Another useful operation in Stata is when we want to generate a new variable. For example, in our dataset, we have a variable called *gpa*. This variable records students overall grades out of 100. What if I wanted a variable that would contain the grades in letter form? Basically, an A would be a student with a *gpa* that is 90 or higher, a B would be a student with a *gpa* that is between 80 and 90, and so on. Let us call this variable *grade_letter*. We create it using this command:

generate grade_letter = ""

This command tells Stata to generate a new variable called grade_letter and to fill in all the values with an empty string. We cannot tell Stata just to create a variable. We need to tell it the value that is to be stored in the variable. We have instructed Stata to fill in the values with an empty string because what I want to do is to have a string variable in our dataset, since currently there is none. If you look at the right-hand side of the screen, you can see that a new variable has been generated.

Now that we have generated the new variable, we need to populate it with the proper values. To do that we use the **replace** command. What we need to do is to replace the value of *grade_letter* by "F" for those observations with a *gpa* that is less than 60. This is done using the following command:

replace grade_letter = "F" if gpa < 60

Stata tells us that no changes have been made. This is because our dataset contains no observations where the *gpa* is less than 60. Using the same logic, we execute the following series of command:

replace grade_letter = "D" if gpa >= 60 & gpa < 70

replace grade_letter = "C" if gpa >= 70 & gpa < 80

replace grade_letter = "B" if gpa >= 80 & gpa < 90

replace grade_letter = "A" if gpa >= 90 & !missing(gpa)

Notice that in the last command, we specified the not missing option. As you should already know by now, a missing value is considered to be larger than any number, so we need to tell Stata not to include missing observations when performing the last operation.

We can now use the **tabulate** command to look at our new variable:

tabulate grade_letter

We see that most students are C-students followed by B-students.

An interesting piece of information is that there exists a command in Stata that allows us to do all of the above in a single line of code. However, since the purpose of this section is to show users how to create and modify variables, it is more instructive for the student to follow through these steps to get a better understanding of how Stata works. A good exercise would be for you to search for this command online and to try to use it by yourself.

There is one more thing that is missing from the new variable, and that is the label. Remember, it is important to have a well-labelled dataset. Therefore, we execute the following command in order to label the new variable:

label variable grade_letter "The letter grade of the GPA"

I think that this is a nice variable to have in the dataset so I would like to save this version of the dataset:

save, replace