Exploratory Data Graphics for Repeated Measures Data

ROBERT WEISS*

Department of Biostatistics

UCLA School of Public Health

Los Angeles CA 90024-1772 U.S.A.

January 29, 1997

Abstract

Repeated measures (RM) is a common data structure in many fields. They are a special form of multivariate data that makes multivariate graphics both practical and useful. A set of basic graphics for RM data is introduced in the context of small to moderately sized balanced data sets where a random effects model is under tentative consideration. The graphics emphasize looking at as much of the data as possible. Graphical considerations in constructing the plots are discussed. Graphics for checking the mean structure, variance structure and relationship to covariates are mentioned. An approach to modeling is introduced.

Key Words: Covariance Selection; Hierarchical Models; Longitudinal Data; Model Specification; Random Effects Models.

1   Introduction

Repeated measures (RM) data are multivariate observations where each case $Y_i = (y_{i1}, \ldots, y_{in_i})$ consists of repeatedly measuring the same quantity such

as IQ score or blood pressure over time on a single subject. Observations y_{ij} are taken at times t_{ij} with $t_i = (t_{i1}, \ldots, t_{in_i})^t$. Graphics for general multivariate data are difficult to construct and interpret because comparisons between such as $y_{i1} - y_{i2}$ between different variables are not interpretable, only differences $y_{i1} - y_{j1}$ and relative differences (slopes) are interpretable. For RM data even simple comparisons such as $y_{i1} - y_{i2}$ and $y_{i1} - y_{j1}$ are both interesting and useful. The graphics should make these comparisons as easy as possible.

RM data have a hierarchical structure, with time an essential component of the data. Compared with regression data, this structure means that more graphics and more complicated graphics are needed to understand the basic structure of the data. Additionally, common RM models, such as random effects (RE) models (Laird and Ware 1982) fit to RM data contain several layers of assumptions, and these assumptions must be checked: each layer requires its own graphics. The major features of RM data are the mean of Y_i over time, the covariance structure of Y_i, and the relationship between covariates and the mean and covariance. The ED graphics in this paper are aimed at understanding these features of the data. In contrast, most graphics are developed as inferential tools, not for data analysis. That is, most graphics display the conclusions of an analysis, and are not geared towards developing the conclusions. Generally, diagnostic plots may be designed to display one particular piece of information well, or they can be omnibus capable of showing a range of problems with the data, but perhaps not quite as clearly as a specifically targeted plot. In this paper, I concentrate on omnibus graphics: these plots contain as much of the data as possible, and they permit the discovery of unanticipated structure. The purpose of the plots in this paper are to permit graphical analysis of repeated measures data. The goal is to prevent major blunders in the analysis: no claim is made for optimality. Rather the discussion is about what plots might usefully replace, surplant or extend the familiar scatterplot of linear regression data.

Limited work has been done on graphics for repeated measures data. The parallel plot or profile plot is a standard tool for RM data (Weiss and Lazaro 1992). For each case i, plot y_{ij} against t_{ij}, and connect the points at times $t_{i(j-1)}$ and t_{ij}. Do not connect the dots between other times or between cases. Examples are given in Figures 1, 3 and 4. Weiss and Lazaro (1992) discuss the interpretation of these plots and examples of parallel plots can be found in Crowder and Hand (1990), Gregoire, Schabenberger and Barrett (1993), Izenman and Williams (1989), Draper (1986) and Lange and Laird (1986) among many other places. Plots of random effects can be found in Waternaux, Laird and Ware (1989); and in Lange, Little, and Taylor (1990). Jones and Rice (1992) plot a small selection of profiles and Segal (1994) extends regression trees to repeated measures and plots profiles for sets of cases with similar profiles in separate plots. Weighted Quantile-Quantile (QQ) plots for checking the normality of the random effects are analyzed in Dempster and Ryan (1986) and Lange and Ryan (1989).

The paper is organized into several sections. The next section discusses the random effects model, estimates, residuals, modeling, and goodness of fit. The examples used in the paper are introduced in section 3. Sections 4 though 6 discuss plots. Section 4 contains a discussion of how to draw parallel plots when over-plotting is of concern, section 5 discusses plots involving covariates; and section 6 discusses direct checking of the variance-covariance specification. The paper ends with discussion in section 7.

2 Repeated Measures Models, Estimation and Residuals

2.1 Models

A class of models is needed to guide interpretation and construction of plots. RE models are used as tentative candidate models particularly in developing residuals, but consideration cannot and should not be limited to those

models. The basic repeated measures random effects model is

$$
\begin{aligned}
Y_i &= X_i\alpha + Z_i\beta_i + \epsilon_i \\
\beta_i &\sim N(0, D), \\
\epsilon_i &\sim N(0, \sigma^2 I),
\end{aligned}
\tag{1}
$$

for $i = 1, ..n$; where $Y_i = (y_{i1}, \ldots, y_{in_i})^t$ is the n_i by 1 vector of repeated measures on subject i taken at times $t_i = (t_{i1}, \ldots, t_{in_i})^t$; X_i, n_i by p, and Z_i, n_i by q are known vectors of covariates; α is a p by 1 parameter vector of fixed effects, β_i is a q by 1 parameter vector of random effects with prior mean 0 and prior covariance matrix D, and σ^2 is the sampling variance from trial to trial. Very commonly, Z_i has columns which are polynomials in t_i, $q < p$, and X_i has a submatrix of q columns equal to Z_i. The entire vector Y_i will be referred to as a subject, while an individual element y_{ij} will be called an observation. The total number of observations is $N = \sum n_i$, cases are independent, and β_i is independent of ϵ_i.

Integrating β_i in (1) gives the marginal model

$$
Y_i \sim N(X_i\alpha, V_i)
\tag{2}
$$

where $V_i = Z_i D Z_i^t + \sigma^2 I$, and I is the appropriately sized identity matrix. Other important models substitute different matrices for V_i. The AR-1 model has $Y_i \sim N(X_i\alpha, V_i^{(AR)})$, where the jk^{th} element of $V_i^{(AR)}$, $V_{ijk}^{(AR)} = \sigma^2 \rho^{|t_{ij} - t_{ik}|}$. Also worth considering is the independent increments model, where $V_{ijj}^{(II)} = \sigma_{ij}$ and $V_{ijk}^{(II)} = V_{ikj}^{(II)} = \sigma_{ij}$ when $j < k$.

This paper takes an ecumenical approach to estimation. Examples of Bayes, empirical Bayes and ad hoc estimation are used. Experience with parallel plots suggests that the method of estimation is secondary to the value of the plots.

Empirical Bayes estimation works from model (2). Estimates of (α, σ^2, D) are produced by maximizing the likelihood resulting from (2), and estimates of β_i come from maximizing the resulting conditional likelihood in (1) given estimates of (α, σ^2, D). Bayes estimation using Gibbs sampling (Gelfand,

4

Racine-Poon, Hills and Smith 1990) gives posterior mean estimates of the parameters $(\alpha, \beta_i, \sigma^2, D)$. Ad hoc approaches can be used, particularly in early stages when the model is simple. An example used with the weight loss data occurs when $X_i = Z_i = 1$, the vector of ones of length n_i. In this case, $\alpha + \beta_i$ can be estimated by the mean of the n_i observations Y_i. Also possible is REML estimation (Laird and Ware 1982).

Any of these methods of estimation can be used to calculate the residuals discussed in the next subsection. Let a hat over a parameter, for example, $\hat{\alpha}$, $\hat{D}$, $\hat{\beta}_i$ and $\hat{\sigma}^2$ denote estimated parameters. Particular methods of calculation will be mentioned as part of the examples.

2.2 Residuals

At least three types of residuals have been defined in the literature (Weiss and Lazaro 1992). The E-residuals follow from model (1) and are estimates of ϵ_i

$$E_i = Y_i - X_i\hat{\alpha} - Z_i\hat{\beta}_i$$

are from Weiss and Lazaro (1992). The cross-sectional (Louis 1988) or R-residuals are

$$R_i = Y_i - X_i\hat{\alpha}.$$

and the random effect estimates $\hat{\beta}_i$ are also called residuals by Waternaux, Laird and Ware (1989). The R_i residuals are the sum of E_i and $Z_i\hat{\beta}_i$.

