

MAS1802: Problem Solving II (Statistical Computing with R)

Dr Lee Fawcett

School of Mathematics, Statistics & Physics

Semester 2, 2019/2020

Part VII

Revision

Post-COVID assessment for MAS1802

- NUMBAS pass/fail exam covering **all** aspects of MAS1802 – problem solving **and** computing
- Computing group project was cancelled – hence computing included in the exam
- The three computing practicals you submitted before Easter will still stand
- Around 40% of the marks in the exam will come from questions from my part of the module
- Multiple attempts will be allowed, and the test will be available over an extended period

Post-COVID assessment for MAS1802

- NUMBAS pass/fail exam covering **all** aspects of MAS1802 – problem solving **and** computing
- Computing group project was cancelled – hence computing included in the exam
- The three computing practicals you submitted before Easter will still stand
- Around 40% of the marks in the exam will come from questions from my part of the module
- Multiple attempts will be allowed, and the test will be available over an extended period

Post-COVID assessment for MAS1802

- NUMBAS pass/fail exam covering **all** aspects of MAS1802 – problem solving **and** computing
- Computing group project was cancelled – hence computing included in the exam
- The three computing practicals you submitted before Easter will still stand
- Around 40% of the marks in the exam will come from questions from my part of the module
- Multiple attempts will be allowed, and the test will be available over an extended period

Post-COVID assessment for MAS1802

- NUMBAS pass/fail exam covering **all** aspects of MAS1802 – problem solving **and** computing
- Computing group project was cancelled – hence computing included in the exam
- The three computing practicals you submitted before Easter will still stand
- Around 40% of the marks in the exam will come from questions from my part of the module
- Multiple attempts will be allowed, and the test will be available over an extended period

Post-COVID assessment for MAS1802

- NUMBAS pass/fail exam covering **all** aspects of MAS1802 – problem solving **and** computing
- Computing group project was cancelled – hence computing included in the exam
- The three computing practicals you submitted before Easter will still stand
- Around 40% of the marks in the exam will come from questions from my part of the module
- Multiple attempts will be allowed, and the test will be available over an extended period

The R part of the exam

Questions split across the following broad areas:

1. Dataframes and data manipulation

- See Chapters 2 and 3 of the lecture notes and slides
- See Practical 1 (Tuesday 10th March), especially question 4!
- Material can be found in the 'Introduction' and 'Data Summaries' folders in BB

2. Functions and 'Control Flow'

- See Chapter 4 of the lecture notes and slides
- See Practical 3 (Tuesday 24th March) – all questions, but especially the assignment questions!
- Material can be found in the 'Control statements and functions' folder in BB

3. Randomness

- See Chapter 5 of the lecture notes and slides
- Give particular attention to Section 5.4.3!
- Material can be found in the 'Randomness' folder in BB

The R part of the exam

Questions split across the following broad areas:

1. Dataframes and data manipulation

- See Chapters 2 and 3 of the lecture notes and slides
- See Practical 1 (Tuesday 10th March), especially question 4!
- Material can be found in the 'Introduction' and 'Data Summaries' folders in BB

2. Functions and 'Control Flow'

- See Chapter 4 of the lecture notes and slides
- See Practical 3 (Tuesday 24th March) – all questions, but especially the assignment questions!
- Material can be found in the 'Control statements and functions' folder in BB

3. Randomness

- See Chapter 5 of the lecture notes and slides
- Give particular attention to Section 5.4.3!
- Material can be found in the 'Randomness' folder in BB

The R part of the exam

Questions split across the following broad areas:

1. Dataframes and data manipulation

- See Chapters 2 and 3 of the lecture notes and slides
- See Practical 1 (Tuesday 10th March), especially question 4!
- Material can be found in the 'Introduction' and 'Data Summaries' folders in BB

2. Functions and 'Control Flow'

- See Chapter 4 of the lecture notes and slides
- See Practical 3 (Tuesday 24th March) – all questions, but especially the assignment questions!
- Material can be found in the 'Control statements and functions' folder in BB

3. Randomness

- See Chapter 5 of the lecture notes and slides
- Give particular attention to Section 5.4.3!
- Material can be found in the 'Randomness' folder in BB

The R part of the exam

Questions split across the following broad areas:

1. Dataframes and data manipulation

- See Chapters 2 and 3 of the lecture notes and slides
- See Practical 1 (Tuesday 10th March), especially question 4!
- Material can be found in the 'Introduction' and 'Data Summaries' folders in BB

2. Functions and 'Control Flow'

- See Chapter 4 of the lecture notes and slides
- See Practical 3 (Tuesday 24th March) – all questions, but especially the assignment questions!
- Material can be found in the 'Control statements and functions' folder in BB

3. Randomness

- See Chapter 5 of the lecture notes and slides
- Give particular attention to Section 5.4.3!
- Material can be found in the 'Randomness' folder in BB

The R part of the exam

What will the computing questions look like?

