SESUG Poster 220-2022: Using a Hash Table to Add Diagnosis Related Information to a Health Claims File, Roberta Glass, Abt Associates

ABSTRACT:

PROC SORT is one of the most "expensive" routines in SAS in terms of resources required. When working with health claim files with millions of records, the last thing you want to do is sort the data multiple times. This paper presents a method of adding information from a file of diagnosis (dx) codes and their estimated probability of being an avoidable emergency room visit onto a claims file which does not require sorting the claims by diagnosis code. A hash table in SAS® base is used to create a look-up table of dx codes and their associated probability estimates. When processing the claims file, the probability of the dx code on the claim is located in the hash table and saved to the claims file. In addition to eliminating the need to sort, a hash table allows for the look-up of multiple diagnosis codes on one claim. This technique is intended to serve as an example for the SAS® user who needs to efficiently add information to a large data set.

SUMMARY:

Hash tables are an efficient solution when you need to add information to a large data file. By avoiding the need to sort the data to ready it for a merge, run a merge data step, and then sort the data back into the original order you write less code – saving your time. You also use fewer system resources – saving run time and endearing yourself to your colleagues by not hogging memory.

CONCLUSION/IMPLICATIONS:

Hash tables save the time and resources needed to sort your data for a merge and then sort it back into the order you need for the next step of your analysis. This technique can be generalized for use in many situations, for example you could use a hash table of nursing home characteristics to look-up the number of beds and staffing ratios to add to an analysis file of nursing home stays and use another hash table of patient characteristics to add demographics to the same file. It is worthwhile to consider using a hash table any time you need to avoid sorting a large file multiple times.

CONTACT INFO:

Your comments and questions are valued and encouraged. Contact the author at: Roberta Glass,
Roberta_glass@abtassoc.com, (978) 460-1458