

THE CHINESE UNIVERSITY OF HONG KONG  
STAT2005 Programming Languages for Statistics  
Take Home Test

**Answer ALL questions.**

**Use R programming language to answer Question 1-2 and save your answers in a script file named "[Your\_Student\_ID].r", where [Your\_Student\_ID] is your 10-digit student ID.**

**Question 1 (14 marks)**

The use of `det ( )` function is prohibited in this question.

Let  $A$  be a  $n$ -by- $n$  matrix and  $|A|$  be its determinant,  $|A|$  could be computed using the following algorithm.

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| - a_{14}|A_{14}| + \dots + (-1)^{n+1}a_{1n}|A_{1n}|,$$

where  $A_{ij}$  is a  $(n-1)$ -by- $(n-1)$  matrix that results from deleting row  $i$  and column  $j$  of  $A$ . Write a function named `my_det ( )` which takes a square matrix and returns the value of its determinant. Your function should check if the input matrix is a square matrix and return the error message "Input matrix is not a square matrix." when necessary.

**Question 2 (14 marks)**

Secant method is an alternative root-finding algorithm in numerical analysis to solve  $f(x) = 0$ . Given two initial values  $x_0$  and  $x_1$ , the successive terms are defined recursively as follows.

$$x_n = x_{n-1} - f(x_{n-1}) \frac{x_{n-1} - x_{n-2}}{f(x_{n-1}) - f(x_{n-2})}, \quad n = 2, 3, \dots$$

Implement the Secant method as a R function in the following form

```
secant <- function(f, x0, x1, n, err) { ... }
```

where `f` is a function to evaluate  $f(x)$ , `x0` and `x1` are the values of  $x_0$  and  $x_1$  respectively, and `err` is the error tolerance in absolute value of  $f(x)$ . Try to solve

$$3e^{-y \times 0.5} + 3e^{-y \times 1.0} + 3e^{-y \times 1.5} + 103e^{-y \times 2.0} = 98.39$$

using your function and compare the number of iterations of the Secant method with the bisection method as discussed in the lecture notes.

Use SAS programming language to answer Question 3-4 and save your answers in a script file named “[Your\_Student\_ID].sas”, where [Your\_Student\_ID] is your 10-digit student ID.

### Question 3 (36 marks)

We have a data set of four variables on the murder rates (per 100,000 inhabitants) for cities from South, North, and West in USA in 1980 as follows:

city	state	coast	rate
-----	-----	-----	-----
Denver	CO	West	9
Los Angeles	CA	West	23
San Diego	CA	West	10
Atlanta	GA	South	14
Dallas	TX	South	18
Washington	DC	South	11
Chicago	IL	North	14
Cleveland	OH	North	16
Detroit	MI	North	16
Madison	WI	North	2

Write SAS programs to read the murder rates data in (a) - (c) as in-stream data. Store the data in a temporary file, called Q3.

You can copy and paste the data lines from Q3.txt. ‘#’ in the data lines are not part of the data.

The order of the variables in the file must be CITY, STATE, COAST, and RATE.

(a) (12 marks)

```

Denver CO      West      9 #####
Los Angeles   CA        West     23
San Diego    CA         West     10
Atlanta      GA         South    14
Dallas TX     South 18
Washington   DC      South    11
Chicago      IL      North     14
Cleveland    OH        North    16
Detroit      MI        North     16
Madison      WI        North      2

```

(b) (12 marks)

```
Denver      CO      West      9 Los Angeles  CA
  West 23 San Diego  CA      West 10
Atlanta GA      South 14 Dallas      TX      South
18 Washington  DC      South 11
Chicago      IL
  North 14 Cleveland OH      North 16
Detroit      MI      North 16 Madison
WI      North 2
```

(c) (12 marks)

```
Denver CO      ## West      9 #####
  Los Angeles  CA # West 23
San Diego CA      ## West 10
  Atlanta GA # South 14
Dallas TX      # South 18
  Washington  DC # South 11
Chicago IL ## North 14
  Cleveland OH # North 16
Detroit MI # North 16
  Madison WI ### North 2
```

Hint: For part (a), you can use the following template.

```
data Q3;
  * Your input and/or length statements here;
cards;
Denver CO      West      9 #####
  Los Angeles  CA      West 23
San Diego CA      West 10
  Atlanta GA      South 14
Dallas TX      South 18
  Washington  DC      South 11
Chicago IL      North 14
  Cleveland OH      North 16
Detroit MI      North 16
  Madison WI      North 2
run;
```

Part (b) and (c) are similar by replacing the in-stream data between the `cards;` and `run;` statements.

#### Question 4 (36 marks)

Write a SAS program to read the following information about Tom's Tuesday appointments as in-stream data.

Time	With	Place	Subject	Length_Meeting	Confirmed?
11:00	Li Lan	Room 30	Personnel review	45 minutes	Yes
13:00	Leung Mei Fai	Leung's office	Marketing	30 minutes	No
15:00	Mak David	Lab	Test results	20 minutes	Yes

The SAS data set should be a permanent one locating in D:\ with filename Q4. The data set should contain seven variables: TIME, LAST\_NAME, FULL\_NAME, PLACE, SUBJECT, LENGTH\_MEETING, CONFIRM.

TIME and LENGTH\_MEETING are numeric and all other variables are character.

The data are listed below in (a) - (c).

You can copy and paste the data lines from Q4.txt. '#' in the data lines are not part of the data.

##### (a) (12 marks)

11:00	Li Lan	Room 30	Personnel review	45 minutes	Yes
13:00	Leung Mei Fai	Leung's office	Marketing	30 minutes	No
15:00	Mak David	Lab	Test results	20 minutes	Yes

##### (b) (12 marks)

11:00	Li Lan	Room 30	Personnel review	45	Yes#
13:00	Leung Mei Fai	Leung's office	Marketing	30	No
15:00	Mak David	Lab	Test results	20	Yes

##### (c) (12 marks)

11:00	Li Lan	Room 30	Personnel review	45 mins	Yes
13:00	Leung Mei Fai	Leung's office	Marketing	30 mins	No
15:00	Mak David	Lab	Test results	20 mins	Yes

**End of Questions**