1. Consider a straight line model for individual behavior as in Equation (9.1) of the notes, which for unit \( i \) is of the form

\[
Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij},
\]

where \( Y_{ij} \) is the random variable representing the observation that might be seen for unit \( i \) at time \( t_{ij} \); \( j = 1, \ldots, n_i \) indexes the time points for unit \( i \); \( \beta_{0i} \) and \( \beta_{1i} \) are the unit-specific intercept and slope, respectively, dictating the “inherent trajectory” for unit \( i \); and \( e_{ij} \) is a mean-zero random deviation representing how \( Y_{ij} \) deviates from the inherent trajectory. Let \( \beta_i = (\beta_{0i}, \beta_{1i}) \) be the vector of unit-specific parameters for individual \( i \) in model (1), and let \( Y_i = (Y_{i1}, \ldots, Y_{in_i})' \) denote the random vector of observations on \( i \), with \( e_i \) defined similarly.

(a) If we write (1) in the form \( Y_i = Z_i\beta_i + e_i \), give the form of \( Z_i \) if \( n_i = 4 \).

(b) Now suppose that units arise from 3 populations, labeled \( A, B, \) and \( C \). Write down a second-stage population model that allows each population to have its own mean intercept and slope \( \beta_{0,k} \) and \( \beta_{1,k} \), respectively, where \( k = A, B, \) or \( C \) about which unit-specific intercepts and slopes vary in each population. Express your model in the form in Equation (9.5) in the notes; i.e.,

\[
\beta_i = A_i\beta + b_i,
\]

where \( \beta = (\beta_{0,A}, \ldots, \beta_{0,C}, \beta_{1,A}, \ldots, \beta_{1,C})' \). Define \( b_i \) and give the form of \( A_i \) when unit \( i \) is from each of populations \( A, B, \) and \( C \), respectively.

(c) As shown on pages 321–322 of the notes, the model under the conditions in (a) and (b) can be expressed as

\[
Y_i = X_i\beta + Z_i b_i + e_i.
\]

Give the form of \( X_i \) for a unit \( i \) with \( n_i = 4 \) if the unit is from populations \( A \) and \( C \), respectively.

2. Consider the random coefficient model with individual first stage model

\[
Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij},
\]

where individual \( i \) is observed at times \( t_{i1}, \ldots, t_{in_i}, \) \( e_i = (e_{i1}, \ldots, e_{in_i})' \), and

\[
\text{var}(e_i) = \sigma^2 I_{ni};
\]

and population second stage model

\[
\beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1 + b_{1i}, \quad b_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix},
\]

where

\[
\text{var}(b_i) = D = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix},
\]

and \( b_i \) is statistically independent of \( e_i \) as on p. 320 of the notes.
(a) Use results on variances and covariances covered earlier in the course to demonstrate that
\[
\text{var}(Y_{ij}) = D_{11} + D_{22}t_{ij}^2 + 2D_{12}t_{ij} + \sigma^2, \quad \text{cov}(Y_{ij}, Y_{ik}) = D_{11} + D_{22}t_{ij}t_{ik} + D_{12}(t_{ij} + t_{ik}),
\]
thus verifying a generalization of the result at the top of p. 329 of the notes.

(b) Suppose that \( D_{12} = 0 \), so that \( b_{0i} \) and \( b_{1i} \) are uncorrelated. Are \( Y_{ij} \) and \( Y_{ik} \) correlated under this condition? Explain.

(c) Suppose instead that \( \text{var}(e_i) = \sigma^2_{1} \Gamma_i + \sigma^2_{2} I_{n_i} \), where \( \Gamma_i \) is the \((n_i \times n_i)\) Markov correlation model with parameter \( \rho > 0 \). Find \( \text{var}(Y_{ij}) \) and \( \text{cov}(Y_{ij}, Y_{ik}) \) in this case, where all the other conditions given above still hold.

