a control, in which no training was .employed (CONT). The second program utilized a weight training system in which the number of repetitions of the exercise was increased incrementally with time (RI). In the third program, the amount of weight was increased incrementally over time (WI). Data are shown in Table 1, with means and standard deviations in Table 2. Mean profiles for the three programs are plotted in Figure 1.

The exercise therapy study is typical of many repeated measures designs in which subjects are randomly assigned to "treatment" groups, and a response variable is measured repeatedly over time.

A model for data from this type of study is

$$y_{ijk} = \mu_{ik} + e_{ijk},$$

where y_{ijk} is the response of the jth subject in the ith treatment at the kth time, and μ_{ik} is the population mean for treatment i at time k. The errors e_{ijk} are assumed normally distributed with mean zero and $V(e_{ij}) = V$, where $e_{ij}=(e_{ij1},...,e_{ijt})'$.

A key issue with repeated measures data is the structure of the covariance matrix V. Here are five particular structures in terms of mathematical conditions on $\sigma_{kk'}$, the element in row k, column k', of V, that play a role in repeated measures data:

Structure		Mathematical Condition
1. Unstructured	no restrictions	$V = (\sigma_{kk'})$ positive definite
2. Spherical	equal variances zero covariances	$\sigma_{kk} = \sigma^2$, $\sigma_{kk'} = 0$
3. Compound symmetric	equal variances equal covariances	$\sigma_{kk} = \sigma^2$, $\sigma_{kk'} = \delta\sigma^2$
4. Huynh-Feldt	unrestricted variances restricted covariance	$\sigma_{kk} = 2\tau_k+\phi$ $\sigma_{kk'} = \tau_k+\tau_{k'}$
5. Autoregressive	covariance function of time interval between repeated measures	$\sigma_{kk'} = \Theta_{|k-k'|}$

A "univariate" analysis of variance (split plot in time analysis of variance) for the exercise therapy data is shown in Table 3. The asterisk next to the P-values for Time and Time*Program F values is a flag that the validity of these F tests depends upon V having a certain mathematical structure. Many sources state that the required condition is compound symmetry, but in fact it is the Huynh-Feldt (H-F) condition. See Huynh and Feldt (1970). However, there is probably little practical distinction between the two conditions. The H-F condition is equivalent to any set of orthonormal contrasts of the repeated measures having a spherical distribution.

Both the compound symmetry and H-F conditions are questionable from a practical point of view. There is little evidence to indicate that either of these conditions should occur in nature. A more realistic condition for repeated measures in time is the first order autoregressive structure. It accommodates the feature that measures close in time will be more highly correlated than measures far apart in time; in particular, covariance between two repeated measures is a function only of their distance apart in time.

Multivariate methods can be used to test the Time and Time*Treatment effects which do not require any particular structure of V. These can be constructed by computing sets of contrasts among the repeated measures and applying multivariate analysis of variance techniques (Cole and Grizzle, 1966). Many of the multivariate tests can also be obtained directly by some computer programs; in particular, the REPEATED statement in the ANOVA and GLM procedures in the SAS System. The generality of the multivariate methods tends to come at a cost in statistical power. For the exercise therapy study, multivariate tests (based upon Pillai's trace) produce the results in Table 4.

It is not uncommon to obtain very different results regarding Time*Treatment interaction from the multivariate test and the univariate test. The univariate test in Table 3 produced P=0.0005 and the multivariate test in Table 4 produced P=0.1943. One must know which, if either, of these P-values to believe. A general approach for analysis of variance of repeated measures data is to use the univariate tests if the H-F condition holds, and use the multivariate tests if the H-F condition does not hold. The problem is in knowing whether the H-F condition holds.

Tests of significance have been recommended to determine if the H-F condition holds; i.e. to determine if the mathematical conditions are met which are required for validity of the univariate analysis of variance F statistics for the Time and Time*Treatment interaction. One of these tests, which is available in the SAS System under the name "Mauchly's criterion," is a likelihood ratio test for the null hypothesis that the H-F condition holds versus the alternative hypothesis that V is unstructured. Tests for the validity of the H-F condition have recently been discredited, however. (See Looney and Stanley, 1989). This leaves this general approach of choosing between the univariate and multivariate tests in a state of limbo.

