

Lab 2 – Chemistry 101 at BSU

At Big State University, 182 students are enrolled in Chemistry 101. The class is divided into three sections that meet in the morning, afternoon, or evening. It is now the end of the semester and the student grades and other information is located in a .txt file. Here is what the first few lines of data look like:

class	time	review	hw	midterm	final
fresh	morning	no	508.6	82.5	74.0
fresh	morning	no	464.4	72.0	67.0
fresh	afternoon	no	469.5	69.5	68.0
fresh	morning	no	519.8	84.5	81.0
soph	evening	no	389.1	41.0	37.0
fresh	afternoon	no	389.2	50.0	51.5

Variables:

Class of student (fresh, soph, junior, senior)

The time of day the student took the course (morning, afternoon, evening)

Did the student attend an optional review the day before the Final? (yes, no)

Online hw total (out of 550 points)

Score on the midterm exam (out of 100 points)

Score on the final exam (out of 100 points)

For example, the first student in the data set is a freshman enrolled in the morning section, did not attend the optional review, earned a 508.6/550 on online hw, 82.5/100 on the midterm, and 74/100 on the Final.

Professor Li Zhang would like to add two more variables to the data set:

The percent according to how the hw and exams are weighted. According to the syllabus this variable should be computed as follows:

$$\text{percent} = 25 * (\text{hw}/550) + 35 * (\text{midterm}/100) + 40 * (\text{final}/100)$$

According to the official grading scale used at BSU, the coursegrade is:

A if $90 \leq \text{percent} \leq 100$

B if $80 \leq \text{percent} < 90$

C if $70 \leq \text{percent} < 80$

D if $60 \leq \text{percent} < 70$

F if $0 \leq \text{percent} < 60$

For example, the first student in the data set should have a percent of $25 * (508.6/550) + 35 * (82.5/100) + 40 * (74/100) \approx 81.59$ and a coursegrade of B.

The first thing you need to do is read the data into R and SAS. Download the file chem101.txt from the Canvas page and save it in a convenient place on the computer you're presently working on. The pathnames you see in the following SAS and R code will need to be changed.

SAS

With SAS, we are able to create the new variables percent and coursegrade immediately after reading in the data ☺. You can either type in the SAS program you see below, or download the chem101.sas file from Canvas.

```
data chem101;
infile '/home/campus21/kendall/Desktop/sas desktop/chem101.txt'
firstobs=2;
input class$ time$ review$ hw midterm final;
percent=25*(hw/550)+35*(midterm/100)+40*(final/100);
if (percent >= 90) then coursegrade="A"; else
if (percent >= 80) then coursegrade="B"; else
if (percent >= 70) then coursegrade="C"; else
if (percent >= 60) then coursegrade="D"; else
coursegrade="F";
proc print data=chem101;
run;
```

The SAS program above will print out the entire data set for you. To answer the questions on the following pages, you will need to add more procs to the program.

R

```
> chem101=read.table("/home/campus21/kendall/Desktop/sas
desktop/chem101.txt",header=T)
> attach(chem101)
```

Within R, more work is required to create the new variables percent and coursegrade.

First, take a quick look at the data frame you've created named chem101. You can look at the entire data frame if you want by just typing in `> chem101`, or just check out the first few or last few lines by using the head or tail function.

`> head(chem101)` or `> tail(chem101)`

Similarly, also take a quick look at some of the individual variables in the data frame. They are named class, time, review, hw, midterm, and final.

For the variable `percent`, you'll have to create a new vector then add it to the data frame named `chem101`.

```
> percent=25*(hw/550)+35*(midterm/100)+40*(final/100)
```

Take a quick look at `percent` by just typing in `> percent`. Does it look like a numerical vector of length 182? Try `> length(percent)`.

Now add the `percent` vector to the `chem101` data frame using the `cbind` function.

```
> chem101=cbind(chem101,percent)
```

To determine the course grade for each student, you'll have to create a new function based on the BSU grading scale. Name the function `grade`.

```
> grade=function(x)
{
  if (x>=90) {result="A"} else
  if (x>=80) {result="B"} else
  if (x>=70) {result="C"} else
  if (x>=60) {result="D"} else
  result="F"
  return(result)
}
```

Test the function to make sure it works correctly. Confirm it gives you correct answers for some arbitrary values.

```
> grade(86.3)
> grade(62.9)
```

Finally, use the **`sapply`** function to apply your new function `grade` to all the values in the vector `percent`. Name this new vector `coursegrade` and add it to the data frame `chem101`.

```
> coursegrade=sapply(percent,grade)
> chem101=cbind(chem101,coursegrade)
```

To confirm the data frame `chem101` has been correctly and completely assembled, type in `> head(chem101)`.

SAS and R

Using what you've learned on the SAS and R handouts so far, answer the following questions using SAS, then do it all over again using R and confirm you get the same answers. Round off to the 2nd decimal place:

the mean among all 182 final exam scores = _____

the variance among all 182 midterm scores = _____

the standard deviation among all 182 hw scores = _____

Look at the QQ Plot for the hw scores. Based on what it looks like, does the data set of hw scores appear to be weak-tailed, heavy-tailed, or approximately normal?

Fill in the following tables with counts:

coursegrade

A	B	C	D	F	total
					182

time

morning	afternoon	evening	total
			182

Two-way table of counts: class \times time

	morning	afternoon	evening	totals
fresh				
soph				
junior				
senior				
totals				182

Draw a picture of what the pie chart looks like for the variable class:

A student named Aditya finished with a higher percent than any other student in Chemistry 101 at the end of the semester.

What was Aditya's percent? _____

Professor Li Zhang would like to compare final exam scores between students who attended the optional review the day before the final exam and those who did not. Look at side by side boxplots and fill in the following table (round off to the 2nd decimal place):

attended review?	# of students	mean score	median score
yes			
no			

How many students are freshman who attended the review and earned an A in Chemistry 101?

How many students are not freshman, had a midterm score greater than 70, and were enrolled in the evening section?

How many students were enrolled in the evening section or had a final exam score less than 80?

Be sure to practice importing data from different file types.
You will need this skill for future assignments!

SAS

Read the data set in again, but this time use the file chem101.csv
Read the data set in again, but this time use the file chem101.xlsx
(See pages 2 and 3 of the SAS3 handout.)

R

Read the data set in again, but this time use the file chem101.csv
(See page 5 of the R3 handout.)

**After you have completed this handout,
complete the Canvas quiz titled:
Lab 02 – Chemistry 101**