3. Recall the lead level study from Homework 3, Problem 3. Suppose that a new group of investigators studying treatment of lead exposure asked the original investigators for their data. This new group took a different approach to modeling these data. In particular, as an initial model, they ignored the age and gender variables and considered the random coefficient model with straight-line first stage

\[
Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}
\]

for child \( i \), where we may define \( \beta_i = (\beta_{0i}, \beta_{1i})' \) for child \( i \). They assumed that, for treatment \( k = 1, 2, 3 \), where \( k = 1 \) is placebo, \( k = 2 \) is low-dose succimer, and \( k = 3 \) is high-dose succimer, \( \beta_{0,k} \) is the “typical” mean value of intercepts \( \beta_{0i} \) and \( \beta_{1,k} \) is the “typical” mean value of slopes \( \beta_{1i} \) for children receiving treatment \( k \).

Define \( \beta = (\beta_{0,1}, \beta_{0,2}, \beta_{0,3}, \beta_{1,1}, \beta_{1,2}, \beta_{1,3})' \). Then the investigators assumed the second stage population model is

\[
\beta_i = A_i \beta + b_i, \quad b_i = (b_{0i}, b_{1i})',
\]
and \( A_i \) is the appropriate design matrix for child \( i \) that “picks off” the correct mean intercept and slope from \( \beta \) corresponding to the treatment \( i \) took.

The investigators were ultimately interested in learning whether the patterns of blood lead levels over the study period were different depending on treatment. In particular, they were interested in whether there is evidence that the “typical” mean slopes were not all the same.

(a) From the spaghetti plots shown in Homework 3, do you think that the assumption that blood lead levels for children in each treatment group follow “inherent trajectories” that may be represented by child-specific straight lines seems reasonable?

(b) The investigators were willing to assume the following:

(i) The assay used to ascertain blood lead levels from blood samples collected from the children committed errors whose magnitude is unrelated to the lead level in the sample being measured; and

(ii) Lead level samples were taken sufficiently far apart in times that correlation due to local within-child fluctuations in lead levels was negligible, and the magnitude of such fluctuations was constant over time for all treatments. The magnitudes of such fluctuations are independent of the magnitude of the true lead levels.

In developing their model further, the investigators wanted to investigate the following:

(iii) whether the magnitudes of within-child fluctuations in lead levels are the same for all treatments (they constant for all treatments, but are they the same?)
(iv) whether the way in which child-specific intercepts and slopes vary and co-vary are the same under the three treatments.

Using \textit{proc mixed}, fit using REML three different versions of the random coefficient model, all of which incorporate assumptions (i) and (ii) above but allow different assumptions about (iii) and (iv), namely:

- Magnitude of within-child fluctuations in lead level and the way child-specific intercepts and slopes vary/co-vary are both \textit{the same} under all three treatments.
- Magnitude of within-child fluctuations in lead levels are possibly \textit{different} under different treatments, but the way child-specific intercepts and slopes vary and covary is \textit{the same}.
- Magnitude of within-child fluctuations in lead levels is \textit{the same} under all treatments but the way in which child-specific intercepts and slopes vary/co-vary are possibly \textit{different}.
- \textit{Both} the magnitude of within-child fluctuations in lead levels and the way in which child-specific intercepts and slopes vary/co-vary are possibly \textit{different} across treatments.

(c) From inspection of \textit{AIC} and \textit{BIC} for each model fit, which set of assumptions on within-child fluctuations and among-child variation/covariation in intercepts/slopes do you prefer?

(d) Under the model that embodies the assumptions you chose in (c), is there evidence to suggest that the “typical” mean slopes of blood lead level patterns for the three treatments are not the same? To address this, include an appropriate \textit{contrast} statement in the fit of your preferred model and obtain the Wald test statistic. State the value of the statistic, the associated p-value, and your conclusion regarding the strength of the evidence supporting the contention that the mean slopes differ.

4. A common way of treating patients with cardiovascular disease is by surgical intervention. In particular, such patients may arrive at a hospital with symptoms such as unstable angina or suspected myocardial infarction (heart attack), requiring that physicians perform an invasive procedure called a percutaneous coronary intervention (PCI) to investigate the extent to which coronary arteries might be blocked. During such an investigation, the blockage may be treated using a balloon to dislodge the blockage and widen the artery (“balloon angioplasty”); in addition, a device known as a stent may be inserted to prop the artery open.