Alternatively, adjusted versions of the univariate tests can be employed. These procedures entail adjusting the significance probability of the univariate analysis of variance F tests according to the degree of departure from the H-F conditions. Box (1954) proposed a measure ϵ of the degree of departure from the H-F conditions. The value of ϵ is between $(t-1)^{-1}$ and 1, where t is the number of time points. The value $\epsilon=1$ indicates the H-F condition holds, while decreasing values of ϵ indicate increasing departure from the H-F condition. Box (1954) showed that significance probabilities of the univariate F tests for the Time and Time*Treatment effects can be adjusted by discounting the numerator and denominator degrees of freedom by multiplying them by ϵ. Thus $\epsilon=1$ would produce no adjustment of the P-value. On the other extreme, $\epsilon=(t-1)^{-1}$ would produce the most severe adjustment, reducing the degrees of freedom for time from t-1 to 1, for Time*Treatment from $(t-1)*(a-1)$ to a-1 (where a is the number of treatment groups), and for Error from $(n-a)*(t-1)$ to n-a (where n is the number of units altogether). However, in practice one would not usually know the value of ϵ. A conservative adjustment is to adjust the significance probabilities as if $\epsilon=(t-1)^{-1}$. Another approach is to estimate ϵ from the data. Greenhouse and Geisser (1959) and Huynh and Feldt (1970) proposed estimates of ϵ. See Milliken and Johnson (1984) or Milliken (1990) for good discussions of these methods. Looney and Stanley (1989) make recommendations concerning a combination of the adjusted, unadjusted and multivariate tests for the Time and Time*Treatment effects.

· Univariate, multivariate and adjusted univariate P-values for the exercise therapy data are summarized in Table 5. All three tests for the Time effect are significant with P < 0.0001. However, this is the "main effect" of Time; that is, the effect of Time averaged across programs. The plot in Figure 2 suggests that the Time trends are not parallel. This is to be expected because RI and WI are active programs and should produce strength measurements which increase with Time, whereas CONT is inactive and should not result in strength increasing with Time. Therefore, the significance of the Time trend averaged across Programs is of little practical value. The nature of the Time*Program interaction is a more important issue. Recall that the univariate test for Time*Program has a P-value of 0.0005 and the multivariate test for Time*Program has a P-value of 0.1943, giving conflicting conclusions from the two tests. Mauchly's test for the validity of the H-F condition yielded P < 0.0001, which indicates the H-F condition does not hold. Therefore, the univariate P-value for Time*Program is unreliable. On the other hand, the multivariate test lacks the power to detect the Time*Program interaction that seems apparent in Figure 1. The Greenhouse-Geisser (1959) estimate of the discounting factor ϵ is 0.4233. When the degrees of freedom of the univariate test are multiplied by 0.4233, the resulting adjusted P-value is 0.0130. This P-value seems in agreement with the visual impression off interaction in Figure 2.

3. REGRESSION METHODS FOR BALANCED REPEATED MEASURES DATA

All of the analysis of variance methods discussed in the preceding section regarding tests of significance for the Time and Time*Treatment effects ignore a basic and essential feature of the variable Time; namely, that it is quantitative. Therefore, as in the case of any quantitative variable, a more effective method of analysis is usually obtained by employing some form of regression methodology. That is, one should think in terms of fitting curves to the Time variable. Typically, one would potentially have a different regression curve for each treatment (in this case, each Program).

The regression approach has many advantages. First, and perhaps most importantly, it directs the data analysis toward a major objective of the study by focusing on trends over time. Second, it provides a reduction in the number of parameters to be estimated; instead of the t means $\mu_{i1},...,\mu_{it}$ for the ith treatment, one would have only the parameters of the mean regression curve, typically two or three in number. Third, inference can be made in terms of the regression parameters. Hopefully, regression functions could be specified so that the parameters have some practical interpretation. Fourth, regression curves "smooth" the response over time. Estimates of differences between treatments can be calculated from fitted curves that will avoid the erratic conclusions often obtained from comparing treatment means at each time point. It is not uncommon, when estimating differences between treatment means at individual time points, to find the differences vacillate between "significant" and "not significant" from one time to the next due to the random variation in the means.