All three residuals are useful. The E_i include variation due to missing random effects and due to unmodeled fixed effects that are, crudely speaking, orthogonal to Z_i. They also contain information for checking on the constant variance specification σ^2, and on the presence of $\epsilon-$outliers, outliers due to large ϵ_{ij}. The $\hat{\beta}_i$ contain information on unmodeled fixed effects in the span of the Z_i, on the constant variance assumption D and on $\beta-$outliers, outlying observations due to large β_{ik}, where β_{ik} is the kth random effect for case i. In particular, in the equicovariance model with Z_i a column of ones, the E_i will contain information about the effects of time trends and time

varying covariates, and the β_i will contain information about the effects of baseline and other time fixed covariates.

2.3 Model development

The modeling plan is a variant on Box (1980), who describes data analysis as a cycle of fitting and criticism. Here the idea is to plot the data; identify what structure we can in the data; and fit the data using a model that incorporates the structure we are aware of. The model structure may incorporate subject matter knowledge and is extended through the graphics. We then plot the data and residuals based on the current model. Additional structure is identified and incoporated into a new model; old structures are modified as needed. The picture is that of a spiral: with increasing cycles more and more of the variability in the data are described. Likelihood ratio tests or Bayes factors might be used to distinguish between models for which the graphics are equivocal. The final inference is the model, rather than point and interval estimates based on the model.

2.4 Goodness-of-Fit

Goodness-of-Fit is the comparison of two sets of information. A goodness-of-fit test in a generalized linear model (glim, McCullagh and Nelder 1989), depends upon the variance in binomial or Poisson regression being determined, once the linear predictors are known. The information about the variance from the fitted model is compared to the variance information in the residuals. In a QQ plot, a sample from one distribution is compared to a theoretical distribution. In either case if the two sets of information disagree, we have lack-of-fit. In graphics, goodness-of-fit requires two sources of information and a method for comparing them. Clearly many possible goodness-of-fit graphics are possible. One possibility is to have a preconceived idea of the shape of a plot, i.e. residual plots in linear regression should show no nonlinear trends. The second possibility is to overlay information from two sources to permit easy comparison. One source might be

from a data fit, the second source might be raw residuals.

3 Examples

3.1 Weight Loss Data

The weight loss data consist of up to eight weekly repeated measures of weight on women enrolled in a weight loss trial. Patients were interviewed the first week and enrolled in the study at the second week. Measures 1, 2, 3 and 6 were clinic visits. At weeks 4, 5, 7 and 8 measures were taken by the women on their home scales at the request of an interviewer calling over the phone. The design is unfortunate in that home weighing and weight loss are confounded: this confounding will be ignored in the sequel, although it would be kept in mind in further modeling. The data are plotted in Figure 1.

Figure 1 shows very little structure, except for the numerous parallel lines. This indicates a random intercept should be included in the initial model. Figure 2 shows a residual plot (Weiss and Lazaro 1992) of the data from the simplest possible model where $X_i = Z_i = 1$. Each subject's mean is $\mu_i = \alpha_0 + \beta_i$, and is estimated in an ad hoc fashion by $\bar{Y}_{i.} = n_i^{-1} \sum (y_{ij})$. Residuals are calculated as $E_i = Y_i - \bar{Y}_{i.} 1$, where 1 is an appropriate sized vector of ones. In stark contrast to Figure 1, the residual plot shows a complicated story reflecting the design of the study. From week 1 to week 2, no weight is lost, if anything, a little weight is gained. From week 2 to week 3, the first week of weight loss, the patients lose quite a lot of weight. At weeks 4 and 5 they continue to lose weight. At week 6, suddenly weight is gained, presumably since patients are weighed under supervision with a properly calibrated scale. At weeks 7 and 8, they continue to lose weight. The overall weight loss indicates that a fixed time effect is needed in the model.

3.2 Pediatric Pain Data

The pediatric pain data consists of up to four observations on 61 children aged eight to ten with up to four repeated measures on a measure of pain tolerance. Pain tolerance is measured as the length of time in seconds that the child can tolerate keeping his or her arm in cold water. Two measures were taken during a first lab visit followed by two more measures during a second visit after a two week gap. The children were classified into one of two groups, attenders (A) or distracters (D) during the first visit according to their style of coping (CS) with the pain. During the second visit, a treatment (TMT) was administered prior to the fourth trial. The treatment consisted of counseling to either attend (A), distract (D) or no advice (N). Interest lies in the main effects of TMT and CS and interactions between TMT and CS. The data are plotted in Figure 3. This figure will be discussed in the next section.

The X_i matrix for each case is 8×4. The first column is a column of ones; the second column is a column of ones or zeros, an indicator of CS; the last six columns are all zeros, except for a single one in the last row in the column that corresponds to the CS/TMT group (AA, AD, AN, DA, DD, DN) the child belongs to.

3.3 Test Data

The test data consists of seven repeated test scores of 13 children at three month intervals. The data are plotted in Figure 4. It appears that a fixed intercept, and possibly a random slope is appropriate. The individual trends for the last few time points are qualitatively different from the shape at the first several times. An analysis (not presented here) suggested that a fixed intercept, random slope, and fixed quadratic were needed to fit the data.

3.4   Catheterization Data

This data set consists of 4 repeated measures on heart rate (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP) for each of 14 infants and children. Measures were taken the morning before a long heart catheterization procedure, immediately after, the evening after, and the following morning. The idea is that the profiles should show changes due to the stress of the procedure followed by a return to baseline.

4   Parallel Plots

Basic interpretation of parallel plots is given by Weiss and Lazaro (1992). Parallel flat lines indicate need for a random intercept, parallel sloped lines indicate the need for a fixed slope as well as a random intercept, and differing slopes indicate the need for random slopes. Univariate outliers and bivariate outliers of the form $(y_{ij}, y_{i(j+1)})$ can be detected easily.

Several problems occur with parallel plots. i.) They can quickly become dense piles of ink, as in the bottom of Figure 3 and parts of Figure 2; ii.) They can apparently show very little interesting structure, as in Figures 1 and 3; and iii.) They can also be sparse, as in the top of Figure 1 where we are unable to see slopes of lines or sparse and waste space as in the upper left and lower right of Figure 4. These problems, particularly the problem of line density have earned parallel plots the appropriate appelation of spaghetti plots.

Several solutions have been tried in the literature. One approach is to summarize the data. For example, Figure 5 displays repeated box plots showing the (min, lower quartile, median, upper quartile, and maximum) of the residual weight data at each time. The box plots are as designed by Tufte (1983, pp. 123-5). These plots lose the blocking information: the information about which cases belong to the same individual. For example, we lose the information that the three largest residuals at week 1 increase their weight substantially at week 2. Researchers often plot the profiles of means

for several groups. This loses both inter- and intra- case variability. The repeated box plot and profile of means can be considered inferential tools: as such they are valid displays. In general, summary plots lose information, but in exchange, concentrate our attention on the remaining information.

Another approach is to plot the data Y_i or residuals E_i without the line segments connecting (t_{ij}, y_{ij}) and $(t_{i(j+1)}, y_{i(j+1)})$. These plots are fatally flawed because they lose the information contained in the line segments between observations within a case. The line-density plots of Miller and Wegman (1991) also do not allow connections to be drawn between points more than 1 time unit apart.

In a different approach, many researchers plot individual profiles on separate plots, hopefully keeping the physical dimensions of each plot and scaling of each axis the same and then look at the set of profiles; this was reported for example, in Gregoire, Schabenberger, and Barrett (1993), but the procedure is much older than that paper; a recent example is in Cleveland (1993). Gregoire, Schabenberger and Barrett (1993) note that this is a collection of small multiples (Tufte 1983, Chapter 4). Figure 6 shows a sample of four residual profiles from the weight loss data. These plots maintain the connections between points within a case, but lose the easy comparison across cases that the overlaid parallel plots have. Similarly, the collection of slopes at a time t cannot easily be compared to the slopes at time t' in the plot. Nor it is simple to compare two cases to see which has a larger slope at a particular time. Of course, slopes are not visible in figure 1 and barely visible for a few cases in figure 3.

The problems identified in this section are i.) overplotting and sparseness of lines, ii.) lack of connection between y_{ij} and $y_{i(j+1)}$, iii.) difficulty in comparing y_{ij} and $y_{i'j}$; iv.) Lack of connection between y_{ij} and $y_{i(j+2)}$; and v.) difficulty in viewing and comparing slopes.

Solutions given below to these problems are varied, and the choice of solution for a particular problem will depend on the sample size of the data set and available software. It is not intended that all of the solutions be

used in any one problem. The solutions given in the sequel borrow graphical ideas from other places and then these ideas are applied to the problems of multivariate data display for repeated measures. Specific ideas used are the shape parameter (Cleveland, McGill, and McGill 1985), the casement plot (Chambers, Cleveland, Kleiner and Tukey 1983), background, and dynamic graphics.

