Instructions: This is a closed book exam. No notes or books are allowed. You may use
a calculator, but you are not allowed to store notes or formulas in the
calculator. Please write your answers on additional sheets of paper. There
is <u>not</u> enough room to write your answers on this exam. Please write your
name on every sheet of paper that you submit. Submit this copy of the
exam with your solutions.

1.   In a study of the amino acid phenylalanine, 27 young male rats weighing 50 to 55 grams
were fed a diet lacking phenylalanine for two days. Then, the rats were randomly
assigned to new diets with three rats assigned to each level of phenylalanine. The
weight gain (or weight loss) after 15 days on the new diet was recorded for each rat.
The results are shown in the following table. Negative values for weight gain indicate a
weight loss.

Percent Phenylalanine in the new diet	Weight Gain (grams)
0	-7.0, -5.5, -1.2
0.2	-6.1, -2.4, 0.7
0.4	-2.0, 1.0, 8.9
0.6	8.8, 14.3, 15.1
0.8	25.6, 26.4, 29.4
1.0	33.3, 35.2, 43.1
1.2	38.5, 46.3, 52.0
1.4	47.9, 49.2, 52.6
1.6	42.1, 52.8, 54.8

Least squares estimation was used to fit the following model to the data

$$Y_{ij} = \beta_0 + \frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 X_i)} + \varepsilon_{ij} \quad ,$$

where

Y_{ij} is the observed 15 day weight gain for the j-th rat fed a new diet
containing the percent phenylalanine given by X_i

and the

ε_{ij}'s   are independent and identically distributed random errors with mean zero
and variance σ^2.

The estimated coefficients are given below along with an estimate of the variance covariance matrix for the estimated coefficients.

$$
\underset{\sim}{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} -6.42 \\ 57.82 \\ 3.57 \\ 4.79 \end{bmatrix} \qquad V = \begin{bmatrix} 7.87 & -10.64 & 1.57 & 1.77 \\ -10.64 & 18.92 & -2.49 & -3.18 \\ 1.57 & -2.49 & 0.45 & 0.55 \\ 1.77 & -3.18 & 0.55 & 0.71 \end{bmatrix}
$$

A graph showing the data and the estimated curve is shown on the next page along with some residual plots.

A. With respect to the relationship between mean weight gain in rats and the level of phenylalanine in their diet, give an interpretation of the parameters β_0, β_1, and β_2 in this model. You are not asked to give a precise interpretation of β_3 which is a coefficient related to the rate at which mean weight gain increases as the level of phenylalinine increases.

B. Assuming that the proposed model is correct, show how to construct an approximate 95% confidence interval for b_1 .

C. Assuming that the proposed model is correct, show how to use b, V and the delta method construct an approximate 95% confidence interval for the level of phenylalanine in the diet corresponding to a 15 day mean weight gain of zero. This is the level of phenylalinine at which an "average" rat will neither gain nor lose weight. (You are not required to complete the numerical evaluation of the confidence limits.)

D. Assuming that the proposed model is correct, explain how you would use a bootstrap procedure to construct a confidence interval for the mean 15 day weight gain for rats with 1.0 percent phenylalanine in their diet. Give only <u>one</u> answer corresponding to what you think is the best way to use the bootstrap. (If you give more than one answer you will receive credit only for your weakest response.)

E. Examine the residual plots. What do they suggest?

F. Show how to construct a test statistic to test the null hypothesis that the conditional mean for the 15 day weight change corresponds to the proposed formula $E(Y \mid X) = \beta_0 + \dfrac{\beta_1}{1 + \exp(\beta_2 - \beta_3 X)}$. Either present a formula for a test statistic and show how to use it, or describe what you would do. You do not have to provide any theoretical justification for your answer.

2.   A certain bioinsecticide is used to control insects that live in soil and damage the roots of plants. This bioinsecticide consists of spores of a certain fungus that are suspended in a solution. After they are introduced into the soil, the spores develop into fungal colonies that infect and destroy the insects. The object of one study was to examine how different levels of irrigation affect the viability of spores in the soil.

Ten different fields were used in this study. These fields were randomly selected from a large set of fields that could have been used in the study. Each field was subdivided into three plots of equal size. Each plot was sprayed with a concentration of the bioinsecticide that deposited about 5 x 10^8 spores per 10 cm^2 of surface area. One plot in each field was irrigated with 2 cm of water per day. Another plot in each field was irrigated with 1.0 cm of water per day. The remaining plot in each field received no irrigation. The three levels of irrigation were randomly assigned to the plots with a different randomization within each field. Soil samples were taken from each plot at 1, 5, 10, and 20 days after the plot was sprayed with the bioinsecticide, and the number fungal colonies (Y) was measured for each soil sample.

