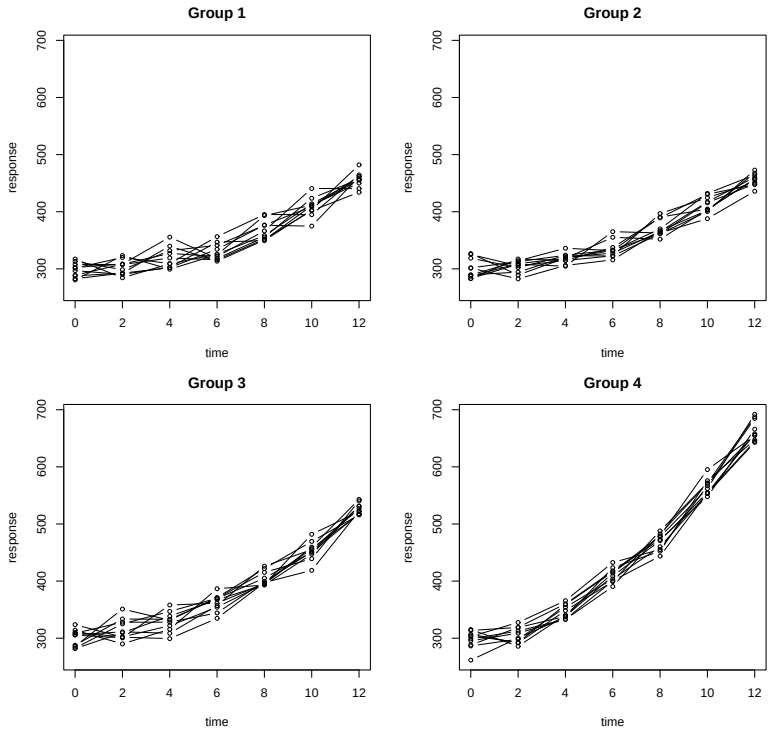
.

ST 732, HOMEWORK 3, SPRING 2007

1. A study was conducted in which $m = 40$ devices were randomized to be operated under 4 different sets of conditions, 10 devices per set of conditions. A response reflecting "performance level" of such devices was measured on each device at time 0, i.e., at "baseline," right before the devices were subjected to their assigned conditions. (The smaller the performance response, the better the performance of the device.) The conditions were then implemented, and the performance each device was measured every two hours thereafter over a 12 hour period. Spaghetti plots for the devices from each of the 4 groups are shown in Figure 1. Figure 2 on the next page shows the sample means for each condition group at each observation time using the group number as the plotting symbol.

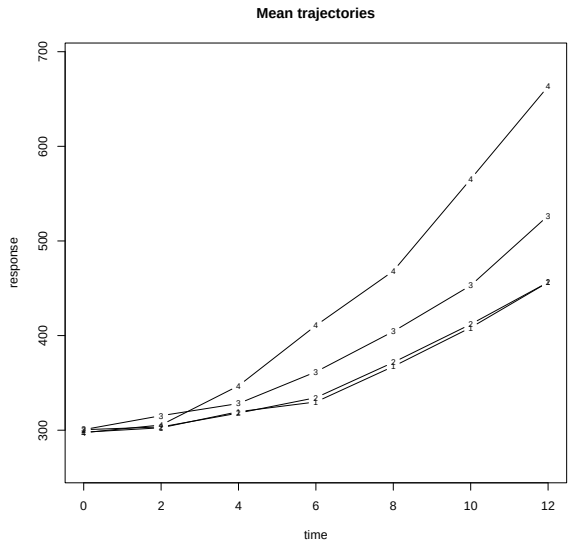
Figure 1: *Performance data for devices under 4 different conditions.*



(a) The experimenters expected that all devices would show degradation of performance over the study period, theorizing that this degradation would accelerate more profoundly as the severity of the conditions increased. In particular, they thought that this behavior would be reflected in mean performance profiles that exhibit upward "curvature" to varying degrees as the study progressed.

Let t_{ij} denote the jth time of observation t_{ij} (hours) for the ith device, and let Y_{ij} denote the corresponding performance response, where $i = 1, \ldots, m = 40$. Based on the investigator's expectations and the visual evidence, write down an appropriate model of the form of those for

Figure 2: *Means of performance data for devices under 4 different conditions.*



Y_{ij} in Chapter 8 of the notes that allows (i) allows for the possibility that the mean response prior to the devices being subjected to their assigned conditions (at baseline) is different in each group and (ii) allows the mean response trajectories during the study period for each group to exhibit "curvature" over time in a way that may have different features for each group. (That is, write a model like that in equation (8.10) that reflects the investigator's expectations and the features suggested by the plots).

