

IMPERIAL COLLEGE LONDON

MSc COURSE IN COMPUTATIONAL METHODS IN ECOLOGY AND EVOLUTION

EXAM 1

For Internal Students of Imperial College of Science, Technology and Medicine

Exam Date: Friday, 15th Jan 2016, 1000 – 1300

Length of Exam: 3 HOURS

Instructions: All sections are weighted equally. It is a three-hour exam, and there are 4 sections, so it is a reasonable guideline to spend about 45 minutes on each section. All sections allow you to choose between two questions, answering one. Read instructions carefully at the head of each section.

PLEASE PUT ANSWERS TO EACH SECTION IN A SEPARATE EXAM BOOK.

WE REALLY MEAN IT. PLEASE PUT ANSWERS TO EACH SECTION IN A SEPARATE EXAM BOOK. THE REASON FOR THIS IS THEN WE CAN PARALLELIZE MARKING AMONG THE DIFFERENT LECTURERS AND YOU GET THE MARKS BACK SOONER.

Section 1: GIS and Genomics

Please select exactly **one question** and answer it. Please indicate clearly in your answer book which question you are answering.

- A. How does deforestation detection from space work, and is global near-real time monitoring of deforestation at high-pixel resolution feasible?
- B. You discover a population of armadillo in the Eastern Congo forest with a distinctive leopard-spotted pattern. The leopard-spotted armadillos are only found to the east of the Elila River, while those to the west are the normal tawny colour.

Be specific and provide justifications in your answers to each of the following questions:

- (i) What genetic, and comparative data would you collect to test whether the leopard-spotted armadillo was differentiated from the tawny armadillo populations? What analyses would you perform to test this? How would you determine whether the Elila River serves as a barrier to gene flow? If there has been gene flow, what would you expect to see? (40%)
- (ii) After performing these analyses you determine that the leopard-spotted armadillos are genetically structured from the tawny armadillos, however there is a tawny population with significant admixture from the leopard-spotted armadillos on the west bank of a shallow stretch of the river. Interestingly, despite the admixture, this population never seems to display the leopard-spotted phenotype. Through some additional analyses you learn that the leopard-spotted phenotype is controlled by the LSPT gene.

How would you test whether there has been selection at the LSPT locus in the leopard-spotted armadillos? What genomic signal would you expect to see if there has been selection? How would you test whether there has been selection for the LSPT locus in the admixed population? What would you expect to see here if so? (60 %)

Section 2: Statistics and Model Fitting

Please select exactly **one question** and answer it. Please indicate clearly in your answer book which question you are answering.

- A. Chris measured the abundance of one bacterial species – an *Epilithonimonas* – after 6 days in communities made up of different numbers of species: 1 (monoculture), 2, 3, 4. In order to investigate the effects of spatial structure on diversity, he grew each community in two treatments: shaken and unshaken. He hypothesized that the abundance of this species would be lower in more species-rich communities because of greater competition, but higher in unshaken cultures because they allow species to occupy distinct spatial niches.

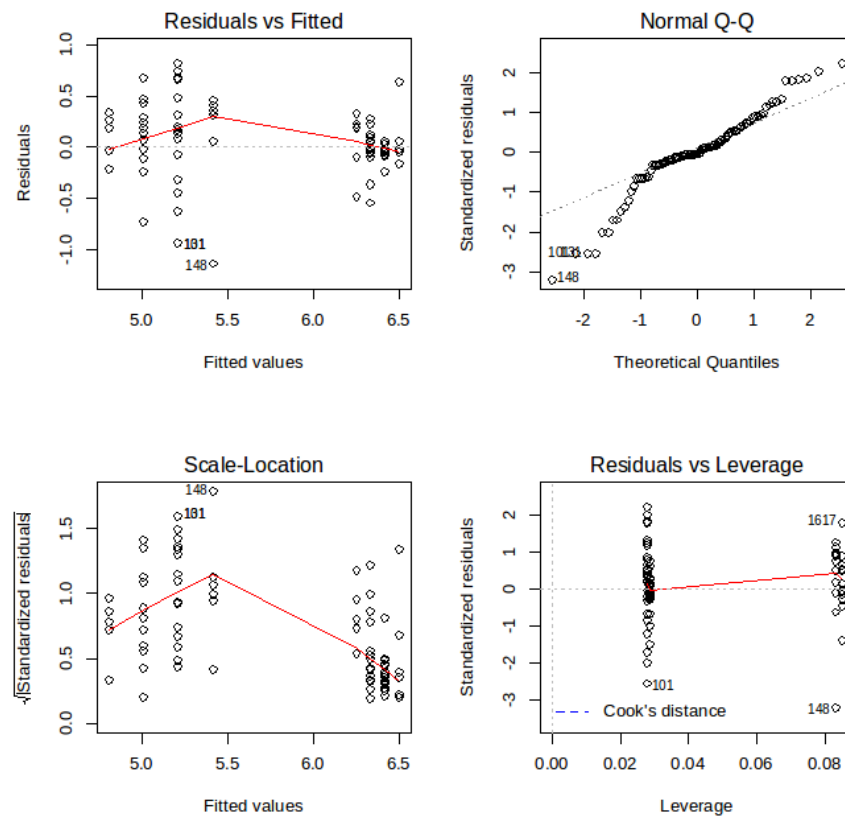
The following output shows the summary of the linear model of log abundance of *Epilithonimonas* against treatment (shaken, unshaken) and richness, and the diagnostic plots.

```
Call:
lm(formula = log.epilithonimonas ~ treatment * richness, data = Epilithonimonas,
    subset = time == 6)

Residuals:
    Min       1Q   Median       3Q      Max
-1.13513 -0.11008 -0.01354  0.19128  0.81248

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.60802    0.16366   28.156 < 2e-16 ***
treatmentunshaken  1.97433    0.23244    8.494 3.88e-13 ***
richness          0.20147    0.06186    3.257 0.00159 **
treatmentunshaken:richness -0.28397    0.08779   -3.235 0.00170 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3711 on 90 degrees of freedom
Multiple R-squared:  0.76, Adjusted R-squared:  0.752
F-statistic: 94.99 on 3 and 90 DF, p-value: < 2.2e-16
```