Consider the model

$$Y_{ijk} = \mu + \beta_i + \alpha_j + \delta_{ij} + \tau_k + \gamma_{jk} + \varepsilon_{ijk} \,,$$

where

Y_{ijk} is an observation on the number of fungal colonies in a soil sample,

α_j corresponds to the j-th level of irrigation

τ_k corresponds to the k-th sampling date

γ_{jk} corresponds to the j-th level of irrigation and the k-th sampling date

β_i is a random field (block) effect with $\beta_i \sim \text{NID}(0, \sigma_f^2)$

δ_{ij} is a random effect with $\delta_{ij} \sim \text{NID}(0, \sigma_p^2)$

ε_{ijk} is a random effect with $\varepsilon_{ijk} \sim \text{NID}(0, \sigma_e^2)$

and β_i, δ_{ij}, and ε_{ijk} are mutually independent.

A. An analysis of variance table for this model includes the following sources of variation:

Source of Variation	Degrees of Freedom	Expected Mean Square
Fields		$\sigma_e^2 + 4\sigma_p^2 + 12\sigma_f^2$
Plots within fields		$\sigma_e^2 + 4\sigma_p^2$
Error		σ_e^2

Report the appropriate degrees of freedom and use these results to obtain formulas for estimates of the variance components.

B. For the proposed model, show how you would test the null hypothesis that there is no interaction between the levels of the irrigation factor and time (sample dates). Give a formula for your test statistic and explain how you would use it to make an inference about the potential interaction between levels of irrigation and time.

C. Suppose the researchers want to compare the results at 20 days for the irrigation levels correspond to 1 cm of water per day and 2 cm of water per day. Averaging across fields, let $\overline{Y}_{\cdot 24} = \dfrac{1}{10}\sum_{i=1}^{10} Y_{i24}$ denote the average of the observations at 20 days using 1 cm of water per day, and let $\overline{Y}_{\cdot 34} = \dfrac{1}{10}\sum_{i=1}^{10} Y_{i34}$ denote the average of the observations at 20 days using 2 cm of water per day. Find the variance of $\overline{Y}_{\cdot 24} - \overline{Y}_{\cdot 34}$ and show how to construct a 95% confidence interval for the difference in the mean results at 20 days for irrigation levels of 1cm and 2cm of water per day.

D. Which of the following are true statements about the proposed model? Circle any statement that is true.

    i.   Each observation has the same variance.
    ii.   Any two observations taken in the same field have the same level of correlation.
    iii.   Two observations taken from two different fields are uncorrelated.
    iv.   The REML estimates of the variance components are equal to the estimates of the variance components that you were asked to derive in part (A).
    v. $(\alpha_2 + \gamma_{24}) - (\alpha_3 + \gamma_{34})$ is estimable.

E. Suppose that one of the objectives of the study was to show that there is a time effect when plots are irrigated with 2cm of water per day. Consequently, the researchers want to test the alternative that the mean results differ for at least two time points. Show how to construct an F-test for this null hypothesis.

3. A second model was proposed for the data in problem 2. Using the notation from problem 2, this model is

$$Y_{ijk} = \mu + \beta_i + \alpha_j + \delta_{ij} + \tau_k + \eta_{ijk},$$

where β_i is a random field (block) effect with $\beta_i \sim NID(0, \sigma_f^2)$ and the η_{ijk}'s are random effects that are independent of the β_i's and

$$
\begin{bmatrix} \eta_{ij1} \\ \eta_{ij2} \\ \eta_{ij3} \\ \eta_{ij4} \end{bmatrix} \sim NID \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho^4\sigma_1\sigma_2 & \rho^9\sigma_1\sigma_3 & \rho^{19}\sigma_1\sigma_4 \\ \rho^4\sigma_1\sigma_2 & \sigma_2^2 & \rho^5\sigma_2\sigma_3 & \rho^{15}\sigma_2\sigma_4 \\ \rho^9\sigma_1\sigma_3 & \rho^5\sigma_2\sigma_3 & \sigma_3^2 & \rho^{10}\sigma_3\sigma_4 \\ \rho^{19}\sigma_1\sigma_4 & \rho^{15}\sigma_2\sigma_4 & \rho^{10}\sigma_3\sigma_4 & \sigma_4^2 \end{bmatrix} \right)
$$

A. Explain how this model differs from the model proposed in problem 2.

B. Explain how you would perform tests to determine if either this model or the model from problem 2 was appropriate for the data. Report the degrees of freedom for any test you propose and explain how you would use it to make a decision.

Final Exam Score _____

Course Grade _____