Regression curves can be fitted to individual units, providing estimates of the regression parameters for each unit. Then multivariate and univariate statistical analysis of the unit regression parameter estimates can be performed to determine in what manner characteristics of the regression curves are affected by the treatment groups. In terms of a model, we have

$$\mathbf{y}_{ij} = \mathbf{X}\beta_i + \mathbf{e}_{ij},$$

where $\mathbf{y}_{ij}$ is the vector of repeated measures for the jth unit in the ith treatment group, $\mathbf{X}$ is the matrix relating the repeated measures to the time points, β_i is the parameter vector for the ith treatment, and $\mathbf{e}_{ij}$ is the error vector for the jth unit in the ith group. Fitting the model to each unit via ordinary least squares (OLS) yields

$$\mathbf{b}_{ij} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{ij}$$

as the vector of parameter estimates for unit j in group i. Then $\mathbf{b}_{ij}$ has expectation $E(\mathbf{b}_{ij}) = \beta_i$, and

variance $V(\mathbf{b}_{ij}) = (\mathbf{X'X})^{-1}\mathbf{X'VX}(\mathbf{X'X})^{-1}$. Therefore, the $\mathbf{b}_{ij}$ estimates follow a multivariate normal distribution, which we write

$$\mathbf{b}_{ij} \sim \text{MVN}(\,\beta_i,\,(\mathbf{X'X})^{-1}\mathbf{X'VX}(\mathbf{X'X})^{-1}).$$

It follows that standard methods for the general multivariate linear model can be applied. (See Grizzle and Allen, 1969). This provides the basis for a seemingly *ad hoc* but effective methodology for inference. Standard multivariate analysis computer programs can be used to perform the analyses, taking the vector $\mathbf{b}_{ij}$ of parameter estimates for the jth unit in the ith group as the "observation" vector. In fact, as will be pointed out in the next section, there is some basis for using this approach even in the unbalanced case. A tremendous advantage of this *ad hoc* approach is that it does not require knowledge of $\mathbf{V}$. Therefore, questions of the specific form for $\mathbf{V}$ (compound symmetry, H-F condition, etc.) are not issues. The cost of not incorporating knowledge of $\mathbf{V}$ is that the OLS estimate $\mathbf{b}_{ij}$ is not necessarily the best estimate of the respective "true" (random) parameter vector β_{ij} for the jth unit in group i. This often is a negligible cost because ordinarily the between unit variance is larger than within unit variances.

In the case of polynomial regression curves, one could fit successively linear, quadratic, and cubic curves to individual subjects. (Fitting polynomials above degree three rarely produces useful parameter estimates. Even results from fitted third degree polynomials are often of questionable merit.) Analysis of variance can be applied to the subject parameter estimates, typically beginning with the higher order terms, to ascertain the degree of polynomial required. This step is followed by fitting curves of the determined order, and drawing inferences about the treatment groups on the basis of these curves.

Results extracted from using the REPEATED statement in SAS PROC GLM, which provides tests for the significance of parameter estimates from sequentially fitted polynomials, are shown in Table 6. (Tests for higher order effects, which were all nonsignificant, were also printed by the GLM procedure.) These results indicate quadratic models are adequate, with a hint of lack of fit suggested by the average Time Cubic parameter being significantly different from zero at the 0.1163 level. Nonsignificance of the Time Quad*Prog interaction suggests the quadratic models for the three Programs might be forced to have the same quadratic estimate. But this would probably be counterproductive because one would expect the response to the CONT Program to be flat. The Time Linear*Program interaction is significant at the 0.0110 level. This implies that, if straight lines were fitted for the three Programs, they would have significantly different slopes. The significance of the Time Linear effect ($p = 0.0028$) implies the average of the slopes is different from zero. As noted earlier, however, the average trend is probably not of great practical value because the slopes differ between the programs.

Having decided upon the degree of the model, the next step is to obtain equations for fitted models. Unfortunately, the GLM procedure in SAS does not provide parameter estimates directly from this analysis. But parameter estimates for the treatment curves can be obtained by fitting a curve for each subject and then averaging parameter estimates across subjects in the treatment groups. Coefficients for equations so obtained for quadratic curves fitted to the weight training data are shown in Table 7, along with standard deviations of the parameter estimates.

Graphs of the average quadratic curves are plotted in Figure 2, with Program means plotted around them. Predicted values at Times 1-7 from the fitted curves with standard deviations are shown in Table 8. These predicted values were calculated as averages of predicted values from the individual subject curves. It is usually advantageous to make comparisons between the treatment group at individual Times using these average predicted values rather than using the ordinary means because the predicted values have been smoothed by the fitted curve.