4.1 The information in a parallel plot

The first point is that while Figures 1 and 3 are flawed, they do both contain useful information and this information should be used in constructing an initial model and in constructing further plots. Neither figure 1 or 3 is a likely candidate for publication in a paper analyzing these data; rather, each is a useful first exploratory plot for understanding the structure of the data. Figure 1 shows that each woman varies around an average weight; any reasonable first model should include either a random intercept, a baseline weight as a predictor or other similar structure. Figure 3 shows skewness in the responses; also children with generally large responses have large variability. This suggests a needed data transformation and further analyses of these data take place on the log scale. The plot also shows that a few extreme values were censored at t=240 seconds, a problem that will not be dealt with further here but needs to be addressed as part of the modeling.

4.2 The shape parameter

Figure 1 is the default shape of the plot in many statistics packages. Slopes are not visible. In contrast, we can see the slopes in Figure 2. Cleveland, McGill and McGill (1985) suggest choosing the shape parameter of a plot to make slopes of interest easily comparable in the plot. The shape parameter is the ratio of the height H of the x-axis to the width W of the y-axis. Slopes of greatest interest here might either be the slopes

$$s_{ij} = \frac{y_{i(j+1)} - y_{ij}}{t_{i(j+1)} - t_{ij}}$$

or the slopes

$$s_i = \frac{y_{in_i} - y_{i1}}{t_{n_i} - t_1}.$$

The physical slopes PS_{ij} visible in the plot are

$$PS_{ij} = \frac{H * (y_{\max} - y_{\min})^{-1}}{W * (t_{\max} - t_{\min})^{-1}} * s_{ij}$$

where H is the physical length of the y axis in centimeters (cm); W is the physical length of the x axis in cm; $H * (y_{\max} - y_{\min})^{-1}$ is a conversion factor from data units to the physical page dimensions, say cm, for the horizontal axis, and $W * (t_{\max} - t_{\min})^{-1}$ is the conversion factor for the vertical axis, from time to cm. Cleveland, McGill and McGill (1985) suggest setting the shape parameter H/W optimally so that the physical slopes cluster around ± 1. Figure 7 shows the weight data with an appropriate shape parameter chosen to make the average physical slope be approximately 1. Figure 7 encodes information about the relationship between baseline values and shape. In contrast, Figure 2 trades baseline information for an expanded y axis and the effect of n similar segmented lines drawn in close proximity.

The theory of shape parameters partially explains the usefulness of the repeated measures residual plot. The slopes on optimized versions of figures 1 and 2 will be the same. Consequently, the ratio of optimal shape parameters is the ratio of the ranges of the raw Y_i data to the E_i residuals up to error caused by the graphics program prettying up the axis labels. Consider the range of balanced data Y_i with $n_i \equiv n_0$ under the null model $Y_{ij} = \alpha_0 + \beta_i + \epsilon_{ij}$, $\mathrm{Var}(\beta_i) = D$ and $\mathrm{Var}(\epsilon_{ij}) = \sigma^2$. A back-of-the-envelope calculation suggests that the range of the raw data plot will be proportional to $\sqrt{\sigma^2 + D}$, assuming $D \gg \sigma^2$ while the range of the residual plot will be proportional to σ, ignoring a constant $(1 - n_0^{-1})^{1/2}$. Therefore, the residual plot will be a factor of $\sigma/(\sigma^2 + D)^{1/2}$ closer to the optimal scale factor than the raw data plot; a factor equal to the square root of one minus the intraclass correlation coefficient.

12

4.3 Casement plots

Casement plots (Chambers, Cleveland, Kleiner, and Tukey 1983) provide a useful way of reducing clutter on a single parallel plot by stratifying observations according to a covariate. Figure 8 shows a casement plot for the pediatric pain data. The data have been log transformed, the model detailed in section 3.1 fit, and E-residuals $Y_i - X_i\hat{\alpha} - Z_i\hat{\beta}_i$ plotted in 6 subplots corresponding to each combination of the predictors. The parallel plot with all residuals shows little structure, and the casement plot was produced to see if the variation differed by CS or by TMT. The distracters have slightly greater variability than attenders, and the DN group has an extreme outlier. Casement plots are used by Crowder and Hand (1990, chapter 2) although not by name. The physical size and units per physical length of the subplots must be kept the same for all plots. This requires extra work on the part of the graphics author since most graphics programs do not easily allow this type of setup. Xlisp-stat (Tierney 1990) is a possible exception.

4.4 Plot Backgrounds

An extreme version of the casement plot are the plots of individual profiles as illustrated in Figure 6. The problem with the individual profiles is that it is difficult to compare absolute levels or slopes across individuals. The identification of sets of profiles with similar shapes is more difficult with profiles drawn on separate plots. This operation is called assembly by Cleveland (1993). Cleveland (1993) suggests horizontal and vertical grid lines for data of this sort, calling it *a visual reference grid.* Here I recommend an alternative visual reference system composed of a data based summary. Figure 9 shows the same selection of four profiles from Figure 6 against the *background* of the summary box plots of Figure 5. Figure 9 shows that the choice of four profiles was not arbitrary, rather the profiles corresponding to the weights at week one which were the two highest, an approximately median value, and the lowest value.

Technically, background is a visual reference grid, however I prefer to distinguish between a visual reference grid, which may have no statistical content, and background, which has statistical content. Background is also to be distinguished from the *chart junk* of Tufte (1983) which would actively detract from interpreting statistical information in the plots.

The background in figure 9 is based on data summaries. Alternative sources for background could be summaries from a fitted model, in which case the plot serves as a graphic for checking goodness-of-fit.

Any figure where we have a strong opinion about the shape of the plot given the model provides a graphical goodness-of-fit test. In cases where the opinion would be based on the model fit or in situations where the comparison is difficult because our opinion is non specific, using background is a good way to contrast the information directly with data or residuals. For example, in Figure 2, if we expect no fixed or random slopes then the profiles should wander around the zero axis at both short and long times. The plot as actually displayed tells us that our basic model does not fit.

4.5   Dynamic graphics

Dynamic graphics may assist in the viewing of many profiles. First I discuss a few plots which I have found useful, and to direct others towards more fruitful lines of research, some plots which I have found less than useful.

A useful dynamic plot of residuals has been, starting with a blank canvas, to add residual profiles to Figure 2 one at a time, with a 2-3 second pause between additions. Too short of a pause confounds identification of the added profile, too long of a pause causes boredom. This dynamic plot was the plot that first made me realize that Figure 2 indicates the need for a random slope. The profiles, as they are placed on the plot, have similar shapes, but the shapes rock back and forth around a pivot that is roughly located at the median time at week 4. Having discovered the random slope structure in figure 2, it becomes easy to construct a static plot that emphasizes the need for a random slope. Figure 10 shows a set of profiles from

the weight loss data chosen specifically to show differing slopes amongst the data.

Background is not necessarily needed for the dynamic display. However, the display of the profiles was most interesting for roughly the fifth through the 25th to 30th cases, after which there were usually too many cases on the screen to process. I suspect that the first few cases formed a background against which later profiles were compared. This suggests considering a dynamic graph with a fixed number of profiles, and, as profiles are added, some are removed, so that a constant amount of clutter remains on the screen. In large data sets, it may not be necessary to view the entire data set.

A number of variants are possible. One which has on one occasion been useful, did use background. First the complete set of profiles were drawn on the plot in color. Then, one at a time, cycling through cases, each case's profile is redrawn in a clashing color. When the next case is drawn, the previous case is returned to the background color.

I have not found drawing single profiles, erasing them, and then drawing the next profile on the same plot to be useful. The removal of most of the information in the plot followed by adding new information is very distracting.

These plots along with many others were prototyped on the weight loss residual data, and on another much larger study of weight loss over many weeks. Further study of dynamic parallel plots is certainly possible, although the ratio of value to effort is uncertain at best.

5   Plots Against Covariates

This section only begins the discussion of plots that show the relationship between Y_i and covariates X_{ik}. Time pervades repeated measures data analysis, so when possible, profile plots have been taken as the basic plot. The casement plot of figure 8 plots the $E-$residuals against the entire set of

covariate patterns for this data set. For time invariant covariates, Figure 11 shows a histogram of the posterior means of the random effects for the logged data. The histogram is composed of individual bricks. Solid bricks correspond to distracters, while the clear bricks which float on top of the solid ones are for attenders. Unwin (1992) illustrates an earlier histogram of this form in another context. The main features of this histogram are the large attender outlier, and, ignoring the outlier, the differing variance in the two groups. There is also slight skewness in the data and a few gaps suggesting possible non-normality.

