Due: 5 pm, Friday March 29.

1. The data in ryegrass.txt come from a study of pasture management in New Zealand. The study evaluates dry matter production (a good thing) in four varieties of perennial ryegrass (*Lolium perenne*) grown at two levels of fertilization (heavy or average manure). The research area was divided into four blocks, each containing four plots. Plots were randomly assigned to the four varieties in an RCBD. Each plot was divided into two; one half was randomly assigned to heavy manure while the other half received average manure. This factor is called fertilization throughout most of the problem.

   A reasonable model for this study is

   $$y_{ijk} = \mu + \beta_i + \alpha_j + \alpha\beta_{ij} + \tau_k + \alpha\tau_{jk} + \epsilon_{ijk}, \tag{1}$$

   where: y_{ijk} is the dry matter production for variety j in block i receiving fertilization level k,
   β_i are the block effects, $i \in \{1, 2, 3, 4\}$,
   α_j are the variety effects, $j \in \{1, 2, 3, 4\}$,
   $\alpha\beta_{ij} \overset{iid}{\sim} N(0, \sigma_m^2)$ are the main plot errors
   τ_k are the fertilization effects, $k \in \{1, 2\}$,
   $\alpha\tau_{jk}$ are the variety×fertilization interaction effects,
   and $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_s^2)$ are the split plot errors.

   (a) Consider the sequence of models obtained by adding each model term sequentially. Calculate the degrees of freedom associated with each sequential SS when terms are added following their left to right appearance in (1).

   (b) The SS associated with adding variety (α_j) to the model is

   $$SS_{variety} = \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{k=1}^{2} (\bar{y}_{.j.} - \bar{y}_{...})^2$$

   Derive the expected mean square for this SS.

   (c) If blocks were considered random, i.e. $\beta_i \overset{iid}{\sim} N(0, \sigma_b^2)$, would the expected MS for variety change from what you derived in the previous part? Briefly explain why or why not. Hint: You should be able to answer this without rederiving the E MS.

   (d) Some of the other E MS are:

Source	E MS
Block*Variety	$\sigma_s^2 + 2\sigma_m^2$
Fert.	$\sigma_s^2 + Q()$
Fert.×Variety	$\sigma_s^2 + Q()$
Residual	σ_s^2

where $Q()$ indicates a quadratic form involving only fixed effects. Construct the ANOVA table and test the main effects of variety, fertilization, and the interaction.

(e) Pastures will be seeded with a single variety of ryegrass. Hence, land managers are especially interested in the simple effects of manuring on each variety. Estimate the increase in dry matter production caused by heavy manuring (i.e. the difference between heavy and average manure) for each variety. Report each simple effect, its s.e. and a p-value for the test of effect of fertilization on that variety.

(f) The investigators are interested in the difference between varieties S23 and NZ when heavily manured. What are the d.f. for the standard error of this difference? Calculate a 95% confidence interval for this difference.

2. Soils are not solid; in fact they may contain a substantial amount of open space. Porosity describes the amount of open space in a soil. This varies across the landscape at different spatial scales. This problem is based on a study to estimate the variability at three spatial scales: among fields, among sections within fields and among locations within sections. There are 15 fields and 30 sections (2 sections per field). Sections are nested within fields; locations are nested within sections. A reasonable model for these data is:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}$$

where y_{ijk} is the porosity of location k in section j of field i,
$\alpha_i \overset{iid}{\sim} N(0, \sigma^2_{field})$, $i \in \{1, 2, \cdots, 15\}$ describe the variaion among fields,
$\beta_{ij} \overset{iid}{\sim} N(0, \sigma^2_{section})$, $j \in \{1, 2\}$ describe the variation among sections within a field,
and $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2_{location})$, $k \in \{1, 2\}$ describe variation among locations within a section.

(a) Imagine there are 2 locations per section. Consider fitting a sequence of models corresponding to sequentially adding each term in the model (fields, sections given fields, locations given sections and fields). Determine the d.f. associated with sequential SS.

(b) The SS for fields is calculated as $\sum_{i=1}^{15} \sum_{j=1}^{2} \sum_{k=1}^{2} (\bar{y}_{i..} - \bar{y}_{...})^2$. Calculate the E MS for fields.

(c) When the study was carried out, the investigators decided not to sample 2 locations per section. Instead, they randomly chose six sections that were sampled at two locations; the other 24 sections were sampled at only one location. In all, they sampled 15 fields, 30 sections, and 36 locations. Determine the degrees of freedom associated with each sequential SS for the study as actually carried out.

(d) It is time consuming to measure porosity, so imagine the investigators could only sample 36 locations. A second study design measures 9 fields, 2 sections per field and 2 locations in each field. What (if any) are the advantages of the unbalanced design that was actually used, compared to the second design (balanced with 9 fields)?

(e) The expected mean squares for sequential addition of terms to the model are:

Source	EMS
field	$\sigma^2_{location} + 1.1905\,\sigma^2_{section} + 2.381\,\sigma^2_{field}$
section(field)	$\sigma^2_{location} + 1.2\,\sigma^2_{section}$
location(section, field)	$\sigma^2_{location}$

Calculate the sequential (Type I) SS associated with each source of variation, then compute the type I ANOVA estimates of the three variance components, $\sigma^2_{location}$, $\sigma^2_{section}$, and σ^2_{field}.

(f) After examining some simple diagnostics, the investigators realize the data set is incorrect. The value of 2.021 for field 1, section 2 should have been 5.021. The corrected data are in poro2.txt. Compute the type I ANOVA estimates of the three variance components using the corrected data.

(g) Using the corrected data, test Ho: $\sigma^2_{section} = 0$. Report your F statistic, its distribution under the null hypothesis, and the p-value.

(h) What linear combination of Mean Squares provides an appropriate error term to test Ho: $\sigma^2_{field} = 0$.

(i) Calculate the Cochran-Satterthwaite approximate d.f. for the appropriate error term to test Ho: $\sigma^2_{field} = 0$?

(j) Using the corrected data, test Ho: $\sigma^2_{field} = 0$. Report your F statistic, its distribution under the null hypothesis, and the p-value.

(k) How large does $\sigma^2_{section}$ need to be for the test of Ho: $\sigma^2_{section} = 0$ to have 80% for the arrangement of samples and sample sizes used in this study. Assume that $\sigma^2_{location} = 0.3816$, that the numerator and denominator df are the same as the current study and $\alpha = 0.05$.
Hints: The distribution of the F statistic for this type of hypothesis is in the notes. Remember that the power of the test depends on the population value of $\sigma^2_{section}$. If $\sigma^2_{section}$ is close to 0, the power will be close to 0.05; if $\sigma^2_{section}$ is larger, the power will be close to 1. I want you to find, or calculate, the value of $\sigma^2_{section}$ for which the power is 80%.