Stat 133, Spr '08
Due Friday, 25 Apr

v-Fold Cross-Validation

In this assignment you are to write an R function that provides the indices for v-fold cross-validation.

Inputs: The parameters of the function include n the population size and v the fold size. Note both n and v are numeric vectors of length 1. The values for v, the fold size, may range from 2 through n. In addition, parameters are needed for the user to specify how to handle the case where n is not an integer multiple of v as explained in the task paragraph below.

Task: The main task is to divide the indices $1, 2, \ldots, n$ randomly into v sugroups of equal size. If v does not divide evenly into n then a few possibilities are:

- fill the short subgroup with NAs;
- spread the NAs a evenly as possible across all groups;
- use a recycling rule to fill in the short group;
- randomly fill in the short group.

You are encouraged to find other approaches that make sense and include them as possibilities.

Output: The function should return a matrix with v columns, corresponding to the v subgroups. The values in each column are the the random indices assigned to the subgroup.

Analyze your code Use the Rprof and summaryRprof functions in R to profile your code. What tasks is your function spending the most time doing? Try an alternative approach to see if you can speed up your code.

Turn in Turn in the two versions of your code, before and after profiling. Your code should be clearly written, efficient, and well documented. Also, make sure that you provide sensible default values for any function parameters other than n and v. Provide a paragraph explanation of how you changed your function, and relate it to the results of your code profile. Include in the write-up, the profile summaries for the two versions of your function.