Figures 8 and 11 show ways of plotting Y_i or E_i against discrete valued covariates. The time varying TMT assignment is not the most general type of covariate, since while it is assigned at random across children, within children it has a very specific design. I do not have a proposal yet for showing the relationship between a randomly varying discrete covariate and Y_{ij}.

Continuous predictors W_{ij} are possibly more difficult. The relationship between a continuous predictor and a continuous response is often bound up in the modeling of multivariate repeated measures, treating both W_{ij} and Y_{ij} as responses. Figures 12a and b show a line plot for the catheterization data. Lines connect consecutive points (W_{ij}, Y_{ij}) to $(W_{i(j+1)}, Y_{i(j+1)})$ within a case. Points between cases are not connected. The line plot can be thought of as a generalization of the parallel plot in two ways. In the first, time t_{ij} is replaced by a variable W_{ij} which may not be monotone increasing in j, as t_{ij} is assumed to be. Alternatively, consider a three dimensional parallel plot, with points (t_{ij}, W_{ij}, Y_{ij}) and points within a case connected in time sequence. Then the line plot is the edge view of this plot, collapsing over time. Figure 12a shows possibly increasing variance of DBP with increasing HR. Figure 12b shows a bivariate outlier in {SBP, DBP} space that is not an outlier on any single index. It shows the general correlation between these two measures. Symbols may be added to the plot to indicate the beginning (x) and ending (o) times. This tends to clutter the figure, although in the

case of {HR, DBP}, it adds information which was not easily noticed in the parallel plot of DBP nor in the plot of {SBP, DBP}: with one major exception, the DBP tends to be much higher at the last time point than at the beginning.

Another set of plots for covariates are developed by Segal (1994) who adapts regression tree (Breiman, Friedman, Olshen and Stone 1984) methodology for inference. He splits the predictor space into regions with similar response profiles, and then plots profile plots for each of the final nodes in the tree.

6   Variance-Covariance specification

The main plot for the inspection of the variance-covariance model is a scatterplot matrix of Y_i, R_i, or E_i. The (k, k') subplot of the scatterplot matrix plots R_{ik} against $R_{ik'}$, for example. This is clearly sensible for balanced equally spaced data with possibly missing values. The scatterplot matrix produced by most statistical software has a flaw that must be remedied for covariance structures of repeated measures data. The range of the kth axis runs from $\min(R_{ik}, k = 1..n_0)$ to $\max(R_{ik}, k = 1..n_0)$, with possible adjustments to make the tick marks be simple numbers. This means that equal distances on different subplots will normally not be an equal number of data units, and comparisons across graphs become difficult. The scaling will roughly but unfortunately not exactly correspond to standardization by the sample range of the $R_{ik}, k = 1..n_0$, and we will have a crude picture of the correlation structure of the Y_i's. Cleveland, Diaconis and McGill (1982) report that the perceived correlation of a scatterplot also changes with the scaling, so if the scales in different subplots are quite different, this could affect interpretation.

There are several possibilities for the solution of these problems. One is to make sure that each axis has the same data units per inch. If the global minimum and maximum data values are used for all variables, there

may also be information on changes in location across time as well as the covariance, but since the parallel plot satisfactorily displays information about changes in location and variance across time, the main reason for inspecting the scatterplot matrix is to check the correlation. Recommended is to use the same range on each axis, with varying minima. When the variance is constant across time, then this plot de-facto gives a picture of the correlation structure. When the variance is not constant, then the changing variance may overwhelm the correlation information. Thus I recommend plotting standardized data, using different minima for each variable, and constant ranges. Let s_k be the sample standard deviation of the variable at time k, and plot Y_{ik}/s_k for example. Note that the definition of s_k changes as appropriate for Y_i, E_i or R_i.

Figure 13 shows the E residuals for a simulated data set. The data was generated with an AR-1 data structure with parameter $\rho = .8$, while the E residuals were calculated based on a random intercept model. The data were not standardized, since the variances appear constant. Checking a particular model can be aided by the addition of background. The ellipses are based on the sample means, variances, and correlations of the plotted points, and plotted using a normal distribution based on those estimates to produce a 50% sample region. The residuals would be expected to show approximate uncorrelatedness if the true model were indeed a random intercept model. The negative correlation between the lag 2 and 3 residuals shows that subtracting the estimated random effects has over adjusted the residuals so that they now have a strong negative correlation instead of a zero correlation.

Figure 14 shows the E residuals for the test data, using standardization. In fitting the model, the quadratic term has been omitted. The 1,2 cell in particular shows a correlation between the first two time points that is quite strong, indicating that the model is incorrect. The ellipses are 50% prediction intervals for the E residuals based on the fitted model. The ellipses clearly do not agree with the data, suggesting a lack of model fit.

18

The ellipses provide a partial check on normality also, but see Dempster and Ryan (1985) and Lange and Ryan (1989) for Q-Q plots specifically designed for checking normality. Figure 14 provides another example of goodness-of-fit. The lack of overlap between the ellipses and the data indicate a lack of fit of the current model, although unlike other plots discussed, Figure 14 does not readily lend itself to model elaboration.

7 Further Directions

I have only scratched the surface of dynamic plots for profile display and much more study is clearly possible.

For covariates, I have not found color to be useful for encoding covariates. There are three possible reasons for this. One is that the covariates did not distinguish between different cases and thus better examples are needed, the second is that color itself does not work for distinguishing profiles in the examples I used, and the third is that the combination of system which I used and the display types (ie choice of colors, line widths, monitor) were not suitable for the discovery of information. Plotting methods for continuous time varying covariates is a topic of current research.

The suggestion of nonconstant variance in the pediatric pain random effects and also in the E residuals raises the question that perhaps cases with large values of β_i may also have a larger variance of the E_i. A plot for detecting non-independence of the β_i and ϵ_i is desirable. I have not discussed the standardization of residuals prior to plotting. For the parallel plots, for the purposes put to them here, this seems to be unnecessary. For detection of nonconstant variance, this may be more useful. Standardization of residuals is discussed in Rutter and Elashoff (1994).

The plots suggested here are intended for nearly balanced data with small to moderate sizes, and should permit users to avoid terrible modeling blunders. Refinements of all the plots presented are certainly possible, and plots for checking the relationship between covariates and the response need

development. For data collected at random times, the scatterplot matrices need modification. One possibility is to have the (k, k') subplot contain a point from all observations with one observation in the window $t_k^0 < t_{ij} < t_{(k+1)}^0$, and $t_{k'}^0 < t_{ij} < t_{(k'+1)}^0$. The times t_k^0 would be chosen to bracket the range of times in the data, and could be chosen either to be equally spaced or have equal numbers of points in each window. Occasional cases with 2 observations in one or both of these windows must be handled.

Many other plots are useful. For example, the correlogram and variogram (Diggle 1990 chapter 2; Diggle 1988) are also useful for checking the covariance structure.

References

[1] Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *JRSS-A*, 143, 383-430.

[2] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone C. J. (1984). *Classification and Regression Trees.* Belmont, CA: Wadsworth.

[3] Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis.* Belmont, CA: Wadsworth.

[4] Cleveland, W. S. (1985). *The Elements of Graphing Data.* Belmont CA: Wadsworth.

[5] Cleveland, W. S. (1993). A model for studying display methods of statistical graphics (with discussion). *Journal of Computational and Graphical Statistics.* 2, 323-364.

[6] Cleveland, W. S., Diaconis, P. and McGill, R. (1982). Variables on scatterplots look more correlated when scales are increased. *Science*, 216, 1138-1141.

[7] Cleveland, W. S., McGill, M. E., and McGill, R. (1988). The shape parameter of a two-variable graph. *J. of the Amer. Stat'l. Assn.*, 83, 289-300.

[8] Crowder, M. J. and Hand, D. J. (1990). *Analysis of Repeated Measures.* London: Chapman and Hall.

[9] Dempster, A. P. and Ryan, L. M. (1985). Weighted normal plots. *J. of the Amer. Stat'l. Assn.*, 80, 845-850.

[10] Diggle, P. J. (1988). An approach to the analysis of repeated measures data. *Biometrics*, 44, 959-71.

[11] Diggle, P. J. (1990). *Time Series, A Biostatistical Introduction.* Oxford: Clarendon Press.

[12] Draper, D. (1987). Comment: On Exchangeable Judgements in Predictive Modeling and the Role of Data in Statistical Research. *Statistical Science*, 2, 454-461.

[13] Gregoire, T. G., Schabenberger, O., and Barrett, J. P. (1995). Modelling irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Canadian Journal of Forest Research*, to appear.