When such PCI procedures are performed, it is necessary for the subject to be treated with a drug that inhibits the aggregation of platelets in the blood. Informally, platelets are a blood constituent involved in clotting of the blood; clotting occurs when the platelets aggregate together in “clumps.” To ensure that clotting does not interfere with the procedure, inhibition of the clotting mechanism is desirable; clotting during the procedure can lead to complications such as stroke or heart attack. A long-standing issue has been to determine which of two popular drugs elicits the most desirable pattern of inhibition of platelet aggregation.

Accordingly, an experiment was conducted to compare the platelet aggregation patterns of the drugs in such subjects under controlled conditions. Subjects arriving at a major medical center with symptoms of unstable angina or myocardial infarction who were judged to require a PCI procedure were randomized into two groups, one for each of the drugs, with 200 subjects per group. For each subject, at time 0, the assigned drug was administered according to the manufacturer’s recommended dosage; for each drug, this involved giving the subject a large dose by injection to start inhibition of platelet aggregation immediately and simultaneously giving the subject a smaller dose of the drug intravenously at a constant rate over several hours, a method of administration known as an infusion. The purpose of the infusion was to
keep platelet aggregation inhibited over at least a 12 hour period, so that clotting would be 
minimized during the PCI procedure and subsequent recovery for the subject.

For each subject, blood samples were to be taken at 0.5, 2.0, 3.5, 5.0, 8.0, 11.0, and 12.5 
hours. Each sample was to be analyzed for degree of platelet inhibition, characterized by 
the response “percent inhibition,” a value between 0 and 100 representing the percentage of 
inhibition relative to that of an untreated sample (in units of “%µM”). Also recorded for each 
subject was whether the subject had experienced a previous myocardial infarction before the 
current hospitalization (0=no, 1=yes) and gender (0=female, 1=male).

The data from the study may be found on the class web page in the file platelet.dat. Each 
record in the file corresponds to a single observation, and the columns are (1) subject id 
number (1–400), (2) previous myocardial infarction indicator (0 = no, 1 = yes), (3) gender 
indicator (0 = female, 1 = male), (4) time (hours, measured since administration of drug), 
(5) percent inhibition, and (6) drug group indicator (1 or 2). Note that for some subjects, 
the response is not available at all intended time points; some samples were mishandled and 
in some instances study personnel did not follow the instructions and neglected to obtain 
samples. It was determined that the reasons for the missing values had nothing to do with 
the drugs or the patterns of inhibition. The data are depicted graphically in Figure 1.

Figure 1: Percent platelet inhibition for two groups of subjects with cardiovascular disease

From the plot, it appears that over the period of the study, platelet inhibition appears for 
most subjects to follow a rough straight-line trajectory that either stays relatively flat or 
rises, although a few profiles seem to decrease. To represent this, the investigators proposed 
the following model. Because the investigators were particularly interested in the time point 
0.5 hours post-administration, as we will see shortly, they defined time in the model so that 
t = 0 corresponds to 0.5 hours after administration of the drug. That is, they let \( t_{ij} \), the time 
of the \( j \)th platelet inhibition response on subject \( i \), be defined as

\[
t_{ij} = s_{ij} - 0.5,
\]
where \( s_{ij} \) = time of the \( j \)th response on subject \( i \) measured from administration of the drug (so \( s_{ij} \) equals the time value given in the data file). Here is the SAS code to read in the data and redefine the time variable hour:

```sas
data pci; infile ".... platelet.dat";
  input patient mi gender hour percent drug;
  hour=hour-0.5;run;
```

Letting \( Y_{ij} \) be the corresponding platelet inhibition response for subject \( i \) at the \( j \)th time, the model is

\[
Y_{ij} = \beta_0i + \beta_1i t_{ij} + e_{ij},
\]  
(2)

where the parameters \( \beta_0i \) and \( \beta_1i \) describe the percent inhibition trajectory starting at 0.5 hours following drug administration (\( t_{ij} = 0 \)) for the \( i \)th subject, and \( e_{ij} \) represents a mean-zero deviation associated with the \( j \)th inhibition response, assumed to be normally distributed. This model thus allows the pattern after 0.5 hours to follow a straight line for each subject.