4. ANALYSIS OF UNBALANCED REPEATED MEASURES DATA

Unbalanced data occur in repeated measures studies in a variety of ways. Study designs ordinarily call for all subjects to provide measures at the same time points, but complications in the data collection process often cause some subjects to not provide data at some points. In other

studies it might not be feasible to even plan for all subjects to provide data at a common set of time points.

Analysis of variance methods, as generalizations of the split plot in time approach, are discussed by Milliken and Johnson (1984). Applications of these methods usually assume an unbalanced data analog of compound symmetry, because the H-F condition is too complicated to formulate with unbalanced data. With this assumption, analysis of variance of unbalanced repeated measures data falls in the general category of unbalanced mixed linear models. This encompasses a very large and complicated set of problems for data analysis, more than can be discussed here. We therefore refer the reader to Milliken and Johnson. Instead, we focus on procedures specific to repeated measures data.

In order to accommodate unbalanced data, we specify the model

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\beta_i + \mathbf{e}_{ij},$$

where $\mathbf{y}_{ij}$ is the vector of repeated measures and $\mathbf{X}_{ij}$ is the matrix relating the repeated measures to the time points for the jth unit in the ith treatment group, β_i is the parameter vector for the ith treatment, and $\mathbf{e}_{ij}$ is the error vector for the jth unit in the ith group. We must now specify possibly different covariance matrices for each subject, $V(\mathbf{e}_{ij}) = \mathbf{V}_{ij}$. This is necessary to reflect the dependence of the covariance matrix on the number and spacing of the repeated measures. Fitting the model to each unit via OLS yields

$$\mathbf{b}_{ij} = (\mathbf{X}_{ij}'\mathbf{X}_{ij})^{-1}\mathbf{X}_{ij}'\mathbf{y}_{ij}$$

as the vector of parameter estimates for unit j in group i. Then $\mathbf{b}_{ij}$ has expectation $E(\mathbf{b}_{ij}) = \beta_i$, and variance $V(\mathbf{b}_{ij}) = (\mathbf{X}_{ij}'\mathbf{X}_{ij})^{-1}\mathbf{X}_{ij}'\mathbf{V}_{ij}\mathbf{X}_{ij}(\mathbf{X}_{ij}'\mathbf{X}_{ij})^{-1}$. It follows that the group sample means of the unit estimates are unbiased estimates of the group population mean; i.e. $E(\mathbf{b}_{i\cdot}) = \beta_i$, where the *ad hoc* estimators $\mathbf{b}_{i\cdot} = n_i^{-1} \Sigma_j(\mathbf{b}_{ij})$.

The model just described is closely related to the so-called *random coefficient regression* (RCR) model. Following the description of Vonesh and Carter (1987), $\mathbf{y}_{ij} = \mathbf{X}_{ij}\beta_{ij} + \epsilon_{ij}$, where β_{ij} is assumed normally distributed with mean β_i and covariance Γ, and ϵ_{ij} is normally distributed with mean zero and covariance matrix $\sigma^2 I$ of appropriate dimension. This makes $\mathbf{V}_{ij} = \mathbf{X}_{ij}\Gamma\mathbf{X}_{ij}' + \sigma^2 I$. The RCR model is sometimes criticized for the property that $\mathbf{V}_{ij}$ depends upon $\mathbf{X}_{ij}$ in this manner. See Ware (1985). The RCR model is fitted in stages: First, $\mathbf{b}_{ij}$ is computed as defined above for each subject. Next, an estimator of the covariance matrix of the (combined) vector of regression coefficients for all the subjects is obtained. Then (estimated) generalized least squares is applied to obtain estimates of the β_i vectors, along with estimates of their covariances.

Gumpertz and Pantula (1989) proposed using the *ad hoc* estimators $\mathbf{b}_{i\cdot}$. They showed that these estimators have near optimal properties even in unbalanced cases. They also showed that the covariance matrix $n^{-1}S_i$ is an unbiased estimate of $V(\mathbf{b}_{i\cdot})$, where $S_i = (n_i-1)^{-1} \Sigma_j(\mathbf{b}_{ij}-\mathbf{b}_{i\cdot})(\mathbf{b}_{ij}-\mathbf{b}_{i\cdot})'$. Although they stated and proved the covariance result in the context of the random coefficient regression model, it holds for general $\mathbf{V}_{ij}$. This provides a basis for approximate inference using the *ad hoc* $\mathbf{b}_{i\cdot}$ estimators even with unbalanced repeated measures data. Further, the Gumpertz-Pantula covariance result holds for certain other experiment designs.

