Generating Statistically Random Subsets

Overview

This is a simple "Warm-up" exercise. The main purpose of this assignment is to demonstrate your problem-solving ability and basic coding skills.

Specifications

Suppose you are working for a pharmaceutical firm conducting a longitudinal study of users of a particular drug. Your team has been asked to administer a "follow up" phone survey in order to estimate the likelihood of various side-effects in the general population. The data you have access to consists of a large set of records containing patient information. The records are accessed via their unique case numbers, which are simply consecutively enumerated integers.

You want to be able to randomly select a certain number of patients to receive survey calls. For example, you might want to contact 50 people out of a set of 250,000 patients.

Your assignment is to write a program that randomly selects the desired number of *unique* call recipients, say k, from a set of patient records of size n.

There is one additional requirement: there is a physical file folder associated with each patient; to make folder retrieval easier, the pollsters want the k numbers output in order.

Your program should:

- accept a k value and an n value from the user
- output k unique values in the range $1..n$
    - Note: the k values should be output in order from smallest to largest

Notes:

- You may use the programming language/platform of your choice.
- Try to analyze/optimize your solution for running time and/or memory usage.
- Be sure to demonstrate good programming style and practices.

Deliverables

- Submit a hard-copy of your source-code, sample output, and a design document.
- Be prepared to present and discuss your solution in class. What data structures did you employ? What algorithm(s) did you use? What interesting problems (and solutions) did you discover?