(b) Let $\boldsymbol{Y}_i$ be the data vector for the ith device in the study. Express your model in (a) as a model for $\boldsymbol{Y}_i$ in terms of a design matrix $\boldsymbol{X}_i$ and parameter vector $\boldsymbol{\beta}$, defining each. Give the form of $\boldsymbol{X}_i$ for devices in each of the 4 groups.

(c) The experimenter's first question was whether the study was carried out properly in the sense that the devices were similar on average at baseline, as would be the case in a randomized study like this. In the context of your model in (a) and (b), write down a set of null and alternative hypotheses that addresses this question. Express your hypotheses in terms of $\boldsymbol{L\beta}$ for some appropriate matrix $\boldsymbol{L}$, giving the form of $\boldsymbol{L}$.

(d) The next question was whether all groups tend to have degradation of performance that occurs at constant rates (that are possibly different in each group) or whether at least one of the groups exhibits an "acceleration" of degradation of performance on average over the study period (whose nature is possibly different in each group). In the context of your model in (a) and (b), write down a set of null and alternative hypotheses that addresses this question. Express your hypotheses in terms of $\boldsymbol{L\beta}$ for some appropriate matrix $\boldsymbol{L}$, giving the form of $\boldsymbol{L}$.

(e) Another question was whether the mean performance profiles show an identical pattern of change **after baseline** across all four groups, regardless of whether the mean performance was the same in the four groups. Assuming your model in (a) and (b), write down a set of null and alternative hypotheses that addresses this issue. Express your hypotheses in terms

of $\boldsymbol{L\beta}$ for some appropriate matrix $\boldsymbol{L}$, giving the form of $\boldsymbol{L}$.

(f) Group 1 represented "mild" operating conditions, while group 4 represented "extreme" operating conditions, with groups 2 and 3 intermediate. The experimenters wanted to know whether, by 6 hours of operation, the mean performance is different among the four groups. Assuming your model in (a) and (b), write down a set of null and alternative hypotheses addressing this issue. Express your hypotheses in terms of $\boldsymbol{L\beta}$ for some appropriate matrix $\boldsymbol{L}$, giving the form of $\boldsymbol{L}$.

2. A study was conducted in which the response was to be measured on each subject at time 0 (baseline), and then subjects were to return at 2, 4, 8, 12, and 24 months post-baseline for additional responses to be ascertained. Thus, let $(t_1, t_2, t_3, t_4, t_5, t_6) = (0, 2, 4, 8, 12, 24)$ be the intended times of observation for each subject. Consider a subject who is missing responses at months 4 and 12.

   (a) Letting j index **the number of observations** for such a subject i, give the value of n_i, the number of observations actually available for i, and, with this indexing, write down the vector $\boldsymbol{Y}_i$ of observed responses for such a subject and times t_{ij} corresponding to each element of $\boldsymbol{Y}_i$.

   (b) Now let j index the **intended** times of observation. With this alternative indexing scheme, write down the vector $\boldsymbol{Y}_i$ of observed responses for such a subject and the vector of times corresponding to each element of $\boldsymbol{Y}_i$.

   (c) Using the indexing scheme in (b), write down the covariance matrix of $\boldsymbol{Y}_i$, $\boldsymbol{\Sigma}_i = \text{var}(\boldsymbol{Y}_i)$, in terms of the variances of and covariances among pairs of elements of $\boldsymbol{Y}_i$.

