.

ST 732, HOMEWORK 2, SPRING 2007

1. Suppose that we have a situation in which units have been randomized into $q = 4$ groups and each unit is observed at the same $n = 3$ times. As in the notation in the notes, let $\mu_{\ell j}$ be the mean for the ℓth group at the jth time.

   (a) In terms of the $\mu_{\ell j}$, write down the $\boldsymbol{M}$ matrix for this problem.

   (b) Suppose $\boldsymbol{Y}_i$ is the data vector for the ith unit. If we assume the model given in equation (5.11), write down the vector $\boldsymbol{a}_i$ appropriate for a unit in group 1.

   (c) Suppose $\mu_{\ell j} = \mu + \tau_\ell + \gamma_j + (\tau\gamma)_{\ell j}$. If we express the model in the alternative "regression form" given on page 119 of the notes, $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, write down the appropriate vector $\boldsymbol{\beta}$ and "design matrix" $\boldsymbol{X}_i$ for the ith unit in the 4th group.

   (d) Write down the appropriate matrix $\boldsymbol{U}$ that would be used in this situation to characterize a set of contrasts focusing on differences among the $n = 3$ times.

   (e) Write down the appropriate matrix $\boldsymbol{C}$ that would be used in this situation to characterize a set of contrasts focusing on differences among the $q = 4$ times.

   (f) Taking into account the usual constraints imposed on an overparameterized model like (5.11), write down the form of the null hypothesis that the mean profiles are parallel across the $q = 4$ groups in terms of μ, the τ_ℓ, the γ_j and the $(\tau\gamma)_{\ell j}$. Write down appropriate matrices $\boldsymbol{C}$ and $\boldsymbol{U}$ that could be used to express this null hypothesis in the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$.

   (g) Suppose that mean profiles for the $q = 4$ groups are approximately parallel, and we are interested in testing whether they in fact coincide for the 4 groups. Taking into account the usual constraints, write down the null hypothesis relevant to this situation in terms of μ, the τ_ℓ, the γ_j and the $(\tau\gamma)_{\ell j}$. Then write down appropriate matrices $\boldsymbol{C}$ and $\boldsymbol{U}$ that would allow the null hypothesis to be written in the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$. Perform the matrix multiplication $\boldsymbol{CMU}$ and verify that your choice of $\boldsymbol{C}$ and $\boldsymbol{U}$ indeed leads to the correct form of the null hypothesis.

   (h) Suppose that mean profiles for the $q = 4$ groups are approximately parallel, and we are interested in testing whether they in fact constant over the 3 time points. Taking into account the usual constraints, write down the null hypothesis relevant to this situation in terms of μ, the τ_ℓ, the γ_j and the $(\tau\gamma)_{\ell j}$. Then write down appropriate matrices $\boldsymbol{C}$ and $\boldsymbol{U}$ that would allow the null hypothesis to be written in the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$. Perform the matrix multiplication $\boldsymbol{CMU}$ and verify that your choice of $\boldsymbol{C}$ and $\boldsymbol{U}$ indeed leads to the correct form of the null hypothesis.

2. A study was conducted to investigate the effect of different diet and exercise regimens on the growth of laboratory rats. 100 young rats were randomly assigned to one of 4 regimens, 25 rats per regimen group. Rats in group 1 were treated as usual; rats in group 2 were fed a special diet but otherwise treated as usual; rats in group 3 were subjected to a daily treadmill exercise program, but were fed as usual; and rats in group 4 were both fed the special diet and subjected to the exercise program. After 1 month on the assigned regimens (allowing their possible effect to emerge), the body weight of each rat was recorded for 3 consecutive weeks (labeled 0,1,2 weeks) and then 2 weeks later (labeled 4 weeks), and then every 4 weeks thereafter for 8 more weeks (labeled 8 and 12 weeks). A main question of interest to the investigators was whether raising the rats on these different regimens produced subsequent

different patterns of growth; specifically, is it possible to alter the rate of growth of the rats by subjecting them to different regimens of diet and/or exercise? The investigators were also interested in comparing body weights among the four regimens at the end of the study.