(a) The investigators initially wished to assume that, for model (2), mean platelet inhibition at 0.5 hours following administration of drug has the following features within each drug group:

- is associated with whether the subject has had a previous myocardial infarction
- is associated with whether the subject is male or female
- the way in which it is associated with whether the subject is male or female is different depending on whether the subject has had a previous myocardial infarction.

Because the subjects had been on the drugs for 0.5 hours, the investigators assumed that mean platelet inhibition at 0.5 hours and the way the above features occur is different for the two drugs.

The investigators also wished to assume that the “typical” or mean rate of change of platelet inhibition over the study period also has these features, and they wished to allow for the possibility that the mean rate of change of platelet inhibition and its association with prior myocardial infarction and gender could be different for each drug. This would allow the possibility that the drug that is received is associated with the pattern of change of platelet inhibition in different ways for subjects of different genders and prior history of myocardial infarction. (See hint at the end of this question regarding the model statement.)

Let \( m_i = 0 \) if subject \( i \) has not had a previous myocardial infarction and \( m_i = 1 \) if s/he has, and let \( g_i = 0 \) if \( i \) is female, and \( g_i = 1 \) if \( i \) is male. Given these beliefs, write down expressions for \( \beta_0i \) and \( \beta_1i \) for subject \( i \) taking drug \( k \), \( k = 1, 2 \). Be sure to define and fully describe all additional symbols you use. (Hint: for \( \beta_0i \) and for each \( k \) there will need to be 4 fixed effect \( \beta \)'s multiplied appropriately by \( m_i \) and \( g_i \) to be added to the random \( b_{0i} \). And similarly 4 for each \( k \) for the slope, together resulting in a vector \( \beta \) of length 16 like

\[
\beta = (\beta_{10}, \beta_{20}, \beta_{10g}, \beta_{20g}, \beta_{10m}, \beta_{20m}, \beta_{10gm}, \beta_{20gm}, \beta_{11}, \beta_{21}, \beta_{11g}, \beta_{21g}, \beta_{11m}, \beta_{21m}, \beta_{11gm}, \beta_{21gm})'.
\]

(b) In terms of your model in (a)

(i) Give an expression that represents the typical value of platelet inhibition at 0.5 hours after drug administration for male subjects with a previous myocardial infarction taking drug 2.
(ii) Give an expression for mean platelet inhibition for female subjects with no previous myocardial infarction at 12 hours following administration of drug 1.

In the following parts, you will develop a SAS program to carry out several different analyses. You will have to modify your program for each part to obtain desired analyses. Please turn in your final program and output that carries out all necessary analyses.

(c) The investigators were all willing to believe that

(i) the assay used to measure platelet inhibition for both drug groups exhibits constant variation regardless of the true value of platelet inhibition being ascertained,

(ii) within-subject local “fluctuations” in platelet inhibition are of similar magnitude for both drugs and across time for all subjects,

(iii) variation in “inherent,” true platelet inhibition at 0.5 hours is similar for patients in both drug groups, as is variation in the “inherent” rates of change of platelet inhibition over the study period and the way these quantities co-vary.

One of the investigators was concerned, however, that the time points at which platelet inhibition was measured were not sufficiently far apart in time to ensure that measurements within a subject are uncorrelated. He was willing to believe that, if such correlation is present, it “falls off” as the time points get farther apart, but he insisted that an analysis be done to resolve this issue.

Give two different sets of assumptions on the $e_{ij}$, $i = 1, \ldots, n_i$, in (2) and random effects corresponding to $\beta_{0i}$ and $\beta_{1i}$ in (2) that incorporate (i)–(iii). The first set of assumptions should incorporate the investigator’s concern; the second set should represent the case where the investigator’s concern is unwarranted.

Fit the overall model (2) along with your model for $\beta_{0i}$ and $\beta_{1i}$ in (a) under both sets of assumptions using SAS proc mixed. Which set of assumptions is best supported?

(d) From the output for the fit of the model you preferred in (c), write down an estimate of the variance associated with among-subject variation in true platelet inhibition in the population of male subjects with no previous myocardial infarction receiving drug 2 at 0.5 hours post-administration.