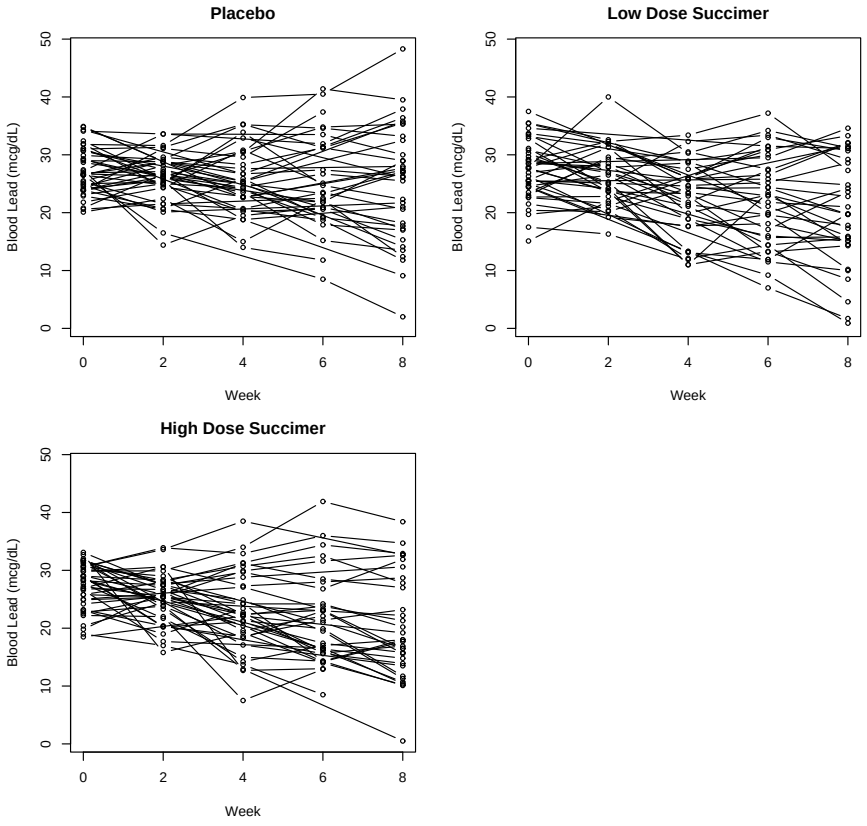
   (d) Suppose that we wish to assume a Markovian covariance structure with the same variance at all time points for any subject. Write down the form of $\boldsymbol{\Sigma}_i$ in this case, defining any symbols you use.

3. Exposure to lead can produce a variety of adverse health effects in infants and children, including hyperactivity, hearing or memory loss, learning disabilities, and damage to the nervous system. Although the use of lead as a gasoline additive has been discontinued in the US, so that airborne lead levels have been reduced dramatically, a small percentage of children continue to be exposed to lead at levels that can produce such health problems. Much of this exposure is due to deteriorating lead-based paint that may be chipping and peeling in older homes. Lead-based paint in housing was banned in the US in 1978; however, many older homes (built pre-1978) do contain lead-based paint, and chips and dust can be ingested by young children living in these homes during normal teething and hand-to-mouth behavior. This is especially a problem among children in deteriorating, inner-city housing. The US Centers for Disease Control and Prevention (CDC) has determined that children with blood levels above 10 micrograms/deciliter (μg/dL) of whole blood are at risk of adverse health effects.

   Luckily, there are so-called chelation treatments that can help a child to excrete the lead that has been ingested. The researchers were interested in evaluating the effectiveness of one such chelating treatment, succimer, in children who had been exposed to what the CDC views as dangerous levels of lead. They conducted the following study. 120 children aged 12–36 months with confirmed blood lead levels of > 15 μg/dL and , 40 mug/dL in a large, inner-city housing project were identified; these lead levels are above the at-risk threshold determined by the CDC. A clinic was set up in the housing project staffed by personnel from the city's Department of Public Health. The personnel randomized the children into three groups: 40

children were assigned at random to receive a *placebo* (an inactive agent with no lead-lowering properties), 40 children were assigned at random to receive a low dose of succimer, and 40 children were assigned at random to receive a higher dose of succimer. Blood lead levels were measured at the clinic for each child at baseline (time 0), prior to initiation of the assigned treatments. Then, assigned treatment was started, and, ideally, each child was to return to the clinic at weeks 1, 2, 4 and 8. At each visit, blood lead level was measured for each child.

Figure 3: *Blood lead levels for three groups of children*



The data are available in the file `lead.dat` on the class web page. The data are presented in the form of one data record per observation; the columns of the data set are as follows:

1 Child id
2 Indicator of age (= 0 if $\leq$ 24 months; = 1 if $>$ 24 months)
3 Gender indicator (= 0 if female, = 1 if male)
4 Week
5 Blood lead level (μg/dL)
6 Treatment indicator (= 0 if placebo, = 1 if low dose, = 2 if higher dose)

You will notice that, although all children were observed at baseline, some children are missing some of the intended subsequent lead level measurements. This might be because was

because some children were unable to come to the clinic for an assigned visit because their caregiver was unable to bring them. The investigators interviewed these children's caregivers and felt comfortable assuming that the the inability of some children to show up for some visits was not related to which treatment they were taking or how they were doing on their assigned treatment. As we will discuss later in the course, the validity of an analysis may be compromised if missingness is related to the thing under study in certain ways.

