

CSU101

Summer 2009

Lab S2

This lab is a study of 15,000+ student record. 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 **StudentData.sav**. Please start with a fresh copy of the StudentData.sav file (don't use the one from class).

When you have completed this lab, you should have a modified .sav file, and a new .spv file for the output of your procedures.

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. The data is from the 2002-03 school year, so someone who graduates in 2003 was a senior when this data was acquired, and thus should be in her fifth year. Someone, in the data, who graduates in 2007 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:

$$\text{year} = 2008 - \text{gradyear}$$

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. You should recode into a new variable, called “college2”, which then gets recoded based on the values of “college”.

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 (or college2) 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 .spv file.

SPV Output

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 .spv 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 .spv file.

Task 8: What percent of students come from each state? This should be the third output to your .spv file.

Task 9: 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 .spv file. The answers should be readily apparent from this output.

Task 10: 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 .spv file.

Submit your two files in the usual way by the due date.