Considerable effort has been expended in recent years to obtain maximum likelihood estimators for repeated measure models with unbalanced data and structured covariance matrices. See Ware (1985) for a review. Some of these methods have been implemented in the computer program BMDP 5V, employing estimation techniques of Jennrich and Schlucter (1984). The BMDP 5V setup assumes that all units potentially had data at the same points (i.e. the data were potentially balanced) but some subjects have missing data at some points. The primary limiting

aspect of this formulation is that relatively few different time points are allowed. A variety of covariance structures are permitted, including compound symmetry, 1st order autoregressive, unstructured, and the random coefficient regression model covariance (which is called random effects covariance). In addition to parameter estimates for the fitted models, BMDP 5V also prints estimates to the covariance matrix under the various assumed structures.

Data from the weight training study were used to make a preliminary comparison of estimates using the *ad hoc* approach with maximum likelihood from BMDP 5V assuming 1st order autoregressive errors. The average regression line $E(y) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$ (across programs) was estimated using each of the methods. (The choice to estimate the average regression line was made because BMDP 5V estimates this directly, along with estimates of differences of individual program parameters from the average.) This was done for the entire data set, and then data were discarded in various patterns: 20%, 40% and 60% were discarded (in sequence) at random in the Subject by Time array. Then 20%, 40% and 60% were discarded randomly across subjects, but with the condition that all missing points are at the end of the series. Thus some subjects have complete data, some are missing the last time, some are missing the last two times, etc. Estimates and standard errors appear in Table 9.

Results in this Table 9 indicate that the ad hoc method performs well relative to ML unless a large fraction of data is missing, on the order of 50% or more. Performance of the ad hoc method is especially poor when a large amount of data is missing at the last end of the series.

REFERENCES

BOX, G.E.P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems," *Annals of Mathematical Statistics*, 25, 290-302.

COLE, J.W.L. AND GRIZZLE, J.E. (1966), "Applications of Multivariate Analysis of Variance to Repeated Measurements Experiments," *Biometrics*, 22, 810-828.

FREUND, R.J., LITTELL, R.C. AND SPECTOR, P.C. (1986), *SAS System for Linear Models*, Cary, NC, SAS Institute Inc.

GREENHOUSE, S.W. AND GEISSER, S. (1959), "On Methods in the Analysis of Profile Data," *Psychometrika*, 24, 95-112.

GRIZZLE, J.E. AND ALLEN, D.M. (1969), "Analysis of Growth and Dose Response Curves," *Biometrics*, 25, 357-381.

GUMPERTZ, M. AND PANTULA, S.G. (1989), "A Simple Approach to Inference in Random Coefficient Models," *The American Statistician*, 42, 203-210.

HUYNH, H. AND FELDT, L.S. (1970), "Conditions under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions," *Journal of the American Statistical Association*, 65, 1582-1589.

JENNRICH, R.I. AND SCHLUCTER, M.D. (1986), "Unbalanced Repeated Measures Models with Structured Covariance Matrices," *Biometrics*, 42, 805-820.

LOONEY, S.W. AND STANLEY, W.B. (1989), "Exploratory Repeated Measures Analysis for Two or More Groups," *The American Statistician*, 43, 220-225.

MILLIKEN, G.A. (1990), "Analysis of Repeated-Measures Designs," in *Statistical Methodology in the Pharmaceutical Sciences*, D.A. Berry, ed., New York: Marcel Dekker, Inc., 83-116.

MILLIKEN, G.A. AND JOHNSON, D.E. (1984), *Analysis of Messy Data*, New York: Van Nostrand Reinhold Company.

WARE, J.H. (1985), "Linear Models for the Analysis of Longitudinal Studies," *The American Statistician*, 39, 95-101.