The investigators had several questions of interest. Broadly stated, the primary focus was on whether succimer, in either low- or high-dose form is effective over an eight week period in reducing blood lead levels in this population of children. They were also interested in whether blood lead levels in this population are associated with the age and/or gender of the child, and whether the effectiveness of succimer in reducing blood lead levels is associated with either or both of these factors. They postulated the following model.

Let Y_{ij} denote the jth lead level measurement on the ith child at time t_{ij} for that child, $j = 1, \ldots, n_i$. Note that the t_{ij} for each child and n_i may be different. Define $a_i = 0$ if subject i's age is ≤ 24 months and $a_i = 1$ if age is > 24. Let g_i indicate the gender of child i ($g_i = 0$ if female, $=1$ if male). The model they considered is

$$\begin{aligned}
Y_{ij} &= (\beta_0 + \beta_{0a}a_i + \beta_{0g}g_i + \beta_{0ag}a_ig_i) + (\beta_1 + \beta_{1a}a_i + \beta_{1g}g_i + \beta_{1ag}a_ig_i)t_{ij} + \epsilon_{ij} \text{ placebo} \\
Y_{ij} &= (\beta_0 + \beta_{0a}a_i + \beta_{0g}g_i + \beta_{0ag}a_ig_i) + (\beta_2 + \beta_{2a}a_i + \beta_{2g}g_i + \beta_{2ag}a_ig_i)t_{ij} + \epsilon_{ij} \text{ low dose} \\
Y_{ij} &= (\beta_0 + \beta_{0a}a_i + \beta_{0g}g_i + \beta_{0ag}a_ig_i) + (\beta_3 + \beta_{3a}a_i + \beta_{3g}g_i + \beta_{3ag}a_ig_i)t_{ij} + \epsilon_{ij} \text{ high dose}
\end{aligned} \tag{1}$$

Note the following features of model (1):

- For each group, there is a common "intercept" term

  $$\beta_0 + \beta_{0a}a_i + \beta_{0g}g_i + \beta_{0ag}a_ig_i$$

  such that mean lead level equals this at week 0. Thus, the model assumes that mean lead level at baseline (before treatment) in this population of children may be associated with age and gender of the child in a way that is different depending on the age-gender combination. Moreover, because the term is identical in all groups, it assumes that these features of baseline mean lead level are **same** in all three groups.

- Similarly, for each group k, there is a "slope" term multiplying week given by

  $$\beta_k + \beta_{ka}a_i + \beta_{kg}g_i + \beta_{kag}a_ig_i.$$

  Thus, the model assumes that the way in which mean lead level changes after baseline (i) possibly depends on the age and gender of the child in way that is different depending on the age-gender combination; and (ii) the "slope" and the way it depends on age and gender are possibly different for each group.

Thus, this is a rather complicated model!

You will want to run one SAS program repeatedly, adding on to the program and rerunning it as you work through the following problems. Turn in the final program that does everything required and its output. Use `PROC MIXED` with the `ML` method to carry out all model fits using the `ML` method. *Whenever you carry out a hypothesis test, be sure to state the level of significance at which you are conducting the test.*

(a) Why do you think that the investigators assumed that the "intercept" is the same in all three groups? Explain.

(b) The first step in the analysis is to investigate the nature of variation, as the validity of subsequent inferences depends on an having an appropriate model for covariance structure. Fit model (1) under the assumption that the covariance matrix of a data vector for a given subject is completely unstructured and **different** in each treatment group. **BE CAREFUL**, as the data are **not balanced** because of the missing values. Have your program print out the estimated covariance and correlation matrix for child 1 (complete data) and child 3 (missing data) in group 1 (placebo) to verify that the imbalance is correctly taken into account. Print out the estimated covariance and correlation matrix for child 44 in group 2 (low dose) and child 82 in group 3 (high dose) as well.

