Stat 430 Homework 4

due: November 23, 2009 at 5pm

Problems marked [optional] are to help you learn the material and prepare for exams, but they are not to be turned in.

1. (a) Take your ANOVA analysis of the PND5 virus data from Homework 3 and compare it to a simple linear regression model for the same data, where the initial conditions are treated as a continuous parameter. Plot the data (x-axis initial condition, y-axis log ratio) and the fitted models (for ANOVA, plot the sample means; for linear regression, plot the fitted line). Is there much difference in the fit?

   (b) The ANOVA model has one more parameter than the linear regression model. What F-like statistic could you use to test for a significantly better fit with the ANOVA model? Does the proposed statistic have an F distribution (I'm not asking for a proof, just an argument by analogy to previous F test derivations)? Assuming it does have an F distribution, perform the test.

2. The **random effects model** for the one-way layout considers the I treatments as sampled from a larger population of treatments. For example, in the PND5 virus, the three initial conditions, 0.1, 0.2, and 1, are the ratio of PND5 and standard virus at the start of the experiment. Thus, the three numbers in the experiment are representative of infinitely many possible initial conditions. The random effects model is

$$Y_{ij} = \mu + A_i + \epsilon_{ij}$$

   where A_i is now a random variable (rather than a fixed effect) such that $E[A_i] = 0$ and $\text{Var}(A_i) = E[A_i^2] = \sigma_A^2$, and all $A_1, \ldots, A_I$ are independent of each other. The measurement errors ϵ_{ij} continue to have $E[\epsilon_{ij}] = 0$ and $\text{Var}(\epsilon_{ij}) = \sigma_\epsilon^2$, and they are independent of the random A effects. Given the assumptions, we know

$$\text{Var}(Y_{ij}) = \sigma_A^2 + \sigma_\epsilon^2$$

   (a) Show that

$$E[MS_W] = \sigma_\epsilon^2 \qquad\qquad E[MS_B] = \sigma_\epsilon^2 + J\sigma_A^2$$

   (b) Use the above to produce estimators of both variance components (σ_ϵ^2 and σ_A^2).

   (c) The virus data does not satisfy the random effects model, because in question 1, you find preliminary evidence that $E[A_i]$ depends on the level i. To demonstrate the random effects model, access the dye data. This dataset tests the strength of dye across batches. To measure strength, the product was used to dye a square of cloth. The resulting fabric was visually assessed and given a numeric score by experts. Large samples of dye were taken from six batches. The samples were well-mixed, and six random subsamples were taken from each large sample. The 36 subsamples were submitted to a laboratory for testing in random order. Estimate μ, σ_ϵ^2, and σ_A^2 for this model.

(d) Suggest and perform a test for $H_0 : \sigma_A^2 = 0$.

(e) [optional] Propose and implement a computer-based resampling test to test the same null $H_0 : \sigma_A^2 = 0$.

3. Female rats exposed to different lighting cycles were dosed with 0 (control saline solution), $10, 50, 250$, or 1250 ng luteinizing releasing factor (LRF). The amount of luteinizing hormone (LH) was measured in the blood at a later time. Use ANOVA to analyze the data to determine the effect of the light regimen and LRF dose on LH release. Perform a non-parametric test as well, and assess your findings.