1. Using R alongside NUMBAS

- One of the questions will require you to follow instructions to download a dataset
- You will then have to use R to query that dataset and then enter your answers to questions within NUMBAS
- Multiple attempts: You will have to re-download the dataset each time, and re-answer the questions – the dataset will change on each attempt!
- Another question will ask you to write some code to find the answer to a problem – just using NUMBAS to record your final answer (no code to be uploaded)

2. Multiple choice questions within NUMBAS

- For example: 'Which line of code matches the task?'; 'Which function contains a bug?'
- You might know the answer straight away...
- ... or you might use R, at the side, to check!

The R part of the exam

What will the computing questions look like?

1. Using R alongside NUMBAS

- One of the questions will require you to follow instructions to download a dataset
- You will then have to use R to query that dataset and then enter your answers to questions within NUMBAS
- Multiple attempts: You will have to re-download the dataset each time, and re-answer the questions – the dataset will change on each attempt!
- Another question will ask you to write some code to find the answer to a problem – just using NUMBAS to record your final answer (no code to be uploaded)

2. Multiple choice questions within NUMBAS

- For example: 'Which line of code matches the task?'; 'Which function contains a bug?'
- You might know the answer straight away...
- ... or you might use R, at the side, to check!

The R part of the exam

What will the computing questions look like?

1. Using R alongside NUMBAS

- One of the questions will require you to follow instructions to download a dataset
- You will then have to use R to query that dataset and then enter your answers to questions within NUMBAS
- Multiple attempts: You will have to re-download the dataset each time, and re-answer the questions – the dataset will change on each attempt!
- Another question will ask you to write some code to find the answer to a problem – just using NUMBAS to record your final answer (no code to be uploaded)

2. Multiple choice questions within NUMBAS

- For example: 'Which line of code matches the task?'; 'Which function contains a bug?'
- You might know the answer straight away...
- ... or you might use R, at the side, to check!

The R part of the exam

What will the computing questions look like?

1. Using R alongside NUMBAS

- One of the questions will require you to follow instructions to download a dataset
- You will then have to use R to query that dataset and then enter your answers to questions within NUMBAS
- Multiple attempts: You will have to re-download the dataset each time, and re-answer the questions – the dataset will change on each attempt!
- Another question will ask you to write some code to find the answer to a problem – just using NUMBAS to record your final answer (no code to be uploaded)

2. Multiple choice questions within NUMBAS

- For example: 'Which line of code matches the task?'; 'Which function contains a bug?'
- You might know the answer straight away...
- ... or you might use R, at the side, to check!

The R part of the exam

What will the computing questions look like?

1. Using R alongside NUMBAS

- One of the questions will require you to follow instructions to download a dataset
- You will then have to use R to query that dataset and then enter your answers to questions within NUMBAS
- Multiple attempts: You will have to re-download the dataset each time, and re-answer the questions – the dataset will change on each attempt!
- Another question will ask you to write some code to find the answer to a problem – just using NUMBAS to record your final answer (no code to be uploaded)

2. Multiple choice questions within NUMBAS

- For example: 'Which line of code matches the task?'; 'Which function contains a bug?'
- You might know the answer straight away...
- ... or you might use R, at the side, to check!

The R part of the exam

What will the computing questions look like?

1. Using R alongside NUMBAS

- One of the questions will require you to follow instructions to download a dataset
- You will then have to use R to query that dataset and then enter your answers to questions within NUMBAS
- Multiple attempts: You will have to re-download the dataset each time, and re-answer the questions – the dataset will change on each attempt!
- Another question will ask you to write some code to find the answer to a problem – just using NUMBAS to record your final answer (no code to be uploaded)

2. Multiple choice questions within NUMBAS

- For example: ‘Which line of code matches the task?’, ‘Which function contains a bug?’
- You might know the answer straight away...
- ... or you might use R, at the side, to check!

