

114-1-Programming and Statistical Software (Midterm) Exam Paper

Date: 2025/10/23(Thu) 9:20~11:30 (140 minutes)

Page 1/8

Important Notes

1. **Time** (Please sit in your usual seat):
 - (a) During the exam, the network will be disconnected (9:20~11:30). You can only use the public computers and your personal USB drive to answer. Laptops, tablets, and mobile phones are not allowed (even with network functions turned off).
 - (b) Exam paper download: Within 5 minutes before the exam, download the exam paper from the teaching website to the public computer and save the file on your USB drive.
 - i. Students who are late or need to (re)download the exam paper after the network is disconnected should bring their USB drive to the teacher's laptop to download (points may be deducted).
 - ii. The download file is: 1141-StatSoft-Midterm.zip. Extract the file to the desktop or USB drive. The resulting folder should be "1141-StatSoft-Midterm". (Don't just double-click the zip file and then get stuck! Don't tell the teacher you don't know how to extract files! Don't tell the teacher you don't know where the file was downloaded!)
 - iii. It contains the exam paper, answer sheet, datasets, and some R packages (zip files). (Don't tell the teacher you don't know how to install R packages locally!)
 - (c) Answer sheet upload: Upload the answer sheet within 5 minutes after the exam. [Students who want to submit early, please bring your USB drive to the teacher's laptop to upload.]
2. This exam must be answered using the R program (Rgui or RStudio). Other programs are not allowed.

3. During the exam, you may consult any materials, including textbooks, teaching handouts, and electronic files (documents, videos/audio (please use headphones)).
4. Communication via FB Messenger, IG, Line, etc., to discuss with others is prohibited. The use of chatGPT or similar AI-assisted tools is prohibited. Cheating (or suspected cheating) is prohibited. [Please!! Don't be the one who spoils it for everyone].

5. **Answer Sheet:**

- (a) Please use the provided answer sheet template: The MS WORD filename should be "Name-StatSoft-Midterm.docx" (change "Name" to your own).
- (b) On the answer sheet, be sure to copy and paste the "executed code and results (including graphs)" from the R Console onto this answer sheet (Courier New, 10pt font, black text on white background). It should not be only the code, nor only the output. (The teacher demonstrated this in class; please refer to the teaching recordings.)
- (c) Please indicate the question numbers in order: e.g., # ex1(a), # ex1(b), # ex2, etc.
- (d) If the answer is incomplete or there are execution errors, also copy and paste the "executed code and results (including graphs)" from the R Console onto the answer sheet. Blank answers cannot receive partial credit.

6. **Uploading the Answer Sheet:**

- (a) Log in to the [Homework and Exam Upload Area] on the course website or go to <http://hmwu.nccu.edu.tw/login.html?lang=tchinese>.
- (b) Account: statsoft, Password: xxxx (announced on the course FB), Folder: "20251023-Midterm".
- (c) If the upload website shows a "blank page", please move the mouse to the "address bar" and press "Enter". If it still doesn't work, try another browser (IE/Edge/Firefox/Chrome).
- (d) Upload the file. After confirming the file size is correct, the submission is complete. Files cannot be deleted or re-uploaded. If there are problems, please contact the teacher.
- (e) Do not tell the teacher that your answer file is missing or has become blank; the teacher cannot solve this either.

- (f) During the network disconnection, students who wish to submit early should bring their USB drive to the teacher's laptop to upload. After uploading the file, you may leave.
 - (g) You can practice uploading any file to the folder: "upload_testing" following the steps above one day before the exam.
7. If there is an unsolvable problem with the public computer, please move yourself to the "Free Area" (the last two rows in Wei San with unassigned public computers).
 8. If you have questions, please try to solve them yourself first, or raise your hand to ask the teacher. Do not contact or discuss with other students.
 9. If the public computer cannot read your personal USB drive, telling the teacher there is a problem won't help solve it. (You should have tested whether it can be read during regular class.)
 10. During the exam, there is only one teacher and one teaching assistant. Students with questions, please remain calm and wait for the teacher or TA to come and understand the problem.
 11. If needed before the exam, please use the restroom. During the exam, you cannot enter or leave the classroom arbitrarily for any reason (unless under special circumstances).
 12. Please be sure to read the important notes carefully first, to ensure a smooth and successful exam process.