[14] Izenman, A. J. and Williams, J. S. (1989), A Class of Linear Spectral Models and Analyses for the Study of Longitudinal Data. *Biometrics*, 45, 831-849.

[15] Jones, M. C. and Rice, J. A. (1992). Displaying the important features of large collections of similar curves. *American Statistician*, 46, 140-145.

[16] Laird, N. M. and Lange, N. (1987), Comment. *Statistical Science*, 2, 451-454.

[17] Laird, N. M. and Ware, J. H. (1982), Random-Effects Models for Longitudinal Data. *Biometrics* 38, 963-974.

[18] Lange, N. and Ryan, L. (1989), Assessing Normality in Random Effects Models, *The Annals of Statistics*, 17, 624-642.

[19] Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1991). Robust statistical modeling using the t-distribution. *J. of the Amer. Stat'l. Assn.*, 84, 881-896.

[20] Louis, T. A. (1988), General Methods for Analysing Repeated Measures. *Statistics in Medicine*, 7, 29-45.

[21] Miller, J. J. and Wegman, E. J. (1991). Construction of Line Densities for Parallel Coordinate Plots. In *Computing and Graphics in Statistics*, eds. A. Buja, and P. A. Tukey, 107-123.

[22] Rao, C. R. (1987), Prediction of Future Observations in Growth Curve Models. *Statistical Science*, 2, 434-471.

[23] Rutter, C. and Elashoff, R. M. (1994). Analysis of Longitudinal Data: Random Coefficient Regression Modelling. To appear, *Statistics in Medicine.*

[24] Segal, M. (1994). Regression trees for repeated measures. *Computational and graphical statistics*, , .

[25] Tierney, Luke (1990), *Lisp-Stat.* New York: Wiley.

[26] Tufte, E. R. (1983). The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press.

[27] Unwin, A. (1992). How interactive graphics will revolutionize statistical practice. *The Statistician*, 41, 365-369.

[28] Waternaux, C., Laird, N. M. and Ware J. H. (1989). Methods for Analysis of Longitudinal Data: Blood-Lead Concentrations and Cognitive Development. *Journal of the American Statistical Association*, 84, 33-41.

[29] Weiss, R. E. and Lazaro, C. G. (1992). Residual Plots for Repeated Measures. *Statistics in Medicine*, 11, 115-124.

Figure Captions

Figure 1. Parallel plot of the Weight Loss data.

Figure 2. Parallel plot of Weight Loss residuals.

Figure 3. Parallel plot of the Pediatric Pain data.

Figure 4. Parallel plot of the Test Score data.

Figure 5. Summary box plots of Weight Loss residuals.

Figure 6. A selection of four profiles of Weight Loss residuals.

Figure 7. Weight Loss data with appropriate shape parameter.

Figure 8. Casement plot of Pediatric Pain Residuals. Attenders are in subplots (a), (c), and (e), Distracters in (b), (d), and (f). The Attend treatment was given to individuals plotted in (a) and (b); the distract treatment to those in (c) and (d); and non-treatment to those in (e) and (f).

Figure 9. Weight loss residuals with summary box plot background.

Figure 10. Selection of weight loss residuals with background to show random slope. The axis of the curves seems to be around week four.

Figure 11. Histogram of random effects with solid blocks for distracters and clear blocks for attenders.

Figure 12. Line plots of a.) Heart rate and Diastolic Blood Pressure and b.) Systolic Blood Pressure and Diastolic Blood Pressure for 13 babies.

Figure 13. Scatterplot matrix of E-residuals from simulated data. The raw data has an AR-1 covariance structure but has been fit using a random intercept model. The ellipses are based on a normal fit using the numbers plotted in each plot.

Figure 14. Mice data: Scatterplot matrix of standardized $E-$residuals from a model without quadratic effect with superimposed circles indicated ostensible fitted 50% content regions.
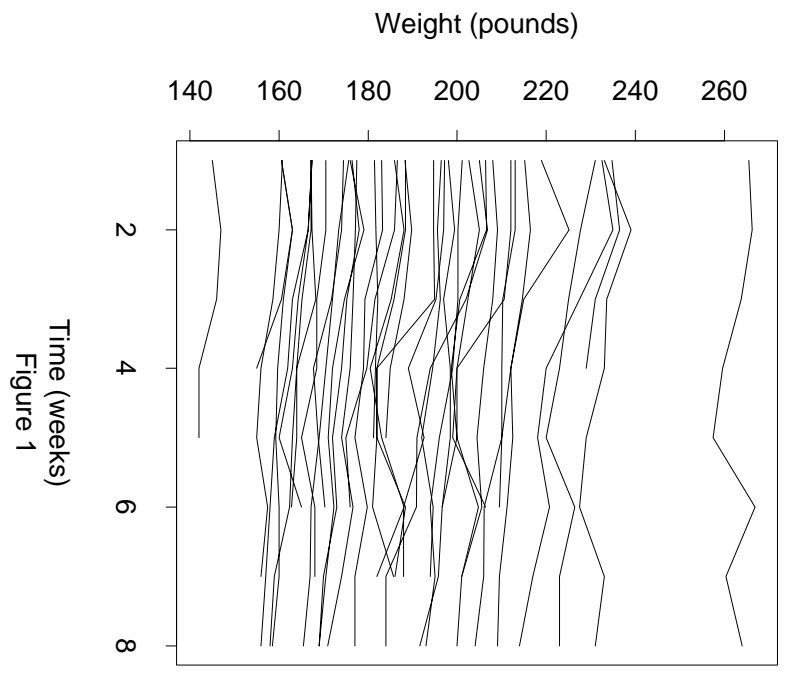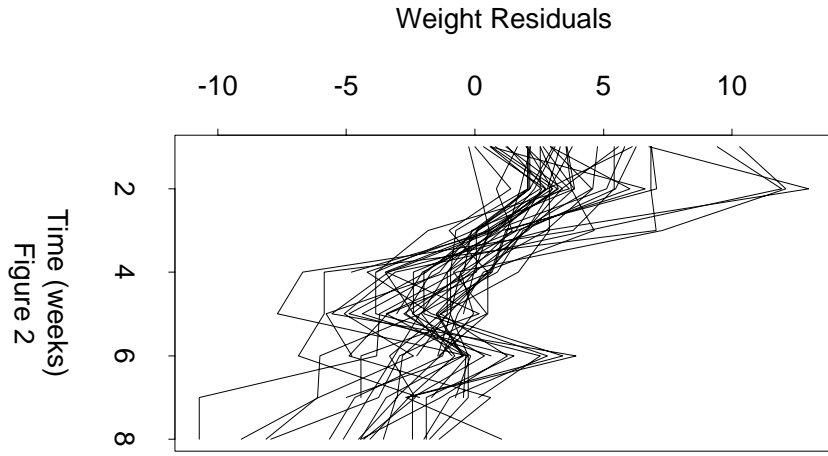
Figure 1