What do the results tell you? Were the hypotheses supported? Can you see any problems with the analysis? If so, what would you change?

- B.** Someone gave you some data on morphometrics in Odonates (Suborder Anisoptera = Dragonflies; Suborder Zygoptera = Damselflies). Having eagerly imported the data into R, you use `str(MyData)` to find:

```
'data.frame':    100 obs. of   16 variables:
 $ Suborder      : Factor w/ 2 levels "Anisoptera","Zygoptera": 1 1 1 1 1 1 1 1 1 1 ←
   ...
 $ BodyWeight     : num   0.159 0.228 0.312 0.218 0.207 0.22 0.344 0.128 0.392 0.029 ←
   ...
 $ TotalLength    : num   67.6 72 78.8 72.4 73 ...
 $ HeadLength     : num   6.83 6.84 6.27 6.62 4.92 6.48 7.53 5.74 8.05 5.28 ...
 $ ThoraxLength   : num   11.8 10.7 16.2 12.5 11.1 ...
 $ AbdomenLength  : num   48.9 54.4 56.3 53.3 57 ...
 $ ForewingLength : num   45.5 46 51.2 49.8 46.5 ...
 $ HindwingLength : num   45.4 45.5 49.5 48.8 46 ...
 $ ForewingArea   : num   370 411 461 469 382 ...
 $ HindwingArea   : num   484 517 574 591 481 ...
```

Now, you would like to build a statistical model to predict Odonate body weights from their body lengths. Figure 1 shows what relevant subsets of the data look like.

You then perform some analyses in R, and obtain the following output with `summary(LmDragonLin)`:

```
Call:
lm(formula = BodyWeight ~ TotalLength, data = subset(MyData,
  Suborder == "Anisoptera"))

Residuals:
```

