

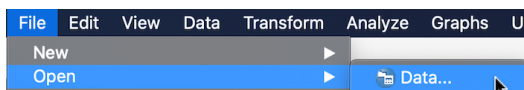
Question: I have to do an analysis using BOOTSTRAP and I've tried but it's taking forever! Is there any way to speed this process up?

Answer: Yes, possibly. Keep in mind that when most procedures run, no matter how many variables are being used, they have to pass ALL the data: Every case and every variable. The more rows and columns used, the more data there is to write and the longer it takes. Many times, we need all the cases; our analysis requires them. However, chances are that we don't need **all** the variables, perhaps just a subset of them. If we can reduce the number of columns to go through, any given analysis will run that much quicker. Sometimes, the difference is dramatic.

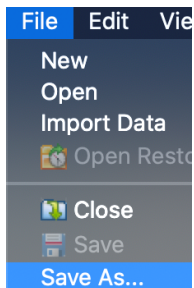
Here is an example, very similar to a real user experience. Suppose that we have a dataset with 16000 cases and 4000 variables. We only need to use 2 variables in our simple REGRESSION problem. If we open the dataset and run REGRESSION with BOOTSTRAP (by default that is 1000 samples, or passing the data 1000 times), it will take REGRESSION about half an hour to display the results. On the other hand, if we reduce the dataset to just the variables we need, the identical analysis takes less than 90 seconds.

Here is what we did:

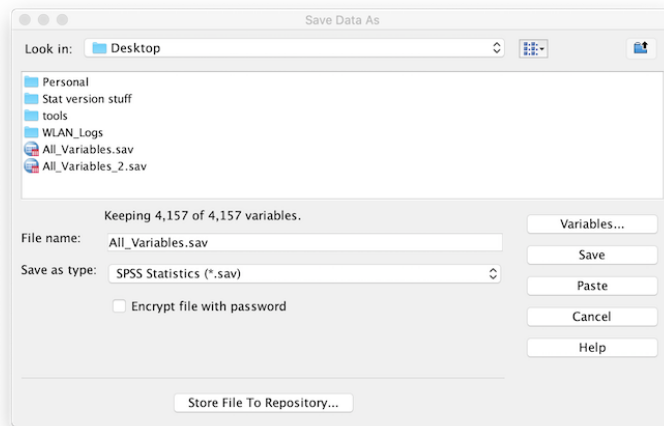
1. Open the original dataset.



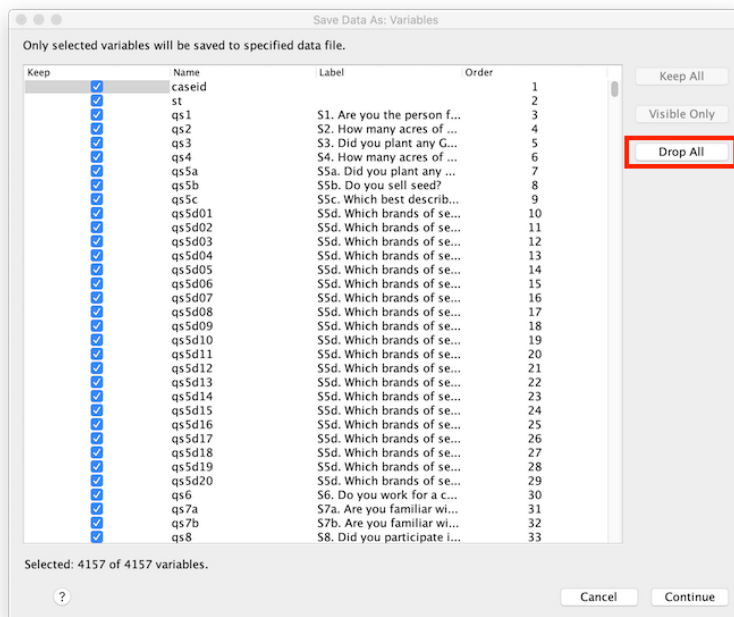
2. Select File > Save As...



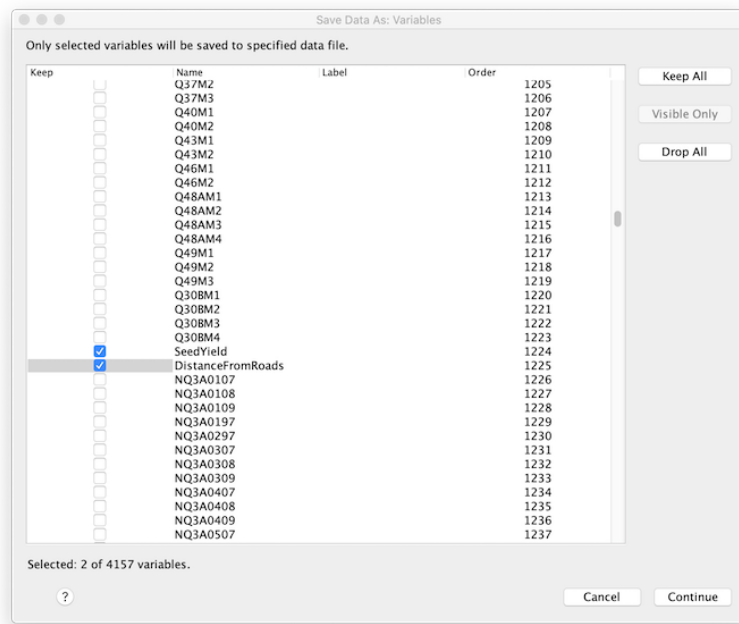
3. Select "Variables..."



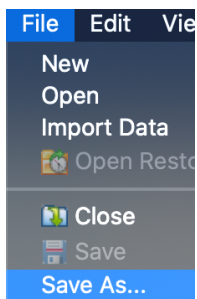
4. Check "Drop All".



5. Find and select just the variables that are required.



6. Save the data to a new file, perhaps in a temporary directory where you can delete it later if you need or want to.



7. Do the analysis on the reduced dataset.