NATIONAL UNIVERSITY OF SINGAPORE
Department of Statistics and Applied Probability

2018/19 Semester 2 **ST2137 Computer Aided Data Analysis** **Tutorial 1**

Three datasets ("tut1htwt.csv", "tut1test.csv", and "tut1htwtfixed.txt") have been uploaded to the course website in the IVLE. They are used for Questions 1 to 7.
There are five variables in the "tut1htwt.csv" and "tut1htwtfixed.txt" datasets. They are

- id: identity of the subject (Columns 1-3)
- gender: gender of the subject (Column 4)
- height: height of the subject in cm (Columns 5-7)
- weight: weight of the subject in kg (Columns 8-9)
- siblings: number of siblings of the subject (Column 10)
  The column numbers in parentheses are the columns occupied by these variables in the "tut1htwtfixed.txt" file.

There are two variables in the "tut1test.csv" dataset. They are

- id: identity of the subject
- test: test score of the subject

1. Create SAS datasets "htwt" and "test" by importing the "tut1htwt.csv" file and "tut1test.csv" respectively into the SAS.

2. Based on the SAS data set "htwt", create a SAS dataset "htwtf" which contains the data for all the female subjects. How many females are there in the dataset "htwt"?

3. Merge the two SAS datasets "htwt" and "test". Let us call this new SAS dataset "htwttest". Identify individuals who are taller than 184 cm by creating a relevant SAS data set. What are the test scores of subjects who are taller than 184 cm?

4. Create a SAS dataset "htwtfixed" by importing the fixed format text file "tut1htwtfixed.txt" into the SAS.

5. Suppose that there was an error in the weight of the Subject 356 in the text file. Obtain a new SAS dataset "htwtfixedremo" by removing the record related to the Subject 356 from the SAS dataset "htwtfixed" .

6. Who is the second tallest female in this group and what are her height, weight, and test score?

7. Create a new variable called "grade" in the data set "htwttest" using the following rule: (1) grade = "A" if test ≥ 80, (2) grade = "B" if $70 \leq$ test < 80, (3) grade = "C" if $60 \leq$ test < 70, (4) grade = "D" if $50 \leq$ test < 60 and (5) grade = "F" if test < 50.
   How many male students get a "F" grade?

8. A chemist has to run an experiment to study the effect of four treatments on the glass transition temperature of a certain polymer compound. Raw material used to make this polymer is in small batches. The material is thought to be fairly uniform within a batch but to vary between batches. The results of the experiment are given below:

Batch	Treatment			
	1	2	3	4
1	303	311	289	270
2	242	290	259	263
3	289	282	277	257

(a) Write a SAS program to create a SAS data set from the above table using do loops. How many variables are there in your SAS data set? What are these variables?

(b) Repeat part (a) without using any do loops. (Hint: How many variables are there in the **input** statements?)

(c) Write a SAS program to create a SAS data set from the table below using do loops.

Batch	Treatment			
	1	2	3	4
1	303	311	289	270
2	242		259	263
3	289	282	277	257

Answers to selected questions

2. 161 females
3. 2 individuals. Test scores are 62 (ID = 160) and 93 (ID = 367).
6. ID = 375. Height = 173 cm, weight = 56 kg, test score = 85.
7. 15 males get a "F" grade