Also fit the model assuming a **common** completely unstructured covariance matrix, and print out the estimated covariance and correlation matrices for the above subjects.

Hint: It will take some figuring to determine an appropriate `model` statement to fit (1). The considerations in Section 8.9 apply, but are harder. You can make things easier by noting that, because age and gender are binary variables coded as 0-1, you can just leave them as numeric variables (do not include them in the `class` statement). The treatment indicator can be declared to be a `class` variable. Try applying the rules in Section 8.9 under these conditions to arrive at your `model` statement.

Based on your subjective inspection of these estimated correlation matrices, which covariance models do you think are likely candidates?

(c) To investigate the covariance assumption more formally, fit (1) assuming the following covariance models (again, **BE CAREFUL** to account for the missingness):

(i) Independence in both groups with the same variance (this may be accomplished by not including a `repeated` statement at all)

(ii) Homogeneous compound symmetry same in both groups and then different in both groups

(iii) Homogeneous one-dependent same in both groups and then different in both groups

(iv) Homogeneous AR(1) same in both groups and then different in both groups

Make a table of AIC and BIC values for these models and those in (b). Based on these results, select the model for which you think the evidence in the data is strongest, explaining your answer. *Adopt the covariance model you think is best for the rest of the problem.*

(d) Previous research on lead exposure says that while age of a child may be implicated in lead levels (older children have had more time to be exposed), gender of the child is not associated with lead levels in this population of children, nor is it associated with the progression of lead levels over time, with or without treatment. Based on these results, the investigators were hoping to simplify model (1) to have no dependence on gender in either the "intercept" or "slope" terms. Write down an appropriate null hypothesis that formalizes this in terms of the model in (1), and express it in the form $H_0 : \boldsymbol{L}\boldsymbol{\beta} = \mathbf{0}$, defining $\boldsymbol{\beta}$ and $\boldsymbol{L}$. Test your hypotheses two ways:

- Re-fit the model in (1) using your favored choice of covariance structure from (c) and include include a `contrast` statement corresponding to the matrix $\boldsymbol{L}$, using the `chisq` option to obtain the Wald test. Write down the value of the test statistic T_L and the associated p-value from the output.

- Fit the "reduced" model implied by your null hypothesis and, based on its output and that from the "full" model (1), obtain (by hand) the likelihood ratio test statistic T_{LRT}. State the degrees of freedom for this test statistic and the appropriate χ^2 critical value from your favorite χ^2 table or the p-value from your favorite software package.

In each case, state the conclusion you draw from the test. Is there sufficient evidence in these data to refute the investigator's claim?

(e) Write down the "reduced" model implied by H_0 in (d). Based on a fit of this model, do whatever you think necessary to address in the context of this model the question of whether there is an association of mean lead level with age in this population of children (i.e., at baseline). (That is, state appropriate null and alternative hypotheses and provide numerical evidence in the form of a test statistic and p-value, interpreting these.) Give estimates (and associated standard errors) for the mean lead level for younger ($\leq$ 24 months old) and for older ($>$ 24 months) children.

(f) Is there evidence to suggest that the age of a child is associated with the change in mean lead level for any of the treatments? Based on a fit of the model in (e), do whatever you think necessary to address this question. (That is, state appropriate null and alternative hypotheses and provide numerical evidence in the form of a test statistic and p-value, interpreting these.)

(g) Based on the fit of whatever model you feel is most relevant, address the general question: Is there a difference in rate of change of mean lead level over time among the three groups? (That is, state appropriate null and alternative hypotheses and provide numerical evidence in the form of a test statistic and p-value, interpreting these.)

