

Lab 9 *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention `LastnameFirstinitial_Lab#.sas` (e.g., `PileggiS_Lab9.sas`).*

The following data sets contain student information for an introductory statistics course. The course had one large lecture and 12 smaller lab sections. There are 14 data sets as follows:

- `LE1`, `LE2`, ... `LE12` contains information from the official course rosters for the 12 lab sections, downloaded on the first day of class
- `survey` contains information asked in a course survey in the first week of class, and contains background information about the students
- `grades` is a record of the student grades at the end of the quarter

You may use as many data steps and/or procs as needed to achieve the following objectives.

1. Create a macro variable called `path` that corresponds to the computer location of these data sets.
2. Write a macro to import the 14 data sets into SAS (you should use your `path` macro variable).
3. Combine all 14 data sets.
4. Create a variable called `type` that classifies students as follows:
 - (a) `dropped` - these students were present on the official course roster, but not present in the grades data set
 - (b) `crashed` - these students were not present on the official course roster, but present in the grades data set
 - (c) `no survey` - these students are present on the official course roster and have grades, but did not fill out the survey
 - (d) `complete` - these students are present on the official course roster, have survey responses, and have grades
5. Identify the number of students in each classification of the `type` variable. Your results should match those below.

type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
complete	227	94.58	227	94.58
crashed	2	0.83	229	95.42
dropped	3	1.25	232	96.67
no survey	8	3.33	240	100.00

- Create a data set that removes the students who dropped. Use this data set for the remaining exercises.
- All recorded grades are presented in terms of percents. However, the grading rubric was in terms of points. There was a miscommunication on one of the lab assignments where the grader accidentally entered too many points, and so the grade recorded as over 100%. Use a SAS procedure to identify the **lab assignment(s)** that has(have) incorrect grades (ie, do not figure this out by viewing/scanning the data). Note that on homeworks and exams students could earn some extra credit, so grades may be over 100% on those assignments. Verify that your re-coding was done correctly.
- Re-code the **lab assignment(s)** with values exceeding 100 percent to be equal to 100 percent. Verify that your re-coding was done correctly.
- When entering grades for homeworks, labs, or quizzes, the TA entered -99 for missing grades. **Using arrays**, recode all values of -99 to zero. Verify that your re-coding was done correctly.
- Calculate each student's average homework, lab, and quiz grade. (Note, there is no lab 1 grade.) Also, calculate each student's overall course grade, weighted as follows: homework 15%, labs 15%, quizzes 5%, exam 1 20%, exam 2 20% and final exam 25%. Verify that your re-coding was done correctly. The grade distribution should match that shown below.

Variable	N	Mean	Std Dev	Minimum	Maximum
ave_hw	237	88.2317429	14.4169234	7.6923077	99.8384615
ave_lab	237	85.9639944	17.3237839	0	100.0000000
ave_quiz	237	71.3010740	18.1503202	0	100.0000000
grade	237	79.2567350	15.6785324	1.1538462	100.5520979

- Create a new variable in the data set that represents the letter grade for each student such that A=90-100, B=80-89.9, C=70-79.9, D=60-70.9, F=0-59.9. Some students didn't actually complete the course, and should get a letter grade of W for withdrawal. These students can be identified as those that did not take the final exam and so have a grade of 0 for the final exam grade. Verify that your re-coding was done correctly. The letter grade distribution should match that shown below.

letter	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	53	22.36	53	22.36
B	96	40.51	149	62.87
C	51	21.52	200	84.39
D	15	6.33	215	90.72
F	10	4.22	225	94.94
W	12	5.06	237	100.00

12. The professor wants to analyze grades by major. Create a new variable called major that takes the classifications of undeclared, biology, psychology, and other. Make sure that students who have a missing program and plan also have a missing value for major. (Any program and plan that has the word “Biology” in it should be classified as a biology major. Then create the following table that summarizes final course grade by major such that there are four rows corresponding to the four majors, and three columns corresponding to the sample size, the average course grade, and the standard deviation of course grade. The results should match that shown below.

Analysis Variable : grade			
major	N Obs	Mean	Std Dev
Biology	43	80.4852491	13.7115120
Other	36	76.7890459	20.3493980
Psychology	8	73.1990836	11.2460208
Undeclared	148	79.7973975	15.2239011