Exam style questions from the review lecture.

Note that the length of this set of questions does not necessarily reflect the length of the exam.

1. Based around the nested random effects model

$$y_{ijk} = \mu + \alpha_i + \gamma_{ij} + e_{ijk}, \; i = 1, \dots, a, \; j = 1, \dots, g, \; k = 1, \dots, r$$

with $\gamma_{ij} \sim N(0, \sigma_g^2)$, $e_{ijk} \sim N(0, \sigma_e^2)$. [Example: α_i is breed of sheep, γ_{ij} is the jth sheep in breed i, and we measure the wool obtained from each sheep k times.]

   (a) What is $\mathrm{cov}(y_{ijk}, y_{i'j'k'})$?

   (b) We can write $SSA = \sum_{ijk}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$. Find and expression for its expectation in terms of the α_i, σ_e^2 and σ_g^2.

   (c) Show that $MSG = \frac{1}{a(g-1)} \sum_{ijk}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2$ has the same expectation as $MSA = SSA/g$ under the null hypothesis that $\alpha_1 = \ldots = \alpha_a = 0$.

   (d) We would like to formally show that we get an F test out of MSA/MSG to do this, we'll write $z_{ij} = \bar{y}_{ij\cdot}$.

      i. Give an expression for z in vector form; show that its covariance can be written as $\tau^2 I$ (and find τ).

      ii. Write SSA and SSG as $z^T A_1 z$ and $z^T A_2 z$ with A_1 and A_2 idempotent and $A_1 A_2 = 0$.

      iii. Hence, show that MSA/MSG has an F distribution. Give its degrees of freedom.

      iv. Bonus: when H_0 is not true, what is the noncentrality parameter in the distribution above?

2. Contrasts. Here we assume that α has four levels and we are using reference coding.

   (a) Find a contrast matrix to test the hypotheses: $\alpha_1 = \alpha_2, \; \alpha_3 = (\alpha_1 + \alpha_2)/2, \; \alpha_4 = (\alpha_1 + \alpha_2 + \alpha_3)/4$

   (b) In fact, the levels $\alpha_1, \dots, \alpha_4$ corresponded to measuring wool in seasons 1, 2, 3 and 4. We expect a linear increase in wool production. Find a contrast matrix to test the hypothesis that $\alpha_1 = \delta_0 + \delta_1, \; \alpha_2 = \delta_0 + 2\delta_1, \; \alpha_3 = \delta_0 + 3\delta_1, \; \alpha_4 = \delta_0 + 4\delta_1$ for some (unknown) δ_0, δ_1.

3. In our model from Part 1,

   (a) Write down the joint distribution of $\bar{y}_{11\cdot}$ and γ_{11}, hence find the distribution of $\gamma_{11}|\bar{y}_{11\cdot}$.
   (b) What is the distribution of $(\gamma_{11} + \gamma_{12})/2|(\bar{y}_{11\cdot}, \bar{y}_{12\cdot})$?
   (c) Despite it being random, a colleague wants to provide a confidence interval for $(\gamma_{11} + \gamma_{12})/2$. What would you provide?

4. Longitudinal models

   In a new experiment each sheep in three different breeds was measured in seasons 1, 2, 3 and 4. The researchers believe that each sheep increases its yield linearly, but with a different linear regression line for each sheep. Different breeds may differ in their average regression lines.

   (a) Write down a model to express this understanding.
   (b) Write down the covariance matrix for the responses from the jth sheep in group i:
   (c) Re-write your model and the covariance with season treated as a category rather than continuous. Can you still estimate an interaction between sheep and season?
   (d) (More abstract and more difficult). In the general longitudinal model $y = X\beta + Zb + e$, $e \sim N(0, \sigma^2 I)$, $b \sim N(0, \sigma^2 G)$
      i. Write down the covariance of $X(I + ZGZ^T)^{-1}y$ and b.
      ii. Hence find the distribution of $b|X(I + ZGZ^T)^{-1}y$.