

CSU101 Introduction to Computer Science with Applications Spring 2008

Lab 8

(No labs will be accepted after the due date.)

This lab is a study of 15,183 records from the Registrar. The data was acquired last winter. If you were a student at NU in February 2007, you are somewhere in the data. Each record consists of a state, zip, year of graduation, division, major and qpa. Some of the fields have special peculiarities which will require us to change them to put them in a form that will be more usable for our purposes. The raw data has already been imported into SPSS and can be found in a file called **studentData2.sav**. Right-click this from the course website and saveAs **yourname.sav**.

When you have completed this lab, you should have a modified .sav file, and a new .spo file for the output of your procedures. Each such file should be named with your last name and the appropriate file extension.

Missing Values

Task 1: From the variable view, assign missing values as follows:

Variable	Missing Values
qpa	0
major	100...199
state	blank

This will tell SPSS not to count majors in the 100s, qpa's of zero or cases with missing state information.

Constructing a new Variable

Task 2: We wish to construct a "College" variable to mark which college each student is from. This will be computed from each student's major. Define the new variable as
$$\text{college} = \text{TRUNC}(\text{major} / 100).$$

Dividing the major code by 100 will turn 265 to 2.65, and truncating this will change 2.65 to 2.0. It is the first digit of the major code that we really want.

Task 3: We wish to construct another **year** variable which will tell us which years a student has completed. Someone, in the data, who graduates in 7 (for 2007) was a senior when this data was acquired, and thus should be in her fifth year. Someone, in the data, who graduates in 11 (for 2011) was a freshman when this data was acquired, and thus should be in her first year. To accomplish this we define a new variable called "year" whose value is as follows:

$$\text{year} = 12 - \text{gradyear}.$$

This will allow us to study people with increasing academic age, instead of backwards.

Recode a variable

Task 4: The college variable must be recoded because some of the distinctions it makes have been collapsed into one college. A few years ago we had separate colleges of Pharmacy, Physical Therapy, and Nursing, but they were put into one college. Therefore, the college variable must be recoded as follows:

Old Value	New Value
5	6
7	6
8	6

The effect of this recoding is to map 5s, 6s, 7s and 8s all to one college, college 6. If we recode into the same variable, our original distinctions will be lost. In general, this is not a good idea, but for our purposes it does not matter. If you wish, instead of recoding in place, you could define a new variable which starts out the same as college, say “college2”, which then gets recoded.

Adjusting the Display of certain Variables

Task 5: (Each of the following sub-tasks can be done from Variable View.)

- Adjust the display of qpa so that only three decimal places are visible (as in 2.342).
- Assign value labels to the college variable as follows:

Value	Label
0	COE
2	CCIS
3	CAS
4	CBA
6	Bouve
9	CCJ

- Adjust the major variable so that it shows without decimals.

We have now finished the data manipulation phase and are ready to run some procedures on the data to come up with some answers. The output of your procedures should go in the .spo file. Label the outputs in the .spo file with the question it is supposed to answer.

Task 6: Run a procedure to reveal the average qpa and standard deviation of qpa for the entire population. Descriptives or Frequencies (without frequency tables) are good choices. This should provide the first output to your .spo file.

Task 7: Does qpa change with years in college? Run a procedure to reveal the average qpa for each year. (Hint: Compare Means is useful for this purpose.) This should be the second output to your .spo file.

Task 8: Does qpa by year vary from college to college? (Use Basic Tables, with college going down, and year going across.) This should be the third output to your .spo file.

Task 9: What is the distribution of men and women in each college? You can use Cross Tabulation, but get the count and row % in each cell. This should be the fourth output to your .spo file.

Task 10: How does the average QPA compare between men and women in each college? Again, you can use Basic Tables, but get the average QPA in each cell. This should be the fifth output to your .spo file.

Task 11: What percent of students come from each state? This should be the sixth output to your .spo file. The data are not really very useful for this question because it is obvious that not everyone's home state is recorded. How can you know this?

Task 12: Which 5 states account for the most students and what is the percent of students from each of them? What is the total percent of the student body from those 5 states? To answer these questions you will need to run a procedure, whose output should appear in the .spo file. The answers should be readily apparent from this output.

Task 13: Each state has a distribution of students studying in the various colleges. Compare the distribution of students in the various colleges from NH and VT. What is the count and percent of students from each state (NH or VT) studying in each college? Use **CrossTabs** after you have **Select Cases** on the Data menu.) This should be the last output to your .spo file.

Submit your two files in the usual way by the due date. Make sure your name is in your spo file.