ST 790, Midterm
Spring 2017

Please sign the following pledge certifying that the work on this test is your own:

"I have neither given nor received aid on this test."

Signature: ______________________________

Printed Name: ______________________________

There are FOUR questions, each with multiple parts. For each part of each question, please write your answers in the space provided. If you need more space, continue on the back of the page and indicate clearly where on the back you have continued your answer. Scratch paper is available from the instructor; just ask.

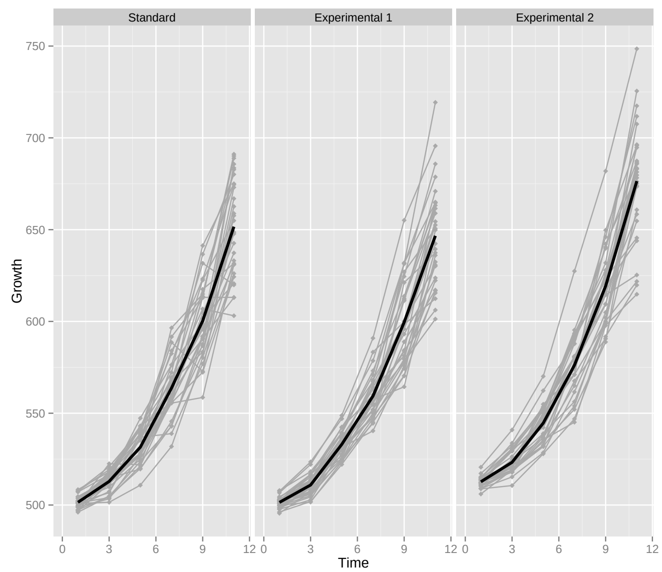
You are allowed ONE (1) SHEET of NOTES (front and back). Calculators are NOT allowed (you will not need one). NOTHING should be on your desk but this test paper, your one page of notes, and any scratch paper given to you by the instructor.

Points for each part of each problem are given in the left margin. TOTAL POINTS = 100.

If you are asked to provide an expression, you need not carry out the algebra to simplify the expression (unless you want to do so).

In all problems, all symbols and notation are defined exactly as they are in the class notes.

1. A crop scientist has conducted an experiment to study the growth of three varieties of hops (*Humulus lupulus*, a plant used in the brewing of beer): a standard variety ("Standard") and two experimental varieties ("Experimental 1" and "Experimental 2"). 30 plants of each variety were planted at week 0, and at weeks 1, 3, 5, 7, 9, and 11, a measure of growth was obtained on each plant. The data are shown below, with the sample means at each measurement time superimposed.



The main goals of the experiment were

(i) To determine if the patterns of change of mean growth are not the same for all varieties.

(ii) To determine if the rate of change of mean growth is not constant for at least one of the varieties.

The crop scientist hopes to address this and other questions based on the following model and the standard assumptions made for it:

$$Y_{h\ell j} = \mu_{\ell j} + b_{h\ell} + e_{h\ell j} = \mu + \tau_\ell + \gamma_j + (\tau\gamma)_{\ell j} + b_{h\ell} + e_{h\ell j}, \tag{1}$$

where $Y_{h\ell j}$ is growth for the hth plant from the ℓth variety at jth measurement occasion, $j = 1, \ldots, 6$; $\ell = 1, 2, 3$ indexes the Standard, Experimental 1, and Experimental 2 varieties, respectively; and the terms on the right hand side of (1) are as defined in the course notes.

The crop scientist presents you with the following output of an analysis based on (1):

```
Source             DF     Type III SS    Mean Square   F Value   Pr > F

 variety            2     24512.63837    12256.31919     15.44   <.0001
 Error             87     69070.26267      793.91107
 week               5     1423820.669     284764.134   1544.26   <.0001
 week*variety      10        2154.105        215.410      1.17   0.3104
 Error(week)      435       80214.903        184.402

Mauchly's Criterion   DF   Chi-Square    Pr > ChiSq
                      14    343.30055        <.0001
```

