Instruction

- Please save your solution in **one single file** with the *.sas* extension.
- Please e-mail me (arthur.x.li@me.com) your assignment by the due date. You only need to send me your SAS code.
- Please do not send me the output and the log files.
- Please do not send me a zip file.
- When writing your homework, please follow the template on the last page of the assignment.

Problems from Chapter 1 - Chapter 4

This homework will be based on the materials from Chapters 1 - 4 of the textbook. I strongly recommend that you do all the exercises at the end of each chapter. However, you are only required to turn in one problem of your choice from each chapter. Since you can obtain the solution from the publisher's website, these four problems will be graded on completion only.

Problem 1

The following data set (casecontrol.sas7bdat) contains hypothetical data for a case control study. Cases have values of 1 in the CASE_CONTROL variable and controls have values of 2. Only cases have data for stage and grade variables.

ID	case_control	grade	stage
1	1	3	Ta
1	2	.	
2	1	2	2
2	2	.	
3	1	1	Ta
3	2	.	
4	1	2	4
5	1	2	1
5	2	.	

Based on this data set, you need to create two variables (GRADE_NEW and STAGE_NEW) by using BY group processing. These two variables are created by assigning the stage and grade of each case to their matched control so that each case-control pair is then essentially matched on grade and stage. The resulting data set looks like the one below:

ID	case_control	grade	stage	grade_new	stage_new
1	1	3	Ta	3	Ta
1	2	.		3	Ta
2	1	2	2	2	2
2	2	.		2	2
3	1	1	Ta	1	Ta
3	2	.		1	Ta
4	1	2	4	2	4
5	1	2	1	2	1
5	2	.		2	1

Problem 2

You will work with the *q2.sas7bdat* data set for this problem. Here are the first and last 5 observations:

Obs	id	date	score
1	1	20JUN2009	A
2	1	06SEP2009	C
3	1	16MAR2010	A
4	1	29OCT2009	B
5	1	02MAY2009	C
...			
...			
46	10	12AUG2009	A
47	10	28MAY2009	C
48	10	21JUL2009	B
49	10	14SEP2009	C
50	10	21JUL2009	B

Write a program to subset the dataset above by taking only the two most recent SCORES for each subject. The resulting data set should look like the one below:

Obs	id	date	score
1	1	16MAR2010	A
2	1	29OCT2009	B
3	2	05APR2010	C
4	2	29MAR2010	C
5	3	21DEC2009	B
6	3	01DEC2009	B
7	4	19MAR2010	A
8	4	04JAN2010	B
9	5	08MAR2010	B
10	5	17OCT2009	B
11	6	09NOV2009	A
12	6	26SEP2009	C
13	7	25MAR2010	A
14	7	22MAR2010	B
15	8	05NOV2009	C
16	8	14SEP2009	C
17	9	14MAR2010	B
18	9	06JAN2010	A
19	10	14SEP2009	C
20	10	12AUG2009	A

```
/*
Name(s): Arthur Li
Due Date: November 11
*/

/*
Exercise 1.1
*/
data example;
    infile ....;
    input .....;
    ...
run;

/*
Exercise 2.2
*/
data another_example;
    ...
run;

proc ...;
run;
```