Table 1. Data for Exercise Therapy Study

OBS	PROGRAM	S1	S2	S3	S4	S5	S6	S7
1	CONT	85	85	86	85	87	86	87
2	CONT	80	79	79	78	78	79	78
3	CONT	78	77	77	77	76	76	77
4	CONT	84	84	85	84	83	84	85
5	CONT	80	81	80	80	79	79	80
6	CONT	76	78	77	78	78	77	74
7	CONT	79	79	80	79	80	79	81
8	CONT	76	76	76	75	75	74	74
9	CONT	77	78	78	80	80	81	80
10	CONT	79	79	79	79	77	78	79
11	CONT	81	81	80	80	80	81	82
12	CONT	77	76	77	78	77	77	77
13	CONT	82	83	83	83	84	83	83
14	CONT	84	84	83	82	81	79	78
15	CONT	79	81	81	82	82	82	80
16	CONT	79	79	78	77	77	78	78
17	CONT	83	82	83	85	84	83	82
18	CONT	78	78	79	79	78	77	77
19	CONT	80	80	79	79	80	80	80
20	CONT	78	79	80	81	80	79	80
1	RI	79	79	79	80	80	78	80
2	RI	83	83	85	85	86	87	87
3	RI	81	83	82	82	83	83	82
4	RI	81	81	81	82	82	83	81
5	RI	80	81	82	82	82	84	86
6	RI	76	76	76	76	76	76	75
7	RI	81	84	83	83	85	85	85
8	RI	77	78	79	79	81	82	81
9	RI	84	85	87	89	88	85	86
10	RI	74	75	78	78	79	78	78
11	RI	76	77	77	77	77	76	76
12	RI	84	84	86	85	86	86	86
13	RI	79	80	79	80	80	82	82
14	RI	78	78	77	76	75	75	76
15	RI	78	80	77	77	75	75	75
16	RI	84	85	85	85	85	83	82
1	WI	84	85	84	83	83	83	84
2	WI	74	75	75	76	75	76	76
3	WI	83	84	82	81	83	83	82
4	WI	86	87	87	87	87	87	86
5	WI	82	83	84	85	84	85	86
6	WI	79	80	79	79	80	79	80
7	WI	79	79	79	81	81	83	83
8	WI	87	89	91	90	91	92	92
9	WI	81	81	81	82	82	83	83
10	WI	82	82	82	84	86	85	87
11	WI	79	79	80	81	81	81	81
12	WI	79	80	81	82	83	82	82
13	WI	83	84	84	84	84	83	83

Table 2. Means and Standard Deviations for Exercise Therapy Data

PROGRAM		STR1	STR2	STR3	STR4	STR5	STR6	STR7
CONT	mean	79.8	80.0	80.0	80.1	79.8	79.6	79.6
	std dev	2.67	2.63	2.75	2.74	3.04	2.95	3.25
RI	mean	79.7	80.6	80.8	81.0	81.25	81.13	81.13
	std dev	3.11	3.22	3.58	3.79	4.12	4.13	4.18
WI	mean	81.0	81.7	81.9	82.5	82.6	82.7	83.1
	std dev	3.11	3.35	3.48	3.04	3.34	3.27	3.35

Table 3. Univariate Analysis of Variance for Weight Training Data

Source of Variation	df	Mean Square	Univariate	
			F	P
Program	2	209.72	3.07	0.0548
Error(a) = Subj.(Program)	54	86.42		
Time	6	8.89	7.43	0.0001*
Program*Time	12	3.58	2.99	0.0005*
Error(b) = Subj*Time(Prog)	324	1.20		

Table 4. Multivariate Tests (Pillai's Trace) for Weight Training Data

Effect	Multivariate F	Significance Probability P
Time	6.46	0.0001
Time*Program	1.37	0.1943

Table 5. Summary of Univariate, Multivariate and Adjusted Univariate P-values

Effect	Univariate	Multivariate	Adjusted Univariate
Time	0.0001	0.0001	0.0001
Time*Program	0.0005	0.1943	0.0130

Table 6. Analysis of Variance of Polynomial Trends

Source of Variation	df	Mean Square	F	p
Time Linear	1	40.514	9.85	0.0028
Time Lin*Prog	2	20.196	4.91	0.0110
Error	54	4.115		
Time Quadratic	1	10.577	8.64	0.0048
Time Quad*Prog	2	0.712	0.58	0.5626
Error	54	1.225		
Time Cubic	1	1.313	2.55	0.1163
Time Cub*Prog	2	0.020	0.04	0.9620
Error	54	0.516		

