

# STAT 480 Assignment #2

(Due July 16th, Thursday, 11:59 PM)

**Instructions:** Read lectures 2.1, 2.2, 3 and sample codes I used in class carefully before starting your homework. You need to submit a .sas file on ANGEL (Lessons > Homework > Homework#2 > dropbox for HW#2).

**1. (List Input and Assignment Statement).** Below is a small dataset of grades for 6 students.

Name	Gender	Homework	Midterm	Final
Beck	Male	90	88	75
Christal	Female	100	85	80
Gavetti	Female	80	77	71
Easton	Male	90	82	76
Mattiace	Male	60	67	53
Chen	Female	85	83	68

Table 1: Grades

- (a) Use **column input** to read the dataset above called *Grades* and store it in the folder **MyLibrary1** (the same folder we used for assignment#1, `X:/STAT480/SAS library/MyLibrary1`).
- (b) Define Two new variables *Score1* and *Score2*, where  
 $\text{Score1} = 40\% \text{ Homework} + 20\% \text{ Midterm} + 40\% \text{ Final}$ , and  
 $\text{Score2} = 40\% \text{ Midterm} + 60\% \text{ Final}$ .
- (c) Calculate the final score. The final score, represented by a new variable *fs*, is defined to be the maximum of the following three variables: Final, Score1 and Score2.
- (d) Curve the final score *fs*. The curved new final score, represented by *newfs*, is defined as

$$\text{newfs} = 0.7 \times \text{fs} + 30$$

- (e) Round the new final score *newfs* to the nearest multiple of 5. (**Hint:** Use round function, `round(argument1, argument2)`, read SAS documentation for more details.)
- (f) Print out the whole dataset, including all the new variables you created except the variable *Gender*. (Use `var` statement).

**2. (Formatted Input).** Below is a dataset of all the games between Argentina and Germany in the FIFA World Cup. The descriptions of the variables are:

- game: the number of the game
- year: the year of the World Cup
- date: the date of the game
- stage: the stage of the game (a character variable)
- Arg: the score of Argentina
- Ger: the score of Germany
- Attendance: the number of audience

game	year	date	stage	Arg	Ger	Attendance
1	1958	08Jun1958	1st Round, Group1	1	3	31,156
2	1966	16Jul1966	1st Round, Group2	0	0	46,587
3	1986	29Jun1986	Final Game	3	2	114,600
4	1990	08Jul1990	Final Game	0	1	73,603
5	2006	30Jun2006	1/4 Finals	1	1	72,000
6	2010	03Jul2010	1/4 Finals	0	4	64,100
7	2014	13Jul2014	Final Game	0	1	74,738

Table 2: Germany vs Argentina in World Cup

- Use **formatted input** to read the dataset above called *games*. When reading the dataset (1) use an absolute pointer control to read *year* and *date*, and use a relative pointer controls to read *stage* and *Attendance*; (2) Use informat *dateW.* for variable 'date' and use informat *commaW.* for variable 'Attendance', where W is an integer which specifies the number of columns that variable will occupy.
- Set the page size to 50, line size to 80, center the output, and suppress the date and time when the output was created.
- Print the data set games. (1) Remember formatted input does not imply formatted output. Use **format** statement to format your output (see the last slide of lecture2.2); (2) Add a title for this dataset.