Min	1Q	Median	3Q	Max
-0.098639	-0.022063	-0.003622	0.020156	0.098628

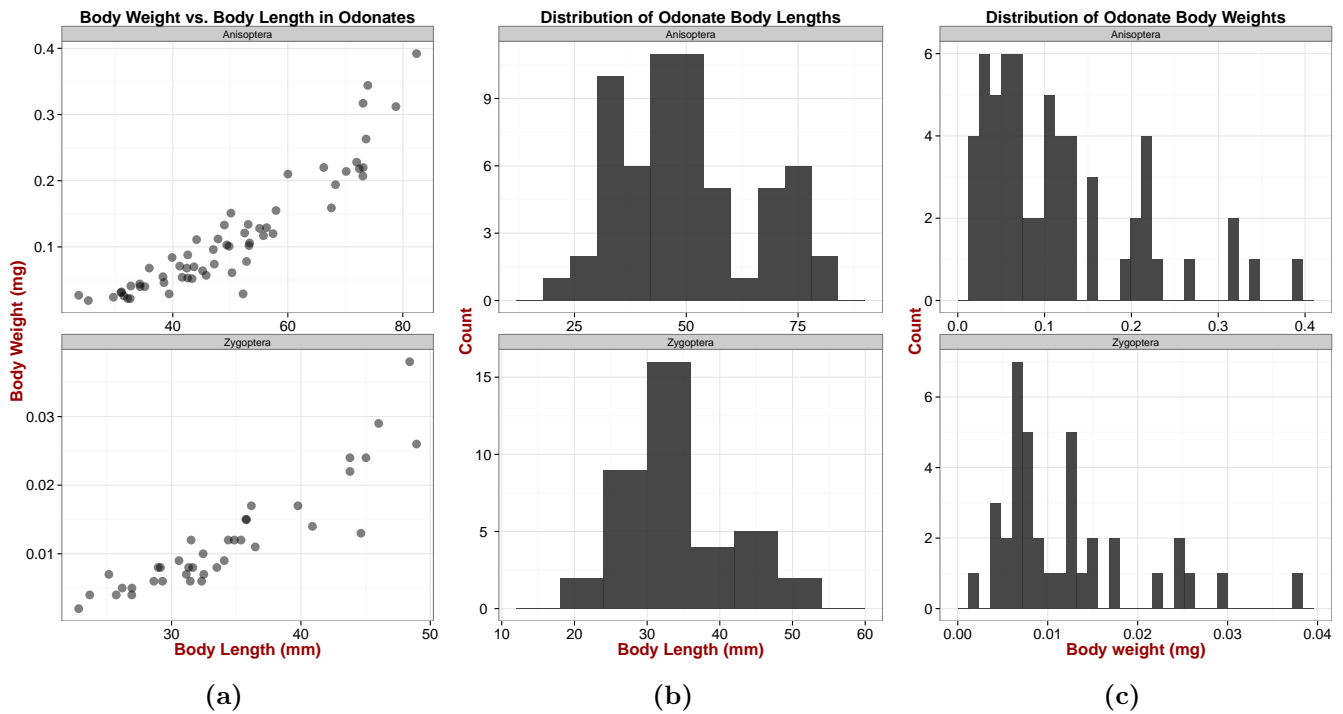


Figure 1: Body weight vs. length in Odonates: (a) Scatterplot; (b) Histograms of body lengths; and (c) Histograms of body weights.

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1592206  0.0160166  -9.941 3.86e-14 ***
TotalLength  0.0054933  0.0003081  17.829 < 2e-16 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1    1

Residual standard error: 0.03496 on 58 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared:  0.8457,    Adjusted R-squared:  0.843
F-statistic: 317.9 on 1 and 58 DF,  p-value: < 2.2e-16

```

Similarly, you get the following output with `summary(LmDamsellLin)`:

```

Call:
lm(formula = BodyWeight ~ TotalLength, data = subset(MyData,
  Suborder == "Zygoptera"))

Residuals:
    Min       1Q   Median       3Q      Max
-0.0097445 -0.0012688 -0.0001252  0.0014485  0.0113548

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.344e-02  2.777e-03  -8.443 4.65e-10 ***
TotalLength  1.035e-03  7.981e-05  12.964 3.96e-15 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1    1