Weight Residuals

-10   -5   0   5   10

Time (weeks)
Figure 2
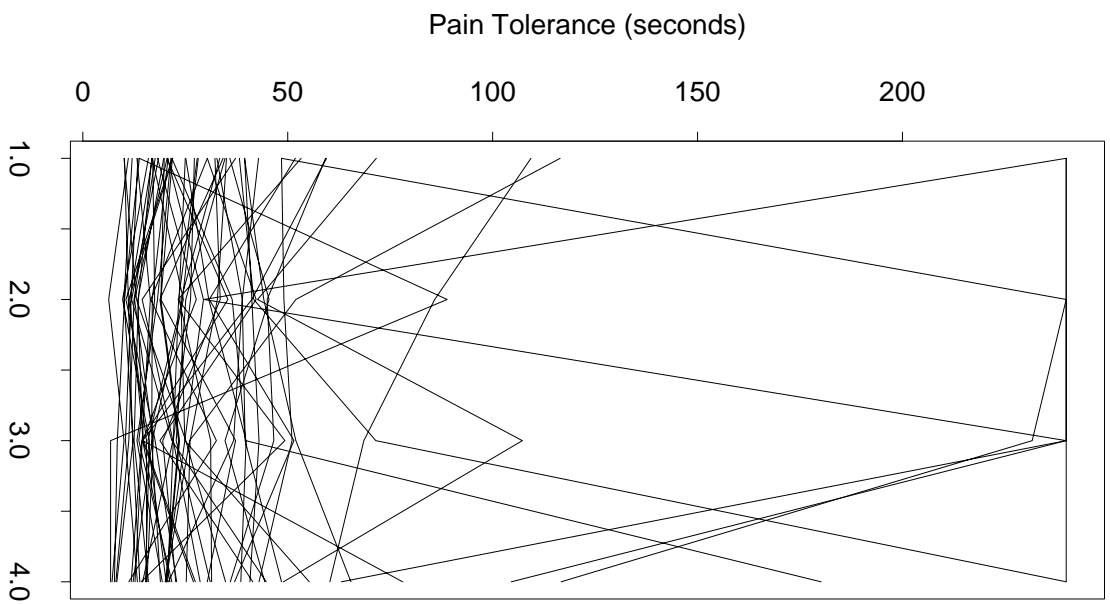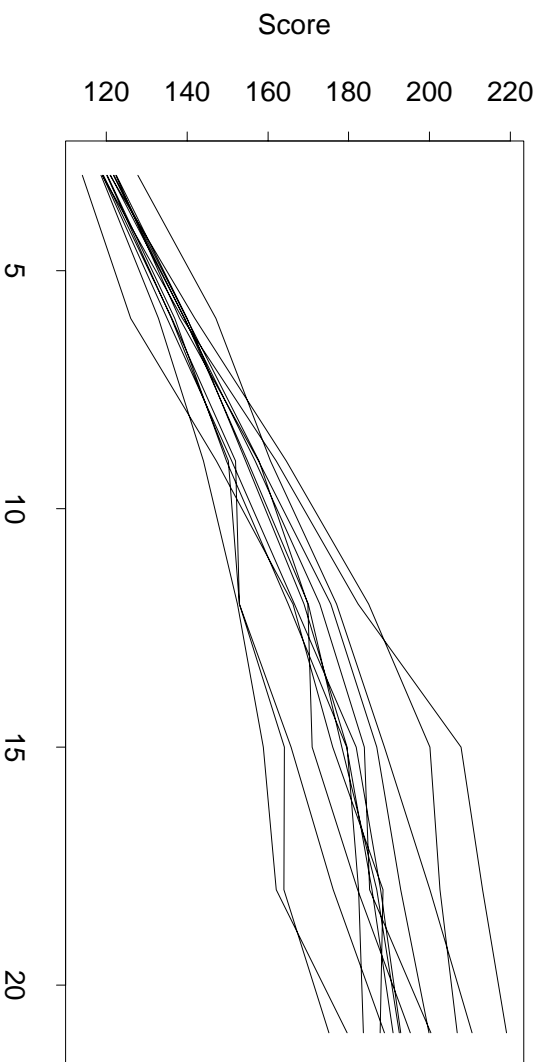
2   4   6   8

Pain Tolerance (seconds)

Trial number
Figure 3

27

Figure 4

Figure 5

weight residual

-20 -10 0 10 20

week
2
4
6
8

weight residual

-20 -10 0 10 20

week
2
4
6
8

weight residual

-20 -10 0 10 20

week
2
4
6
8

weight residual

-20 -10 0 10 20

week
2
4
6
8

30 Figure 6

Figure 7

(a)

(b)

(c)

(d)
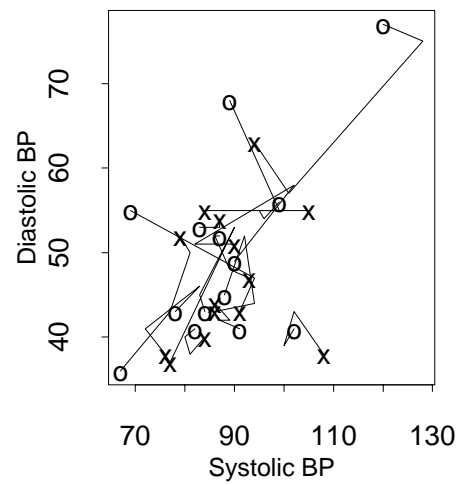
(e)

(f)

Attenders

Distracters
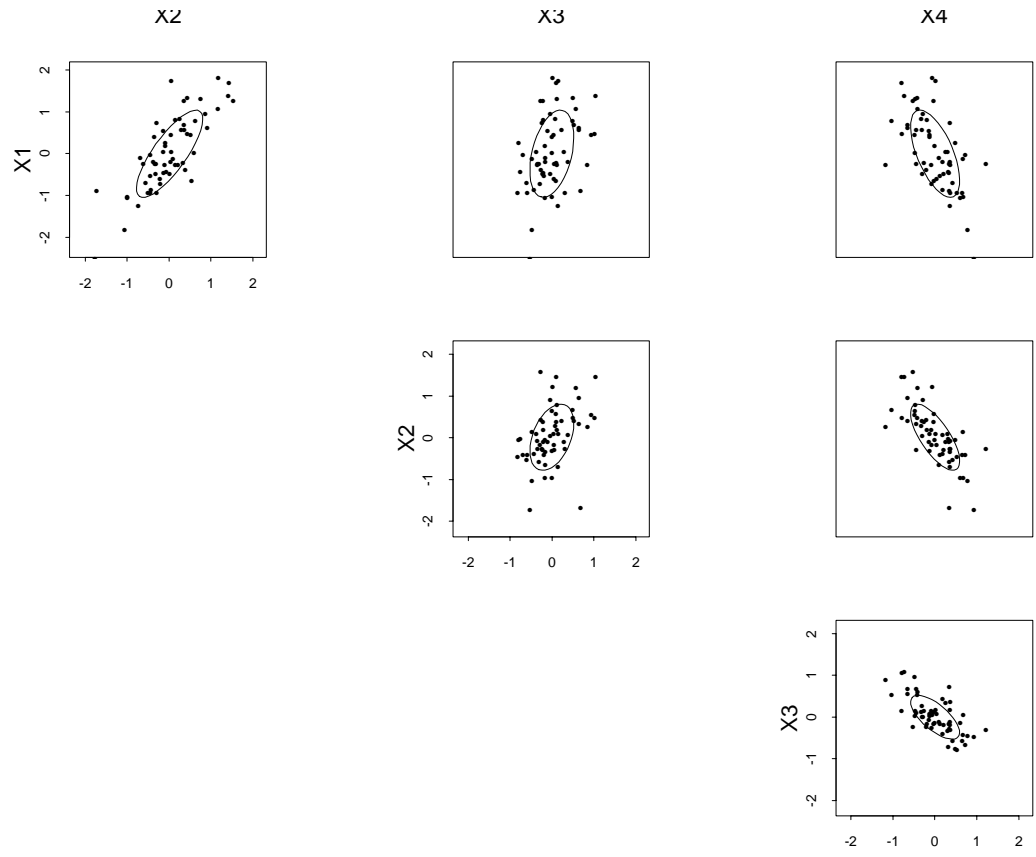
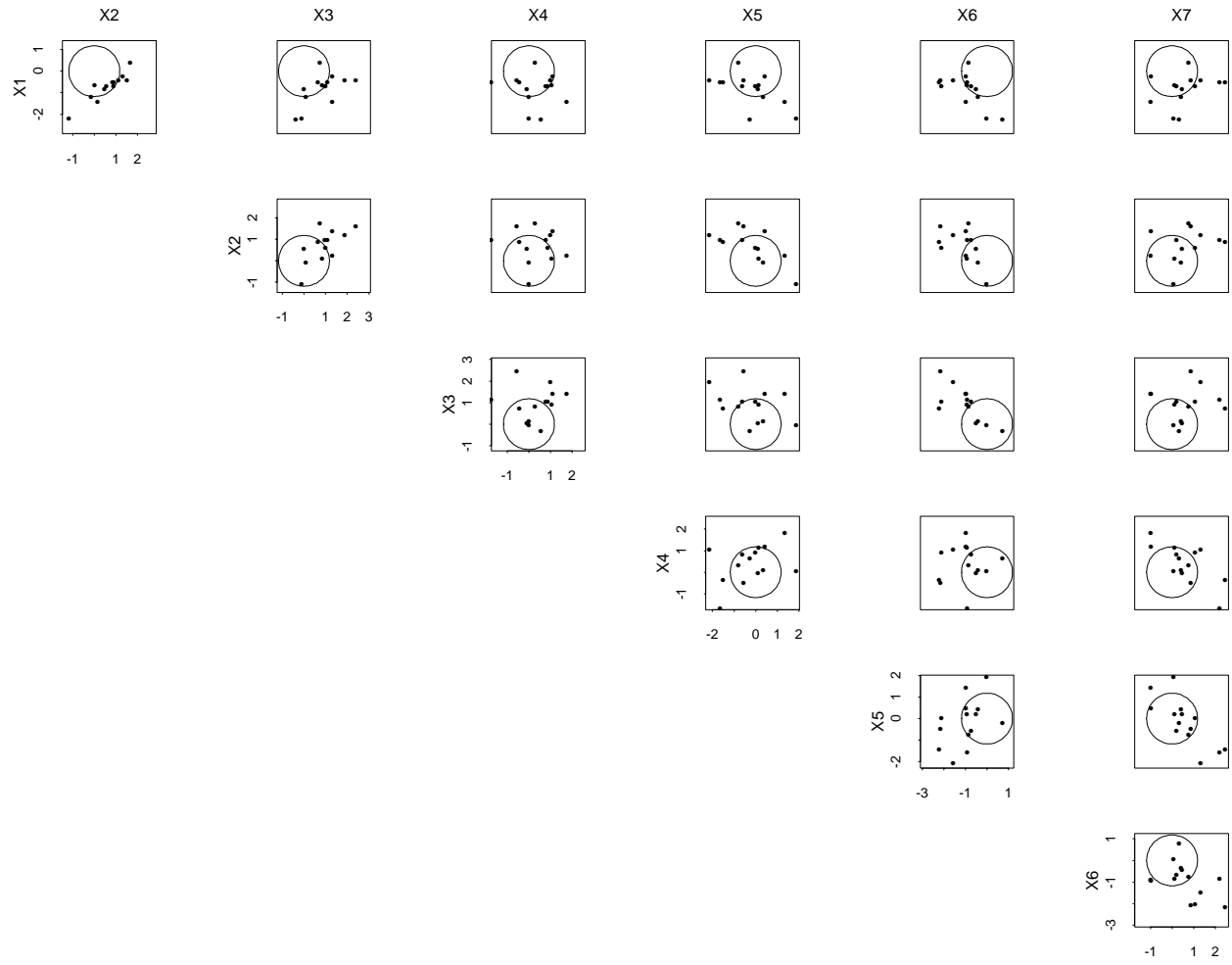Figure 8