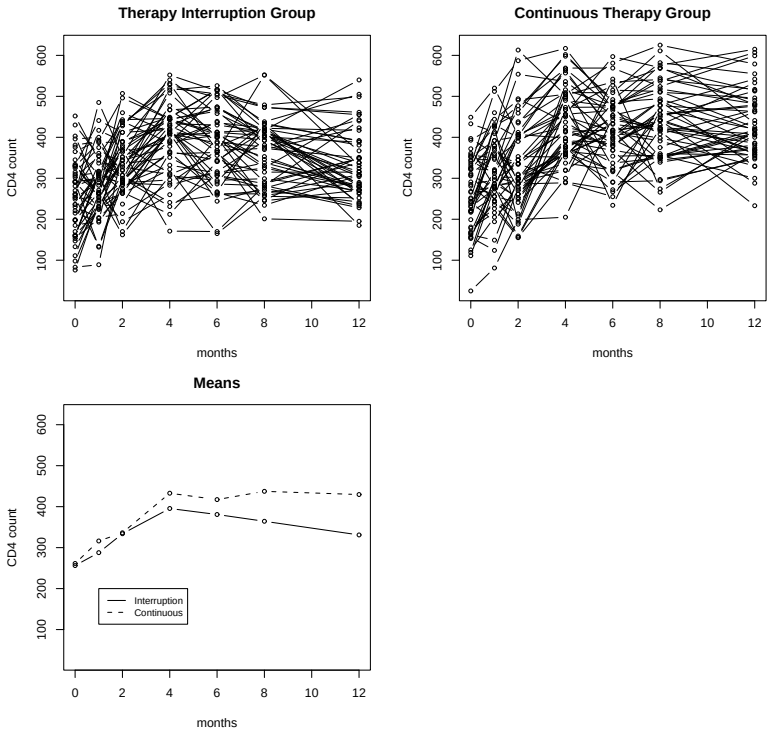
4. *Spline model.* A topic of considerable recent interest in HIV research is whether placing HIV infected subjects on antiretroviral therapy and then "interrupting" the therapy can improve prognosis. A hypothesis is that exposure to treatment followed by withdrawal of treatment can help to "train" the immune system to mount a better defense against the virus on its own. A further benefit of treatment interruption is that it provides patients a break from the side effects, costs, and burden of these therapies.

   The HIV virus attacks the immune system, so prognosis is often gauged by measuring blood constituents such as CD4 count that reflect immune system status (the higher the CD4 count, the better; CD4 counts below 1000 cells/mm^3 suggest that the immune system is somewhat compromised). A study was conducted involving 100 HIV-infected subjects randomized to two groups of 50 subjects each (the "therapy interruption" group and the "continuous therapy" group). All subjects were chronically infected, meaning that they had been infected with HIV for a long time, and had CD4 counts of less than 500 cells/mm^3. All subjects were placed on antiretroviral therapy at baseline (0 months) and followed for 4 months, with CD4 count recorded at 1 and 2 months. At month 4, subjects in the therapy interruption group withdrew from therapy, while subjects in the continuous therapy group continued to take the antiretroviral treatment. All subjects were then followed for an additional 8 months, with CD4 counts measured at months 6, 8, and 12 months post-baseline.

   Here, then, the patients in each group were treated the same up until month 4, at which point they were treated differently. The data from the study are plotted in Figure 4. Miraculously, all subjects showed up for their assigned visits, so there were no missing data.

   These data may be found in the file `interrupt.dat` on the class web page. The data are in the format with one observation per line; the columns are (1) subject id, (2) time in months, (3)

Figure 4: *CD4 count profiles for 100 subjects*



CD4 count (cells/mm^3), and (4) group indicator (1 = therapy interruption, 2 = continuous therapy).

The investigators were interested in understanding whether there were differences in the pattern of CD4 counts between the two groups. Because subjects were treated the same until month 4, at which time therapy was withdrawn from subjects in the interruption group but not in the continuous group, these questions were a bit different from those we might ordinarily see:

- Intuition would suggest that the CD4 patterns for both groups prior to month 4 should be identical, given that all subjects were treated the same. Is there evidence in the data to contradict this?
- One would expect differences in growth patterns to emerge after month 4,. Was there an effect of interrupting therapy relative to continuing it (the investigators were hoping that this would not be the case)? For subjects whose therapy was interrupted, was the pattern different after 4 months from the way it was before 4 months? How about for the continuous therapy subjects?

From the plot of the mean profiles, it appears that it may be reasonable to assume that the mean response pattern up to the end of month 4 can be represented by a straight line in each group; because the groups were treated the same, this straight line might be the same in each group (the plot certainly supports this). After month 4, the mean response seems to "shift" in both groups. In the interruption group, it shows a steady decline that could

follows a rough straight-line trajectory. In the continuous therapy group, it seems to "level off," which could be reasonably represented by a (possibly horizontal) straight line. Thus, a reasonable model for the mean profile in either group might be a straight line with a certain slope up to month 4, at which point it shifts to another straight line with a possibly different slope. Furthermore, these latter slopes may differ between groups.