For each variety, the scientist has also calculated the sample covariance matrix $\widehat{\boldsymbol{V}}$ and associated correlation matrix $\widehat{\boldsymbol{\Gamma}}$ of the observed responses; these are as follows for the Standard (S), Experimental 1 ($E1$), and Experimental 2 ($E2$) varieties:

$$\widehat{\boldsymbol{V}}_S = \begin{pmatrix} 11.1 & 8.5 & 13.4 & 32.7 & 6.6 & 45.3 \\ 8.5 & 30.1 & 23.1 & 50.1 & 26.0 & 26.4 \\ 13.4 & 23.1 & 77.0 & 70.7 & 69.5 & 58.0 \\ 32.7 & 50.1 & 70.7 & 241.8 & 137.0 & 182.7 \\ 6.6 & 26.0 & 69.5 & 137.0 & 406.6 & 146.6 \\ 45.3 & 26.4 & 58.0 & 182.7 & 146.6 & 723.3 \end{pmatrix}, \quad \widehat{\boldsymbol{\Gamma}}_S = \begin{pmatrix} 1.00 & 0.47 & 0.46 & 0.63 & 0.10 & 0.50 \\ 0.47 & 1.00 & 0.48 & 0.59 & 0.23 & 0.18 \\ 0.46 & 0.48 & 1.00 & 0.52 & 0.39 & 0.25 \\ 0.63 & 0.59 & 0.52 & 1.00 & 0.44 & 0.44 \\ 0.10 & 0.23 & 0.39 & 0.44 & 1.00 & 0.27 \\ 0.50 & 0.18 & 0.25 & 0.44 & 0.27 & 1.00 \end{pmatrix}$$

$$\widehat{\boldsymbol{V}}_{E1} = \begin{pmatrix} 8.9 & 13.1 & 13.6 & 26.7 & 44.9 & 67.6 \\ 13.1 & 31.1 & 29.2 & 51.2 & 96.4 & 112.6 \\ 13.6 & 29.2 & 47.2 & 58.4 & 88.6 & 109.3 \\ 26.7 & 51.2 & 58.4 & 142.6 & 170.6 & 223.8 \\ 44.9 & 96.4 & 88.6 & 170.6 & 522.3 & 464.7 \\ 67.6 & 112.6 & 109.3 & 223.8 & 464.7 & 768.8 \end{pmatrix}, \quad \widehat{\boldsymbol{\Gamma}}_{E1} = \begin{pmatrix} 1.00 & 0.79 & 0.67 & 0.75 & 0.66 & 0.82 \\ 0.79 & 1.00 & 0.76 & 0.77 & 0.76 & 0.73 \\ 0.67 & 0.76 & 1.00 & 0.71 & 0.56 & 0.57 \\ 0.75 & 0.77 & 0.71 & 1.00 & 0.63 & 0.68 \\ 0.66 & 0.76 & 0.56 & 0.63 & 1.00 & 0.73 \\ 0.82 & 0.73 & 0.57 & 0.68 & 0.73 & 1.00 \end{pmatrix}$$

$$\widehat{\boldsymbol{V}}_{E2} = \begin{pmatrix} 7.7 & 12.4 & 21.4 & 35.2 & 42.2 & 44.6 \\ 12.4 & 41.7 & 47.3 & 79.7 & 81.5 & 150.4 \\ 21.4 & 47.3 & 99.1 & 133.6 & 167.7 & 223.7 \\ 35.2 & 79.7 & 133.6 & 279.5 & 281.2 & 353.6 \\ 42.2 & 81.5 & 167.7 & 281.2 & 422.7 & 421.2 \\ 44.6 & 150.4 & 223.7 & 353.6 & 421.2 & 993.4 \end{pmatrix}, \quad \widehat{\boldsymbol{\Gamma}}_{E2} = \begin{pmatrix} 1.00 & 0.69 & 0.77 & 0.76 & 0.74 & 0.51 \\ 0.69 & 1.00 & 0.74 & 0.74 & 0.61 & 0.74 \\ 0.77 & 0.74 & 1.00 & 0.80 & 0.82 & 0.71 \\ 0.76 & 0.74 & 0.80 & 1.00 & 0.82 & 0.67 \\ 0.74 & 0.61 & 0.82 & 0.82 & 1.00 & 0.65 \\ 0.51 & 0.74 & 0.71 & 0.67 & 0.65 & 1.00 \end{pmatrix}$$