R Questions (6 major questions, total 140 points)

1. (10 points) Use the `cat`, `scan` commands to print prompts, read in your department, student ID, and name respectively. Based on the input data, print the following pledge.

"I, (Department, Student ID, Name), solemnly pledge to abide by all examination rules.

If violated, I am willing to accept the most severe punishment from the school. So sworn."

2. (10 points) Using only the R command `gl`, perform a completely randomized design in an experimental design according to the following scenario: 2 groups (placebo, treatment), total of 20 subjects, 10 subjects per group. The code is as follows, the "... " is the part the student needs to complete.

```
> set.seed(123) # Set random seed for reproducible results
> ...
> sample(exp)
[1] treatment treatment treatment placebo placebo placebo placebo
[8] treatment placebo placebo treatment placebo treatment treatment
[15] treatment placebo treatment placebo placebo treatment
Levels: placebo treatment
```

3. (20 points) Please write an R function. The input is the "number of layers" `n`. Use `while` (cannot use `for`) to output the following sequence pattern (when this R function is executed with `n` as 4 and 7).

(a) $n = 4$

1
1 2
1 2 3
1 2 3 4

(b) $n = 7$

1
1 2
1 2 3
1 2 3 4
1 2 3 4 5
1 2 3 4 5 6
1 2 3 4 5 6 7

4. (30 points) The test results of a certain exam paper are recorded in "answer.xlsx". The correct answers for the 10 multiple-choice questions in the exam paper are sequentially B, D, B, D, D, A, C, D, C, B.

- (a) Please read this data and print the first and last 5 records.
- (b) If a student's answers are: A, D, B, D, B, A, B, D, C, B, which questions did he answer correctly? If one correct answer earns 10 points, what is the student's total score?
- (c) If one correct answer earns 10 points, please calculate each person's total score and print the score table.

```
> score.table
0 10 20 30 40 50 60 70 80 90 100
9 18 16 9 18 19 27 34 25 10 6
```

- (d) If the top 20% (inclusive) of total scores are defined as the high-score group, and the bottom 20% (inclusive) of total scores are defined as the low-score group, then which students are in the high-score group, which are in the low-score group, and how many people are in each group?
- (e) Calculate the percentage of people who answered each question correctly in the high-score group and low-score group, denoted as P_H and P_L .
- (f) Please calculate the difficulty index (formula $P = (P_H + P_L)/2$) and discrimination index (formula $D = P_H - P_L$) for each question.

5. (40 points) Denote the observed data of two variables as $\{x_i, y_i\}_{i=1}^n$, and perform simple linear regression analysis. In simple linear regression ($y = \beta_0 + \beta_1 x + \epsilon$), the estimators for the slope (β_1) and intercept (β_0) are as follows:

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \text{where } \bar{x} \text{ and } \bar{y} \text{ are the means of } x_i \text{'s and } y_i \text{'s.}$$

- (a) (30 points) Write an R function. The input is the data of two variables, and the output is the estimators for the slope and intercept.

- (b) (10 points) The data for the two variables are as follows: `x <- iris[,1]; y <- iris[,2]`. Use the function from the previous question to calculate the estimators for the slope and intercept, and compare the results with those from `lm(y~x)`.

6. (30 points)

A bag contains 6 white balls and 4 red balls. Randomly draw 3 balls from it (without replacement). If $P(A)$ represents the probability of drawing 2 white balls and 1 red ball, find $P(A)$.

sol:

$$P(A) = \frac{C_2^6 C_1^4}{C_3^{10}} = \frac{1}{2}.$$

Xiao Ming wants to simulate drawing balls using programming to calculate this probability. (That is, we can estimate the probability of an event occurring by calculating the "proportion" of the number of times the event occurs to the total number of experiments).

- (a) (10 points) If `set.seed(123456)` is set, list the result of one experiment of "drawing 3 balls randomly from a bag containing 6 white balls and 4 red balls", and count and print the number of white balls and red balls that appear.
- (b) (10 points) Same as the previous sub-question, repeat the above experiment 10 times, count and print the number of white balls and red balls that appear each time, as shown below.

```
> DrawResult
White_balls Red_balls
1          1      2
2          2      1
3          2      1
4          2      1
5          0      3
6          1      2
7          2      1
8          2      1
9          1      2
10         1      2
```

- (c) (10 points) Same as the previous sub-question, repeat the above experiment 1000 times, and calculate the probability of drawing 2 white balls and 1 red ball.