(e) Previous research has suggested that the way in which platelet inhibition occurs for both drugs over this period may be associated with whether a subject has had a previous myocardial infarction, but there is no evidence to suggest that it is associated with gender in any way. Thus, the investigators planned to base their subsequent analyses not on the model you developed in (a) but on a model that includes no effect of gender either in the representation of mean platelet inhibition at 0.5 hours or in the representation of the “typical” rate of change of platelet inhibition over the study period. Write down this simpler model and fit it using ML and your preferred covariance structure from (c). Based on your preferred fit in (c) and this fit, is there any evidence against doing this? (If you used REML in (c), just rerun it here with method=ml.)

(f) For the rest of the problem, consider the simpler model in (e) with no gender effects. The reason that the investigators were so interested in 0.5 hours post-administration is because another research team had recently published a paper receiving a lot of press, which claimed that the 2 drugs exhibit the same mean platelet inhibition and that, furthermore, mean platelet inhibition on the two drugs is the same for subjects with or without a previous myocardial infarction. This team based their finding on comparing platelet inhibition levels 0.5 hours post-administration. Our investigators felt that comparing platelet inhibition at
a single time point, particularly one so soon after administration, was not very informative. However, their first goal was to examine whether the data from the current study offer evidence refuting the claim of their rival investigators.

Write down a set of hypotheses that addresses the issue of interest to the investigators in terms of the model in (e), and express your null hypothesis in terms of a linear function \( L\beta \), defining \( L \). Using Wald methods, carry out the test at level of significance 0.05 based on a REML fit. State your conclusion as a meaningful sentence. (This is pretty tricky to get through all the words. The hypothesis is simply that the 4 intercepts are all equal. So figure how to do get a 3 degree of freedom \( L\beta \) for that hypothesis and then translate it into the SAS code.)

(g) The investigators’ second goal was to make the point that comparing platelet inhibition at a single point does not tell the whole story. Thus, regardless of how the test in (f) turned out, they wanted to investigate longer time periods and the rate of change of platelet inhibition over them. The first question along these lines was whether the way “typical” rate of change differs between subjects who have had a previous myocardial infarction and those who have not is different for the two drugs. (This is an interaction!)

Write down a set of hypotheses that addresses the issue of interest to the investigators in terms of the model in (e), and express your null hypothesis in terms of a linear function \( L\beta \), defining \( L \). Using Wald methods, carry out the test at level of significance 0.05 based on a REML fit. State your conclusion as a meaningful sentence.

(h) The second question was whether, averaged across patients with and without a previous myocardial infarction, “typical” rate of change of platelet inhibition was different for the two drugs. Write down a set of hypotheses that addresses the issue of interest to the investigators in terms of the model in (e), and express your null hypothesis in terms of a linear function \( L\beta \), defining \( L \). Using Wald methods, carry out the test at level of significance 0.05 based on a REML fit. State your conclusion as a meaningful sentence. (So you have to average slopes over \( m_i = 0 \) and \( m_i = 1 \) for the drugs separately and then subtract them.)

(i) Based on the model in (e), provide estimates (and associated standard errors) of mean platelet inhibition at 12.5 hours after administration of (a) drug 1 in subjects with previous myocardial infarction; and (b) drug 2 in subjects with no previous myocardial infarction.

Note. For (c) there should be 8 different average lines (all combinations of 2 drugs, 2 genders, \( m_i=0 \) or 1). Three equivalent versions are:

class patient drug time;
model percent = drug drug*mi drug*gender drug*mi*gender drug*hour
drug*hour*mi drug*hour*gender drug*hour*mi*gender / noint solution ddfm=kr;

class patient drug time;
model percent = mi gender drug mi*gender mi*drug gender*drug mi*gender*drug
hour mi*hour gender*hour drug*hour mi*gender*hour mi*drug*hour
gender*drug*hour mi*gender*drug*hour / solution ddfm=kr;

drugn=drug-1;
class patient time;
model percent = mi gender drugn mi*gender mi*drug gender*drugn mi*gender*drugn
hour mi*hour gender*hour drugn*hour mi*gender*hour mi*drugn*hour
gender*drugn*hour mi*gender*drugn*hour / solution ddfm=kr;