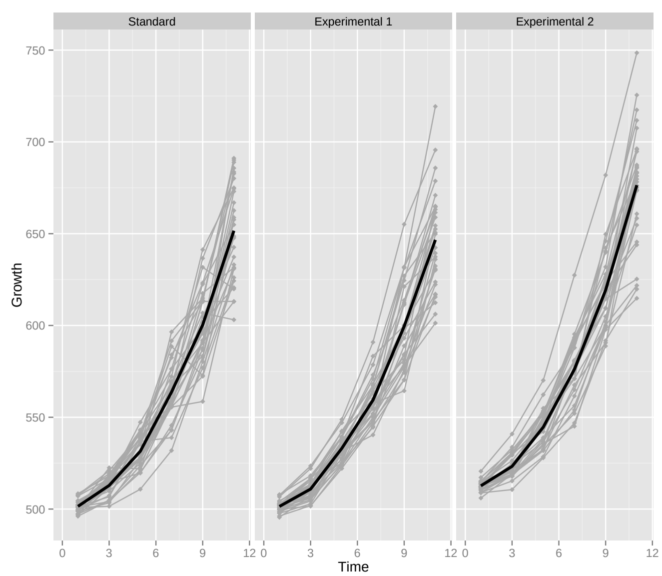
[7 points] (a) Define

$$\boldsymbol{\mathcal{M}} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{16} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{26} \\ \mu_{31} & \mu_{32} & \cdots & \mu_{36} \end{pmatrix}.$$

Give an expression in terms of $\boldsymbol{\mathcal{M}}$ that formalizes the crop scientist's question (ii) (to determine if the rate of change of mean growth is not constant for at least one of the varieties), defining any additional symbols you use, or explain why this is not possible.

[10 points] (b) Based on the information you have, can you address question *reliably* (i) using model (1)? If so, describe how and present a formal statement of the result. If not, explain why not.

2. Consider the hops experiment in the previous problem. Here are the data again:



The goals are

(i) To determine if the patterns of change of mean growth are not the same for all varieties.

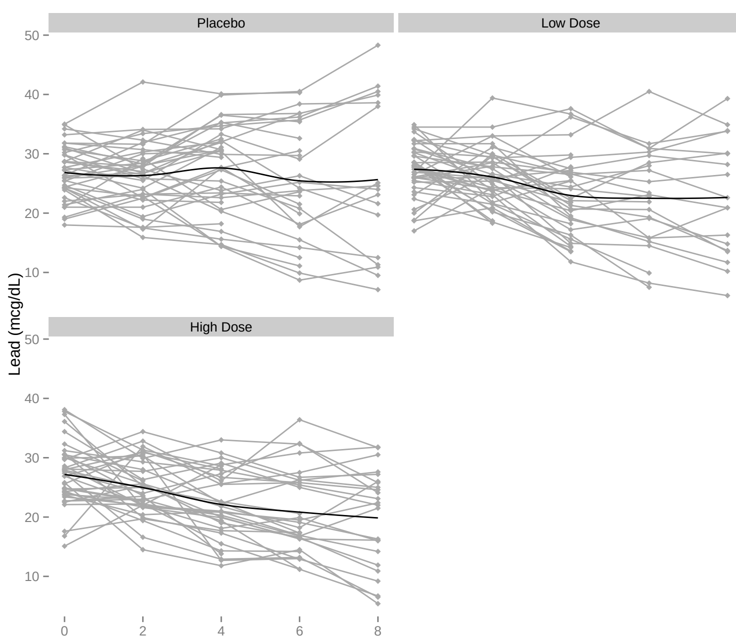
(ii) To determine if the rate of change of mean growth is not constant for at least one of the varieties.

[10 points] (a) Propose a statistical model different from that in (1) in which both (i) and (ii) can be addressed. *Briefly* state any assumptions you incorporate in the model.

[8 points] (b) For your model in (a), write down a vector β that collects all parameters that characterize mean growth for the three varieties. Then provide a matrix $\boldsymbol{L}$ such that you can address the question of whether or not the rate of change in mean growth is *both constant and the same* for all varieties through an expression of the form $\boldsymbol{L}\beta$.