The data are in the file `rats.dat` on the class web page. The data set has 4 columns:

Column	Description
1	Rat ID number
2	Week (0,1,2,4,8,12)
3	Weight (g)
4	Regimen group indicator (1–4)

(a) Using your favorite graphical software, make separate spaghetti plots of the raw longitudinal profiles for each treatment group separately. Give your general impressions on the "overall" observed patterns and how they might compare across treatment groups.

(b) Write a SAS program to obtain the following analyses:

(i) Read in the data.

(ii) Find the means for each group at each time point and plot them for each group on the same graph.

(iii) Calculate the sample covariance matrices and associated estimated correlation matrices for each group separately.

(iv) Calculate the centered and scaled versions of the data separately by group and make scatterplot matrices for each group separately.

(v) Conduct a univariate repeated measures ANOVA using `proc glm` via the "split-plot" specification of the model. Use a `random` statement to get SAS to print out the table of expected mean squares.

(vi) Conduct a univariate repeated measures ANOVA using `proc glm` using the `repeated` statement. Have your program print out the test for sphericity.

(vii) Use `proc glm` and appropriate `repeated` statements to obtain specialized within-unit analyses using the polynomial, profile, and Helmert transformations.

On the basis of the output, answer (c)–(k) below.

(c) Informally, from the plot of the weight means over time for each group from (b)(ii), do you think that the visual evidence supports the contention that the pattern of mean weight over the time period of the study is similar across the regimens? Explain your answer.

(d) From the output of (b)(iii) and the scatterplot matrices from (b)(iv), do you think the assumptions regarding variance, covariance, and correlation underlying the univariate repeated measures analysis of variance seem plausible? Explain your answer.

(e) From the output of (b)(v), write down the numerical value of the F ratio suitable for testing the null hypothesis that the change in mean weight over time is the same for the 4 regimens versus the alternative that is is not. At level of significance $\alpha = 0.05$, is there sufficient evidence to reject the null hypothesis? Provide justification for your answer.

(f) From the output of (b)(v), construct ANOVA estimates of σ_b^2 and σ_e^2 under the assumption of the compound symmetry model, and write down an estimate for the covariance matrix of a data vector from any regimen group. Provide an estimate of the correlation between two weight observations on the same rat from any regimen group.

(g) Returning to the issue in (c), cite appropriate results from the output of (b)(vi) that provide further insight into whether the assumption on covariance and correlation underlying the univariate repeated measures analysis of variance is appropriate. Is there sufficient evidence to conclude that the assumption is inappropriate? Is it necessary to use the "adjusted" F test for parallelism here?

(h) Does it make sense to consider the test of whether mean weight differs among the regimens averaged across time? Why or why not?

(i) Regardless of your answer to (h), write down the numerical value of the F ratio that would be appropriate for addressing this issue and state the conclusion that would be drawn.

(j) Does it make sense to consider the test of whether mean weight differs over time averaged over regimen groups? Why or why not? Regardless of your answer, give the numerical value of the F ratio that would be appropriate for addressing this issue and state the conclusion that would be drawn.

(k) From the output for (b)(vii), give the numerical value of the F test statistic that would be appropriate for testing whether the groups differ in terms of linear trend over time. Is there sufficient evidence to conclude at level $\alpha = 0.05$ that the mean weight patterns for the four groups differ in this way? From the plot in (ii), does the result of this test seem to agree with the visual evidence? Explain your answer.

(l) Before the experiment, the investigators were expecting that rats that neither ate the special diet nor exercised would not only increase their weights on average at a faster rate than rats that did at least one of these things, but they would in fact show an "accelerated trend" of weight gain toward the end of the study as their bad habits caught up with them! That is, the investigators expected that the mean profile for group 1 would "pull away" from the others toward the end of the study and show an increasing rate of change that might make it curve upward. They wanted to obtain some formal evidence that the patterns of mean weight differ across regimens in this kind of way.

From the specialized analyses performed in (b)(vii), identify one that you think most nearly tries to address this issue, and explain why you chose it. Give the numerical value of the F test statistic that would be appropriate for testing this issue. Is there sufficient evidence to conclude at level $\alpha = 0.05$ that this phenomenon occurred on average? From the plot in (ii), does the result of the test seem to agree with the visual evidence? Explain your answer.