(a)

(b)

(c)

(d)

33
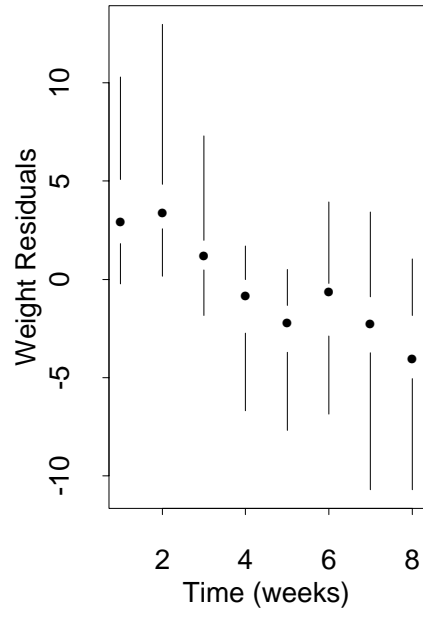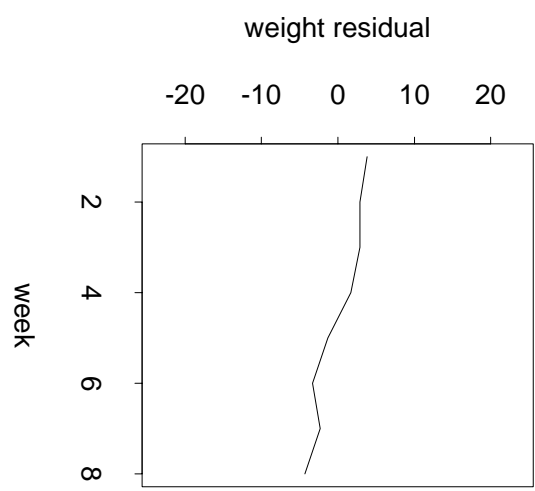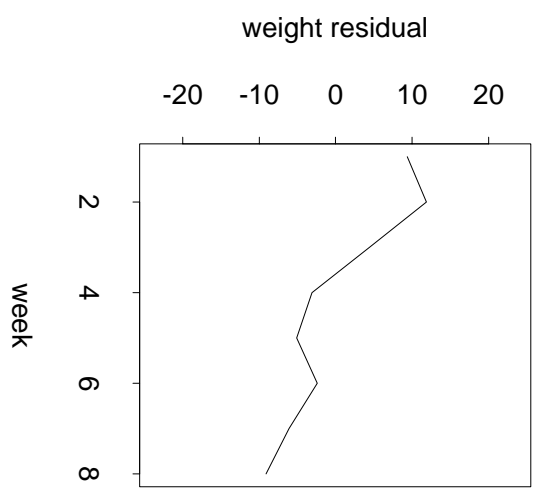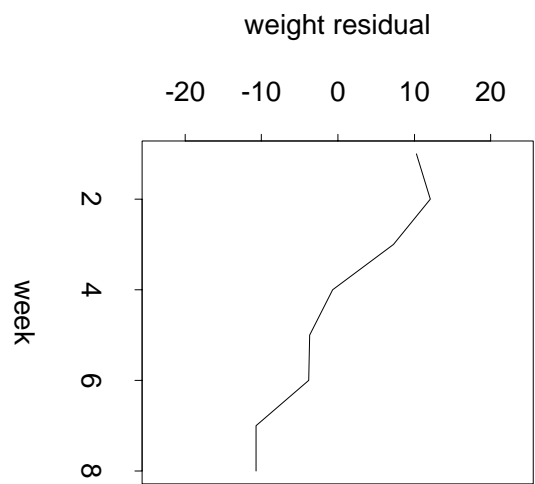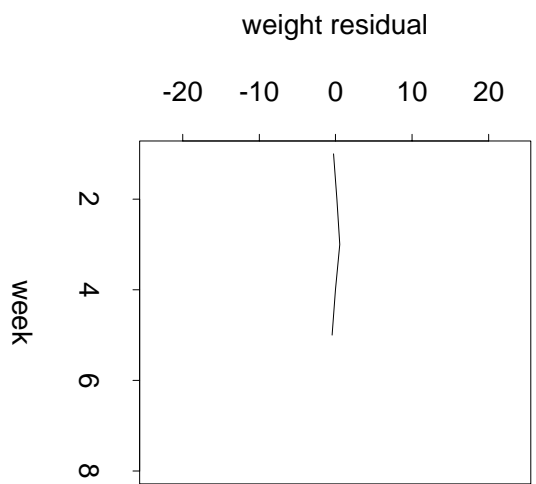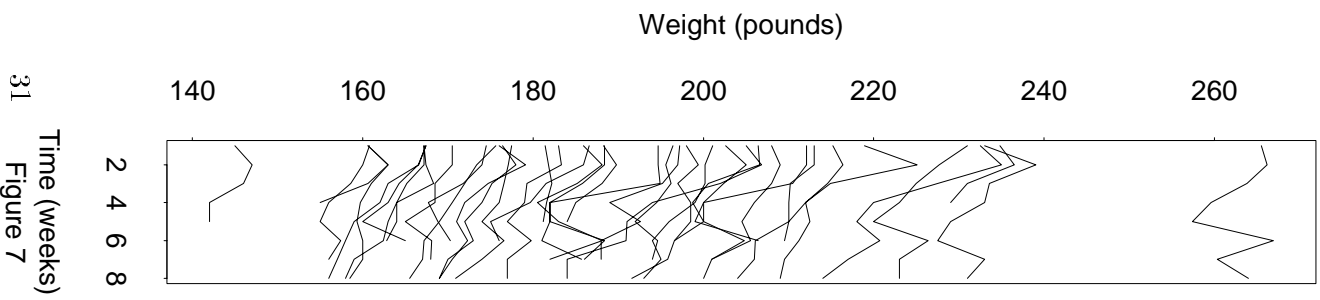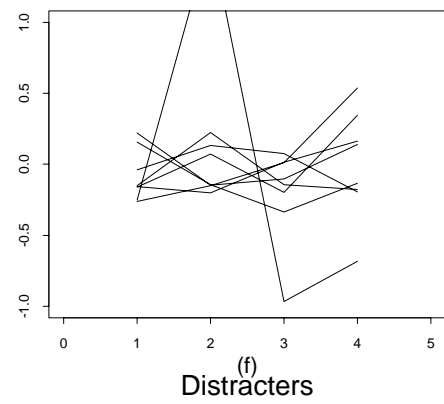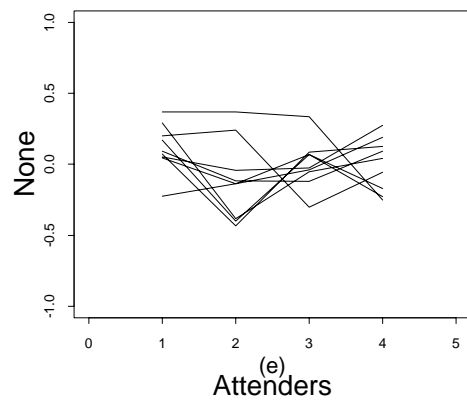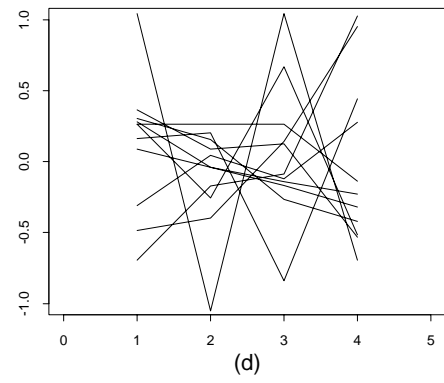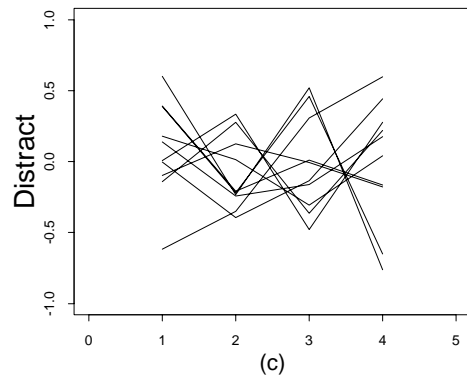