Table 7. Average Parameter Estimates and Standard Deviations

PROGRAM	INTERCEPT	LINEAR	QUADRATIC
CONT	79.61(3.13)	0.223(1.153)	-0.034(0.137)
RI	79.08(3.14)	0.793(1.192)	-0.073(0.134)
WI	80.52(3.39)	0.601(0.813)	-0.035(0.089)

Table 8. Estimates and Contrasts of Quadratic Curves

PROGRAM		STR1	STR2	STR3	STR4	STR5	STR6	STR7
CONT	mean	79.8	79.9	80.0	80.0	79.9	79.7	79.5
	std dev	2.70	2.60	2.67	2.77	2.85	2.96	3.21
RI	mean	79.8	80.4	80.8	81.1	81.2	81.2	81.1
	std dev	3.08	3.27	3.53	3.56	3.92	4.05	4.23
WI	mean	81.1	81.6	82.0	82.4	82.7	82.9	83.1
	std dev	3.22	3.19	3.21	3.24	3.25	3.26	3.31
CONTRAST								
CONT VS TRT	P-value	0.44	0.21	0.11	0.06	0.03	0.02	0.01
RI VS WI	P-value	0.20	0.23	0.25	0.24	0.20	0.15	0.10

Table 9. Comparison of Maximum Likelihood and *ad hoc* Parameter Estimates

Missing data pattern	Estimation Method	α_0	α_1	α_2
Complete	ML	79.66(0.46)	0.55(0.14)	-0.05(0.02)
	ad hoc	79.74(0.43)	0.54(0.14)	-0.05(0.02)
random 20%	ML	79.59(0.48)	0.63(0.16)	-0.06(0.02)
	ad hoc	79.68(0.45)	0.65(0.19)	-0.07(0.02)
random 40%	ML	79.67(0.05)	0.66(0.17)	-0.06(0.02)
	ad hoc	80.11(0.65)	0.54(0.23)	-0.05(0.03)
random 60%	ML	79.69(0.55)	0.59(0.21)	-0.05(0.03)
	ad hoc	79.78(0.83)	0.73(0.33)	-0.07(0.02)
last 20%	ML	79.61(0.46)	0.61(0.15)	-0.06(0.02)
	ad hoc	79.61(0.42)	0.65(0.17)	-0.07(0.02)
last 40%	ML	79.58(0.47)	0.70(0.19)	-0.08(0.03)
	ad hoc	79.63(0.44)	0.74(0.20)	-0.09(0.04)
last 60%	ML	79.54(0.49)	0.74(0.28)	-0.08(0.06)
	ad hoc	79.22(0.67)	1.34(0.44)	-0.22(0.12)

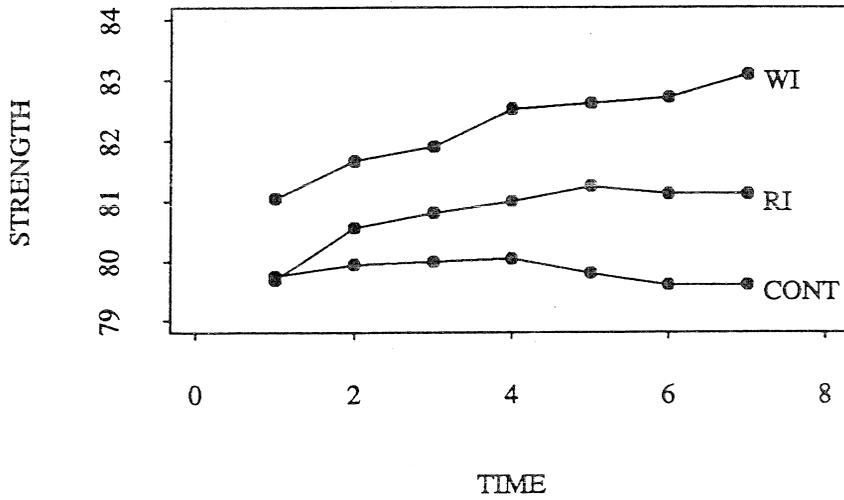Figure 1. Means for Exercise Therapy Programs

Figure 2. Quadratic Regression Curves for Exercise Therapy Programs