A reasonable model for mean response for each group that tries to describe these features is that of a straight line from baseline to month 4 connected to another straight line, with possibly different slope, from month 4 to month 12. A model with this feature is a so-called "linear spline model" with one "knot" at month 4. Letting Y_{ij} be the CD4 count for the ith subject at the jth time of observation, a general form of such a model is

$$\begin{aligned} Y_{ij} &= \beta_{01} + \beta_{11} t_{ij} + \beta_{21}(t_{ij} - 4)_+ + \epsilon_{ij}, \quad i \text{ from group 1,} \\ &= \beta_{02} + \beta_{12} t_{ij} + \beta_{22}(t_{ij} - 4)_+ + \epsilon_{ij}, \quad i \text{ from group 2.} \end{aligned} \tag{2}$$

In (2), the notation "x_+" has a special meaning; in particular,

$$\begin{aligned} x_+ &= x \quad \text{if } x \geq 0 \\ &= 0 \quad \text{if } x < 0. \end{aligned}$$

Thus, if t is time in months, $(t-4)_+$ is equal to 0 for $t = 0, 1, 2, 4$ and equal to 2,4,8 for $t = 6, 8, 12$, respectively. Focusing on group 1 for definiteness (the other group is similar), note then that before month 4, the model says that mean response follows the straight line

$$\beta_{01} + \beta_{11} t$$

with slope β_{11}. For $t \geq 4$, mean response looks instead like

$$\beta_{01} + \beta_{11} t + \beta_{21}(t - 4),$$

which may be rewritten, by adding $\beta_{11}(4) - \beta_{11}(4)$ (which is equal to zero and does not change anything),

$$\{\beta_{01} + \beta_{11}(4)\} + (\beta_{11} + \beta_{21})(t - 4). \tag{3}$$

If we considered month 4 to be the "origin" for time, with time subsequent measured as time since month 4, i.e. $(t-4)$, (3) represents a straight line with "intercept" (value at origin 4 months) of $\beta_{01} + \beta_{11}(4)$ and "slope" $(\beta_{11} + \beta_{21})$. Note that at month 4 both the pre- and post-month-4 models have value of the mean response equal to this "intercept." Note also that if $\beta_{21} = 0$, then (3) reduces to the same straight line model prior to month 4 (slope does not change after month 4).

Thus, we see that (2) gives a model for each group where the mean response is a straight line with some slope up to month 4 and then changes to a straight line with possibly different slope at month 4, where the change in slope is governed by the "β_{2k}" terms in the model for each group $k = 1, 2$.

Such a model is referred to as a "spline" model. In general, such models are formed by linking together several different functions at each of several "knots" (time points), thus allowing the mean response to exhibit rather complicated patterns over time. They are especially handy in situations like this one, where mean response is expected to change behavior at certain, known time points (at least in group 1). In (2), the spline is "linear" because the functions

are straight lines, and there is one "knot," at month 4. Such functions are also often called "piecewise linear," for obvious reasons.

Note that (2) allows the possibility that the first straight line, prior to month 4, is different across groups, although we might expect otherwise, as all subjects are treated identically up to that point.

You will want to run one SAS program repeatedly, adding on to the program as you work through the problems below. Turn in your final, complete program (that carries out all necessary calculations) and its output along with your interpretation. Use `proc mixed` with the method of **restricted maximum likelihood** to carry out all model fits.

(a) To obtain valid inferences, it is of course necessary to fit the model under an appropriate assumption on the covariance matrix of a data vector. To begin investigating this, fit model (2) under the assumption that the covariance matrix of a data vector for a given subject is completely unstructured but identical for all subjects, regardless of group. Obtain the estimated covariance and correlation matrices implied under this assumption, and write out the latter. Based on your subjective inspection of these estimated matrices, which covariance models do you think are likely candidates to represent the true pattern of variation in a data vector?

Note: To fit the model, you will need to represent it in a SAS `model` statement. To do this, you will want to create a new variable corresponding to $(t-4)_+$. This may be accomplished in a `data` step with the following two lines of code:

```
monthplus = month-4; if monthplus<0 then monthplus=0;
```

You may then treat `monthplus` as simply another covariate in the model. It would be wise to use `proc print` at some point to print the contents of the data set you create to be sure you have formatted things and defined `monthplus` correctly.