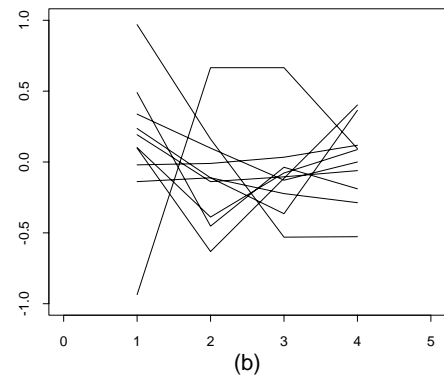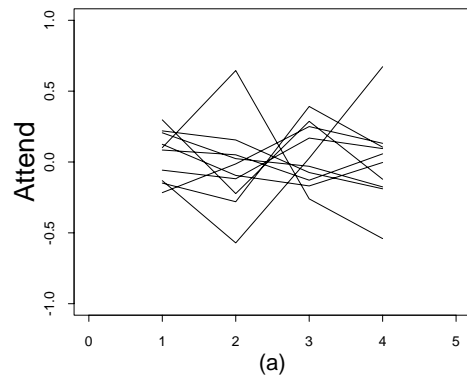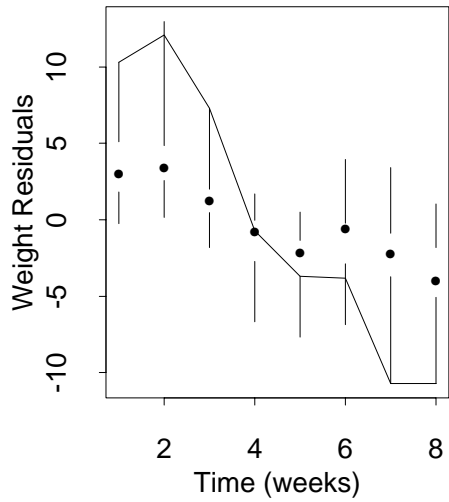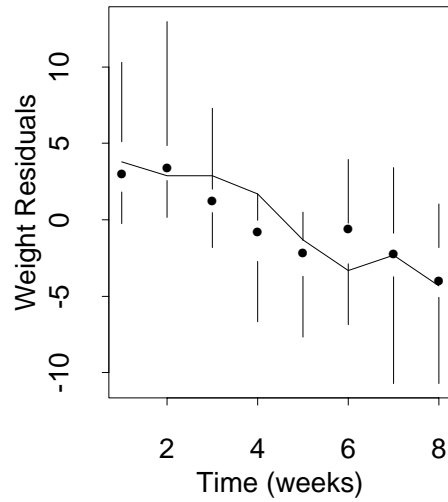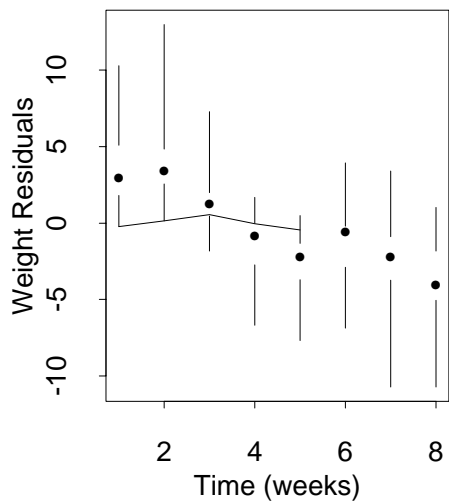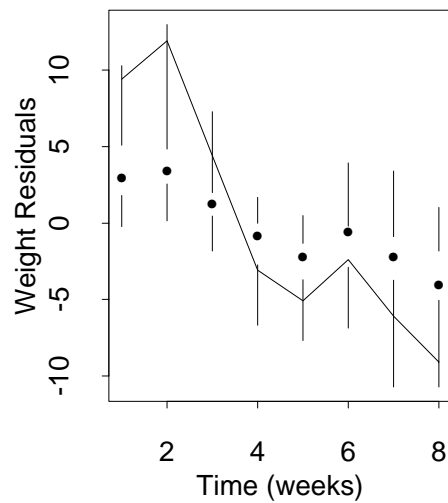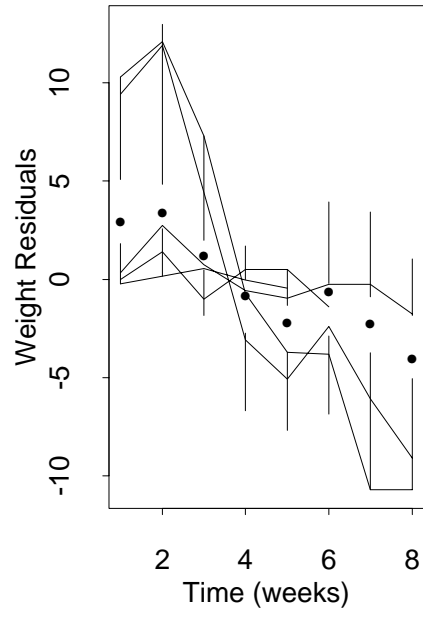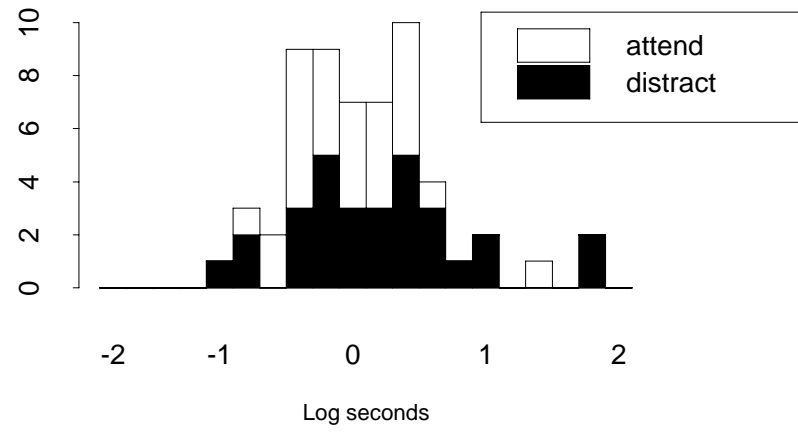
Figure 9
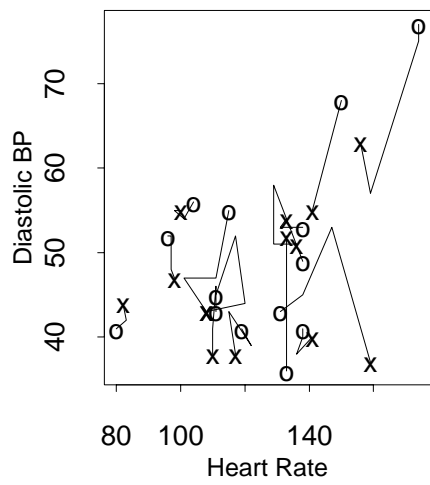
Figure 10

Figure 11

(a)

(b)

Figure 12

Figure 13

Figure 14