STAT 244 FINAL ASSIGNMENT SPRING, 2011

due 5/13/11

1 Binary Trees

1. Write a program which uses a binary tree to order and count a list of 1000 random words or numbers. Include a few entries more than once to check that your program is correctly finding duplicates. Record the time it takes to build the tree.

2. Repeat part a), this time sorting the list of words or numbers first. (See the manual entry for the UNIX sort command.) How does the time compare with the time in part b)?

3. Repeat parts a) and b), this time using the balanced binary tree routines in `~s244/samples/bt` or from the class web page. Compare the performance of the regular and balanced binary tree algorithms.

2 Analysis of Variance

Write a program to perform a general factorial analysis of variance using the reduction technique for data with an arbitrary number of factors. Your program should be able to handle data sets with up to (at least) eight factors. Use binary trees to keep track of the levels of the various factors, so that the only input required should be the number of factors to be read, and possibly the number of observations. Generate all the factorial effects in the model, and for each effect calculate degrees of freedom, sum of squares, mean square, F-ratio, and probability that a F-distributed variable would be larger than the F-ratio obtained. You can use the `dcdflib` libraryto find the F probabilities, and `sas` programs from the samples directory to verify your calculations.
For extra credit, you can add additional features to the program, such as analysis of co-variance, weighted least squares or the ability to remove selected terms from the analysis.