Assignment #11 — Due Friday, December 3, 2010, by 4:00 P.M.
Turn in homework to your TA's mailbox using this sheet as the cover page.

Fill in your name and also circle the *lecture section in which you are registered* and circle the *discussion section you expect to attend* to pick up this assignment.

Name:

Lecture 1 (Larget).	**311:** Tu 1:00 - 2:15pm	**312:** Th 8:00 - 9:15am	**313:** We 1:00 - 2:15pm
Lecture 2 (Hanlon).	**321:** Tu 1:00 - 2:15pm	**322:** We 2:30 - 3:45pm	**323:** We 1:00 - 2:15pm

Please answer the following questions.

1. The purpose of this exercise is to gain experience using R to analyze data, specifically in the context of ANOVA. Therefore, your analysis should use the raw data provided (and not the summary measures). You might find it helpful to review the R code used in the lecture notes to analyze the cuckoo bird data. Use the data set from p.427, Problem 22, which you can enter by hand or find in the file `lodgepole.csv`. But answer these questions.
   (a) Make a dotplot of cone size versus the group. Do the samples look like they were drawn from populations with equal means? Do the samples have low skewness and similar standard deviations?
   (b) Find the ANOVA table. From this table, what is the numerical estimate of the common standard deviation in all populations?
   (c) Use ANOVA to test the equality of mean cone size of lodgepole pines among the three groups. Interpret your results in the context of the problem.
   (d) Use Tukey's honestly significant difference (HSD) to compute simultaneous 95% confidence intervals for all pairwise differences.
2. Use the context and raw data from p. 425, Problem 13, to answer these questions.
   (a) Make a dotplot of resistance versus cyanobacterium density. Comment on features of the data.
   (b) Use the R command `lm()` to fit a linear model that uses the cyanobacterium group to predict resistance and display the fitted model with `anova()`.
   (c) Assuming that deviations from model assumptions are minor, interpret the results in the context of the problem.
   (d) What characteristics of the data might cause you to question the validity of the previous answer?
3. Use the data set from p.426, Problem 18. But answer these questions.
   (a) Using the summary of the data provided in the text, complete the ANOVA table for the problem.
   (b) Compute R^2 for this analysis.
   (c) Compute a 95% confidence interval for the mean difference between Strain A and Strain B.
   (d) Use ANOVA to test the equality of means among the four strains. Interpret the results within the context of the problem.
4. p. 427, Problem 20 in the textbook.

5. p. 427, Problem 21 (a,b,c) in the textbook.

6. Independent samples of sizes 10, 15, 20, 20, and 25 are taken. The sample variance of the combined sample of 90 observations is 9.0, and the five sample variances are respectively 4.4, 6.2, 3.7, 5.1, and 3.8. Complete an ANOVA table for this data and use the R function `pf()` to find the p-value. *Hint: use sample size information to find the degrees of freedom, the combined sample variance to find the total sum of squares, the separate sample variances to find the error sum of squares (or residual sum of squares), and then use relationships in the table to complete it.*

7. Four populations have means $\mu_1 = 33$, $\mu_2 = 35$, $\mu_3 = 20$, and $\mu_4 = 32$.

   (a) Express these means in a parameterization where $\beta_0 = \mu_1$ and $\beta_i = \mu_{i+1} - \mu_1$ for $i = 1, 2, 3$. Find the numerical values of β_j for $j = 0, 1, 2, 3$.

   (b) Express these means in a parameterization where $\mu_j = \mu + \alpha_j$ for $j = 1, 2, 3, 4$ subject to the constraint that $\sum_{j=1}^{4} \alpha_j = 0$. Find numerical values for μ and α_j for $j = 1, 2, 3, 4$.