[8 points] (c) In terms of $\boldsymbol{\beta}$ you defined in (b), provide a matrix $\boldsymbol{L}$ that allows you to characterize the rate of change of mean growth for each variety at the midpoint of the study period (5.5 weeks) through an expression of the form $\boldsymbol{L\beta}$.

3. The data shown shown below are from a study of an oral treatment for lead exposure in 120 children with blood lead levels greater than 10 μg/dL. The children were randomized to receive placebo, low dose of the active treatment, or high dose of the active treatment, 40 children per group. Blood lead levels were to be obtained from each child at baseline (week 0), prior to initiation of assigned treatment, and then at weeks 2, 4, 6, 8 thereafter. The gender of each child was also recorded (0 = female, 1 = male). Here are the data, with a loess smooth superimposed on each plot.



There was substantial dropout from the study; only 84 of the 120 children returned for the week 6 measurement, and only 59 returned at week 8.

One of the study investigators ("Investigator A") posed the following questions:

(i) Is the average (mean) rate of change in lead levels over the 8 weeks associated with treatment received?

(ii) Is the average rate of change in lead levels over the 8 weeks associated with gender?

(iii) Is average baseline lead level different for males and females?

Another investigator ("Investigator B") disagreed, and restated the questions as follows:

(i) Is the rate of change of average (mean) lead level over the 8 weeks associated with treatment received?

(ii) Is the rate of change in average lead level over the 8 weeks associated with gender?

(iii) Is the average baseline lead level different for males and females?

[12 points] (a) Can you propose a statistical model in which questions (i)-(iii) of both Investigator A and Investigator B can be addressed? If so, write down the model and *briefly* state any assumptions you incorporate in the model. If not, state why not, and write down a model in which the questions of *one of* the investigators can be addressed (state which investigator), including (*briefly*) any assumptions you incorporate in the model.

[7 points] (b) In terms of your model in (a), show how you would address Investigator A's question (i) (Is the average rate of change in lead levels associated with treatment received?) If you cannot, state why not.

[7 points] (c) In terms of your model in (a), show how you would address Investigator B's question (ii) (Is the rate of change in average lead level associated with gender?) If you cannot, state why not.

[8 points] (d) Show how, using your model in (a), you would estimate lead level at week 6 for the *i*th child, a female who received high dose active treatment. If you can not, state why not.

[8 points] (e) Suppose that children who dropped out of the study tended to do so because their previously recorded lead levels were not improving. Would you feel comfortable proceeding with a standard analysis using your model? If so, explain *briefly* how you would conduct the analysis based on your model under this condition. If not, explain *briefly* why not.

[8 points] 4. (a) Consider the model

$$Y_{ij} = \beta_{0i} + e_{ij}, \quad i = 1, \dots, m, \; j = 1, \dots, n, \quad \beta_{0i} = \beta_0 + b_i, \tag{2}$$

where b_i are independent for all i, e_{ij} are independent for all i, j, and b_i and e_{ij} are independent of one another for all i, j, with

$$E(b_i) = 0, \quad \text{var}(b_i) = D \qquad E(e_{ij}) = 0, \quad \text{var}(e_{ij}) = \sigma^2.$$

Suppose we fit (2) by maximum likelihood to data under the assumption that b_i and e_{ij} are all normally distributed using standard software to obtain values for the estimators $\widehat{\beta}_0$, $\widehat{D}$, and $\widehat{\sigma}^2$ and associated standard errors for all three parameters based on model (2).

What conditions would you want to be met to feel comfortable that these standard errors for $\widehat{D}$ and $\widehat{\sigma}^2$ are reliable assessments of the uncertainty of estimation of D and σ^2? Explain (*briefly*) your answer.

[7 points] (b) Now consider the model

$$Y_{ij} = \beta_{0i} \exp(-\beta_1 t_j) + e_{ij}, \quad i = 1, \dots, m, \; j = 1, \dots, n, \quad \beta_{0i} = \beta_0 + b_i, \tag{3}$$

for times $t_1, \dots, t_n$, where b_i and e_{ij} are as in (a).

Statistician A refers to β_1 in (3) as the average rate of decay of the response among individuals in the population, while Statistician B refers to β_1 as the rate of decay of the average response in the population. Which statistician is correct? Explain (*briefly*) your answer.