Residual standard error: 0.003381 on 36 degrees of freedom
Multiple R-squared:  0.8236,    Adjusted R-squared:  0.8187
F-statistic: 168.1 on 1 and 36 DF,  p-value: 3.965e-15

```

Furthermore, `par(mfrow=c(2,2)); plot(LmDragonLin)` gives you the plots shown in Figure 2 for the dragonfly subset of the data.

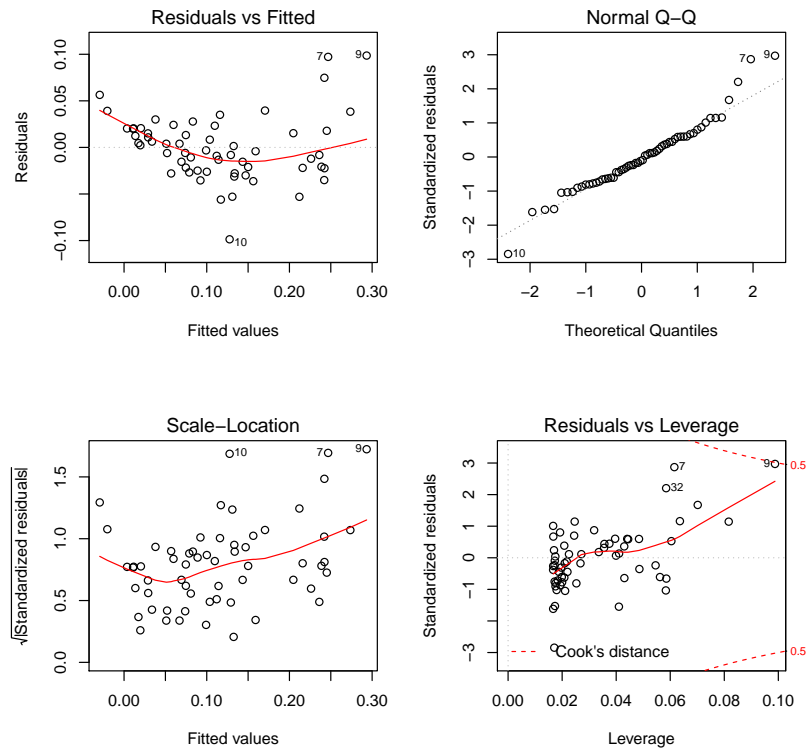


Figure 2: The result of `plot(LmDragonLin)`

Given this information, please answer the following (each question equally weighted):

- (i) Why might you expect a relationship between body weight and length (see Figure 1)? Describe the statistical analysis that was performed. Make sure you state what model was fitted, and write out the mathematical equation for the model, explaining what each of the parameters in the equation/model is. Also write out the representation of the model's equation in R syntax.
- (ii) State the assumptions that the data need to satisfy to allow you to fit the model to these data, and which of assumptions do and do not the data satisfy (how can you tell?). Thus, do you see any problems with using the model that was chosen with these data?
- (iii) Based upon the R outputs above, can you conclude that weight increases more rapidly with length in Dragonflies vs. Damselflies? Why or why not?
- (iv) What would be a more appropriate model than the one that was fitted above? Show it's equation and state what its parameters are.

Section 3: HPC & fractals

Please select exactly **one question** and answer it. Please indicate clearly in your answer book which question you are answering.

A. Answer the following questions. Please be brief in your answers to these questions – bullet points are OK as you are being marked on content:

- (i) What is an embarrassingly parallel problem? (10%)
- (ii) Give an example of a simulation task that is embarrassingly parallel. (10%)
- (iii) Give an example of a simulation task that is NOT embarrassingly parallel. (10%)
- (iv) What is a pseudo random number generator and a random seed? (20%)
- (v) The following question concerns an individual based simulation of organisms living in a community. An R function `stochastic_sim` accepts the community as an input and returns the community one generation later according to some unknown stochastic model. You expect that the species richness in your community will converge to a dynamic equilibrium if you repeatedly apply the function `stochastic_sim` and you wish to find out the mean value of the species richness at this equilibrium. Describe briefly how would you go about solving this problem. Please discuss random number generation, burn in and use of high performance computing in your answer. (50%)

B. Answer the following:

- (i) You have collected some fern leaves on a field trip and you wish to calculate their fractal dimensions. Describe briefly how you could do this using the box counting method. (40%)



- (ii) What is the fractal dimension of the object constructed in stages below (please explain your workings)? (30%)



- (iii) The following R code function generates a fractal along a line of a certain `width` centred on position `x` provided that the `threshold` is small.

```
linedraw <- function(x,width,threshold){
  if (width>threshold){
    linedraw(x-width/3,width/3,threshold)
    linedraw(x+width/3,width/3,threshold)
  } else {
    lines(c(x-width/2,x+width/2),c(0,0))
  }
}
```

Continues on next page

```
}  
}
```

- (a) What will happen if you call `linedraw(x=0,width=27,threshold=27)`? (10%)
- (b) Describe the fractal object that appears as you reduce the threshold, e.g. by calling `linedraw(x=0,width=27,threshold=27)`. You can do this with a series of drawings corresponding to decreasing thresholds. (20%)

Section 4: Maths I

Please select exactly **one question** and answer it. Please indicate clearly in your answer book which question you are answering.

A. Matrices:

Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

- (i) Find the eigenvalues and eigenvectors of A . Find matrices P and P^{-1} such that $D = P^{-1}AP$ is diagonal. Thereby compute the matrix e^{tA} . (50%)
- (ii) Compute the matrix e^{tA} again, but now using the Taylor series of the exponential. Express the answer in terms of $\cos t$ and $\sin t$. (50%)

B. Answer the following questions:

- (i) The following model is used in fisheries to describe the recruitment of fish as a function of the size of the parent stock. If we denote the number of recruits by R and the size of the parent stock by P , then

$$R(P) = \alpha P \exp(-\beta P), \text{ for } P \geq 0,$$

where α and β are positive constants.

Sketch the graph of the function $R(P)$ when $\alpha = 2$ and $\beta = 1$ and find all points on the curve that have a horizontal tangent. (20%)

- (ii) Calculate the derivative of the *inverse function* of the function $f(x) = \tan(x)$. (10%)
- (iii) Calculate the Taylor series of the arctangent function around the point $x = 0$ up to order 3. (20%)
- (iv) Spruce budworms are a major pest that defoliate balsam fir. They are preyed upon by birds. A model for the per capita predation rate is given by

$$f(N) = \frac{aN}{k^2 + N^2},$$

where N denotes the density of spruce budworm and a and k are positive constants.

For which density of spruce budworms is the per capita predation rate maximal? (50%)