(b) To investigate the covariance assumption more formally, fit (2) assuming the following covariance models (assume covariance is the same regardless of group and consider the homogeneous versions of these models. Note that the data are unequally spaced but balanced):

(i) Independence

(ii) Compound symmetry

(iii) Markov

From the results of these fits, make a table of the values of the AIC and BIC criteria, including that for the unstructured matrix from (a). Based on these criteria, for which of the covariance models in (a) or in (i)–(iii) above does the evidence in the data appear strongest? Explain why you selected the model you did on the basis of AIC and BIC values.

For (c)–(i) below, use the covariance structure you selected in (b). In each case, state the level of significance you choose to conduct the test.

(c) Because the 100 subjects were randomly chosen from the same population (subjects with CD4 count < 500) prior to the start of the study and were treated the same up to month 4, we would expect that (i) the (true) means at month 0 would be the same, and (ii) the (true) pattern of change up to month 4 would be the same in both groups. In terms of model (2), write down a set of null and alternative hypotheses addressing this issue. Express the null hypothesis in the form $H_0 : \boldsymbol{L}\boldsymbol{\beta} = \mathbf{0}$, giving the form of $\boldsymbol{\beta}$ and $\boldsymbol{L}$. Assuming the covariance

structure you identified in (b), re-fit model (2). To test H_0, include a `contrast` statement corresponding to the matrix $\boldsymbol{L}$, using the `chisq` option to obtain the Wald test. Write down the value of the test statistic T_L and the associated p-value from the output. Is there sufficient evidence in these data to conclude that the mean profiles up to the end of month 4 differ?

(d) *Regardless of your answer to (c)*, consider the "reduced" model in (c), where the month 0 means and patterns of change up to 4 mnths are assumed identical in both groups. Fit this model using your chosen covariance structure. For each group, (i) use an `estimate` statement to obtain an estimate of the "second phase" slope (slope of the straight line describing mean CD4 after the end of month 4) for the group and a suitable standard error, and (ii) make a spaghetti plot of the raw data using your favorite plotting software, and superimpose the fitted spline model for the group on the plot.

(e) Consider the same (reduced) model in (d). Under this model, is there evidence to support the contention that the slope of the straight line mean profile prior to the end of month 4 for subjects in the interruption group differs from that for the straight line mean profile after the end of week 4? In terms of the reduced model, write down a set of null and alternative hypotheses addressing this question, expressing the null hypothesis in the form $H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$, giving the form of $\boldsymbol{\beta}$ and $\boldsymbol{L}$. Fit the model including a `contrast` statement corresponding to the matrix $\boldsymbol{L}$, using the `chisq` option to obtain the Wald test. Write down the value of the test statistic T_L and the associated p-value from the output. Is there sufficient evidence in these data to conclude that pre- and post-month 4 slopes differ for subjects whose treatment was interrupted?

(f) Consider the same (reduced) model in (d). Under this model, is there evidence to support the contention that the slope of the straight line mean CD4 profile prior to the end of month 4 differs from that for the straight line mean profile after the end of month 4 in at least one of the groups? In terms of the reduced model, write down a set of null and alternative hypotheses addressing this question, expressing the null hypothesis in the form $H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$, giving the form of $\boldsymbol{\beta}$ and $\boldsymbol{L}$. Fit the model including a `contrast` statement corresponding to the matrix $\boldsymbol{L}$, using the `chisq` option to obtain the Wald test. Write down the value of the test statistic T_L and the associated p-value from the output. Is there sufficient evidence in these data to conclude that pre- and post-month 4 slopes differ for at least one of the groups?

(g) Consider the same (reduced) model as in (d). Under this model, is there evidence to support the contention that the slopes of the straight lines that describe the mean response after month 4 differ across the groups? In terms of this reduced model, write down a set of null and alternative hypotheses addressing this question, expressing the null hypothesis in the form $H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0}$, giving the form of $\boldsymbol{L}$. Include in the fit of this model a `contrast` statement corresponding to the matrix $\boldsymbol{L}$, using the `chisq` option to obtain the Wald test. Write down the value of the test statistic T_L and the associated p-value from the output. Is there sufficient evidence in these data to conclude that post-month-4 slopes differ?

(h) Consider the same (reduced) model as in (d). Use an `estimate` statement to obtain an estimate of mean CD4 count for subjects in the continuous therapy group at 12 months post-baseline. Give also a standard error associated with this estimate.