1. Dataframes and data manipulation

Let's try some questions from practical 1...

2. Functions and control flow

Consider the following R code:

```
h = function(x){  
  l = 6  
  u = 11  
  t = 0  
  for(j in 2:u){  
    t = t + x[j]}  
  return(t)}  
h(x)/length(x)
```

Given a numeric vector x , what does the function h return?

$$\sum_{i=6}^{11} x_i$$

Mean of x

$$\sum_{i=1}^{11} x_i - x_1$$

Mean of the 2nd to the 11th values

The length of the vector

2. Functions and control flow

2. Functions and control flow

$$l = 6$$

$$u = 11$$

$$t = 0$$

$$\underline{j}$$

$$2$$

$$3$$

$$4$$

$$11$$

$$\underline{t}$$

$$0 + x_2 = x_2$$

$$(x_2 + x_3)$$

$$(x_2 + x_3) + x_4$$

$$(x_2 + \dots + x_{10}) + x_{11}$$

$$x_2 + x_3 + \dots + x_{11}$$

"distractors": $l, j, h(x) / \text{length}(x)$

2. Functions and control flow

Consider the following R code:

```
p = function(l){  
  x1 = 0  
  for(i in l to -1){  
    x1 = x1+exp(i)/i**2  
  }  
  return(x1)}
```

- (a) There are two bugs in this code. Fix these bugs, and run the code using $l = -5$. What is your answer?
- (b) What is this function computing?

$$x_1 + \frac{e^i}{i^2} \qquad \sum_{i=1}^{-1} \frac{e^i}{i^2} \qquad \sum_{i=-5}^{-1} \frac{e^i}{i^2}$$

2. Functions and control flow

Consider the following R code:

```
Investigate = function(x){  
  a = mean(x)  
  b = sort(x)[(length(x))/2]  
  l = length(x)  
  c = var(x)**0.5  
  X = c(a, b, c)  
  return(X)}  

```

Given a numeric vector x , which elements below does Investigate return?

- (a) The mean, median and variance
- (b) The mean, variance and length
- (c) The mean, standard deviation and something else

2. Functions and control flow

Example

n odd: 1, 2, ②, 4, 8
median = 2

$$\text{Position: } \frac{n+1}{2} = \frac{6}{2} = 3$$

n even: 1, 2, ②, ④, 8, 9

Position:

$$\text{median} = 3$$

$$\frac{n}{2}, \frac{n}{2} + 1$$

$$3, 4$$

Try in R!

3. Randomness

Consider the following R code:

```
Y = sample(seq(0, 1), 1000, replace = TRUE, prob = NULL)
```

What does this R code do?

- (a) Sample 1000 observations from Y , where $Y \sim \text{Uniform}(0, 1)$
- (b) Randomly sample the values 0 and 1 one thousand times
- (c) Randomly sample values between 0 and 1 one thousand times
- (d) Random sample the values 0 and 1, one thousand times, with probability zero

3. Randomness

Suppose you have $Y \sim \text{Poisson}(3)$. Which line of R code would you use to find $P(Y \geq 4)$?

- (a) `1-ppois(3, 3)`
- (b) `dpois(4, 3)`
- (c) `1-dpois(3, 3)`
- (d) `1-ppois(4, 3)`
- (e) `1-pbinom(3, 3, 0.75)`

3. Randomness

Suppose you have $X \sim \text{Binomial}(10, 0.2)$. Which line of R code would you use to find $P(X \leq 4)$?

- (a) `1-pbinom(4, 10, 0.2)`
- (b) `pbinom(4, 10, 0.2)`
- (c) `dbinom(4, 10, 0.2)`
- (d) `pbinom(10, 0.2, 4)`
- (e) None of the above

3. Randomness

You have a coin that you suspect is slightly biased in favour of Heads. You wish to simulate 1000 flips of this coin. Which line of code would do this?

- (a) `sample(c("T", "H"), 1000, replace = TRUE, prob = c(0.48, 0.5))`
- (b) `rbinom(1000, 1, 0.52)`
- (c) `rbinom(1, 1000, 0.52)`
- (d) `sample(c("H", "T"), 1000, replace = TRUE, prob = c(0.75, 0.25))`
- (e) `rbinom(1000, 1000, 0.47)`