NATURAL SCIENCES TRIPOS **Part IA**

Monday 3rd June 2019 9 am to 12 noon

MATHEMATICAL BIOLOGY

You must answer ***eight*** *questions.*

You must answer at least ***one question from each of Sections A to E****.*

You ***must not*** *answer all three questions from* ***Section E****.*

You must begin each answer on a ***separate*** *sheet.*

Attach a ***separate*** *cover sheet to* ***each*** *question.*

The question in Section B that is marked with an asterisk (*) requires knowledge of the last six lectures in the Michaelmas term.

Indicative proportions of marks are given for each question part.

It ***does not matter*** *whether you write on only one side of the paper or on both sides.*

STATIONERY REQUIREMENTS

Script Paper
Rough Work Pads
Blue Coversheets
Tags

SPECIAL REQUIREMENTS

Formulae Booklet
Approved Calculators Allowed

You may not start to read the questions printed on the subsequent pages of the question paper until you have been instructed that you may do so by the Invigilator.

SECTION A

A1

A diagnostic test has a probability 0.95 of giving a positive result when applied to a person suffering from disease A, and a probability 0.10 of giving a (false) positive when applied to someone not suffering from the disease. It is estimated that 0.5 % of the population suffer from the disease.

Suppose that the test is applied to a person about whom we have no relevant information relating to the disease (apart from the fact that they come from this population). Calculate the following probabilities:

(a) that the test result will be positive;

[~15% marks]

(b) that, given a positive result, the person is a sufferer;

[~15% marks]

(c) that, given a negative result, the person is a non-sufferer;

[~15% marks]

(d) that the person will be misclassified.

[~10% marks]

The manager of a laboratory is planning to buy a machine of either type P or type Q.

For each day's operation, the number of repairs, X, that P needs is a Poisson random variable with a mean value of 0.96. The daily cost of operating P is $C_P = 160 + 40X^2$.

For machine Q, let Y be a Poisson random variable with a mean value of 1.12 that indicates the number of daily repairs, such that the daily cost of operating Q is $C_Q = 128 + 40Y^2$.

(e) Assume that the repairs take negligible time and that each night both machines are cleaned so that they operate like a new machine at the start of each day. Which machine minimises the expected daily cost?

[~30% marks]

(f) If on average three samples are run by the chosen machine, find the probability that less than three samples are run on a given day.

[~15% marks]

A2

(a) Consider the following matrices:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 6 \\ -3 & 5 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} 4 & 0 \\ 2 & -1 \end{pmatrix}$$

For this pair of matrices,

(i) show that multiplication of ***A*** and ***B*** is non-commutative;

(ii) show that $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$.

[~20% marks]

(b) Determine the number of solutions to the following linear system:

$$\begin{aligned} x + 2y - 3z + 4u &= 2 \\ 2x + 5y - 2z + u &= 1 \\ 5x + 12y - 7z + 6u &= 7 \end{aligned}$$

[~30% marks]

(c) Find a complete set of eigenvalues and eigenvectors for the following matrix:

$$\boldsymbol{M} = \begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}$$

[~50% marks]

SECTION B

B3

A botanist is interested in investigating whether forests in the Paleotropics (Africa, Asia and Oceania) and the Neotropics (Americas) differ in the amount of carbon they store.

He collects data on Above Ground Biomass (Mg Dry Weight/ha) from ten plots in each region. For the Paleotropics, he obtains a mean value, $\bar{x}_1$, of 401.6, and for the Neotropics a mean value, $\bar{x}_2$, of 242.3. The respective Sums of Squares are SS_1 = 298,698.4 and SS_2 = 178,652.1.

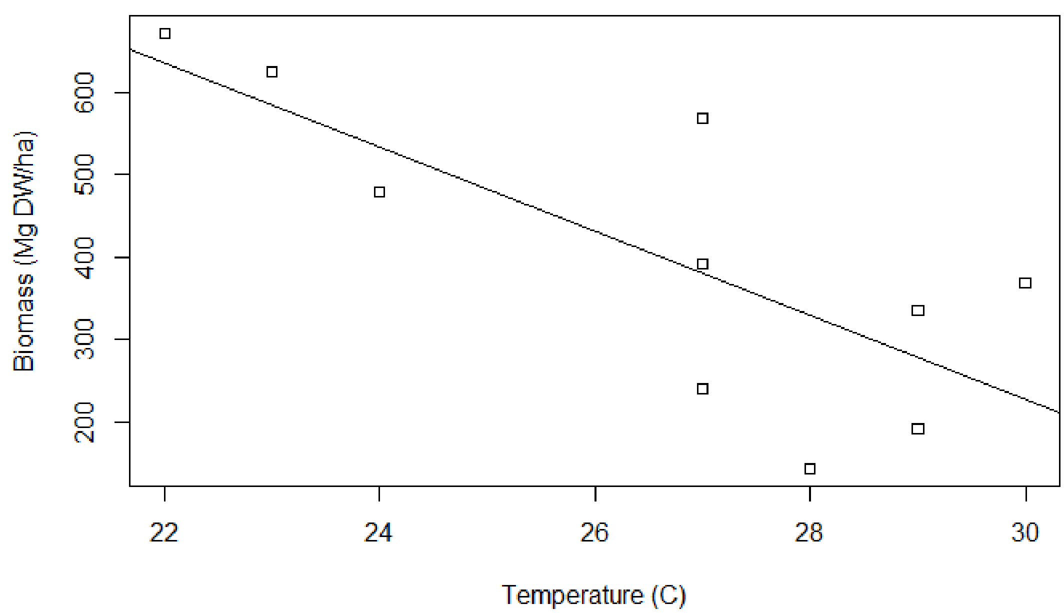
(a) Do forests in the two regions differ in their ability to store carbon?

[~20% marks]

(b) Provide a sentence that you could use to report your findings for part ***(a)*** in a scientific paper.

[~15% marks]

Our botanist also wants to test whether carbon storage potential in a forest might be related to Mean Annual Temperature (°C). For the ten Paleotropic forests, he plots Above Ground Biomass against Mean Annual Temperature.



The line of best fit is 1757.9 – 51.0 * Temperature

(c) What is the predicted Biomass at 24 degrees Celsius and at 28 degrees Celsius?

[~10% marks]

B3 (CONTINUED)

He then tests the relationship between the two variables and gets:

```
                Analysis of Variance Table

      Response: Biomass
                  Df      SS   MS   F
      Temperature  1 172625   ??  ??
      Residuals   ?? 126074   ??
      Total        9      ??
```

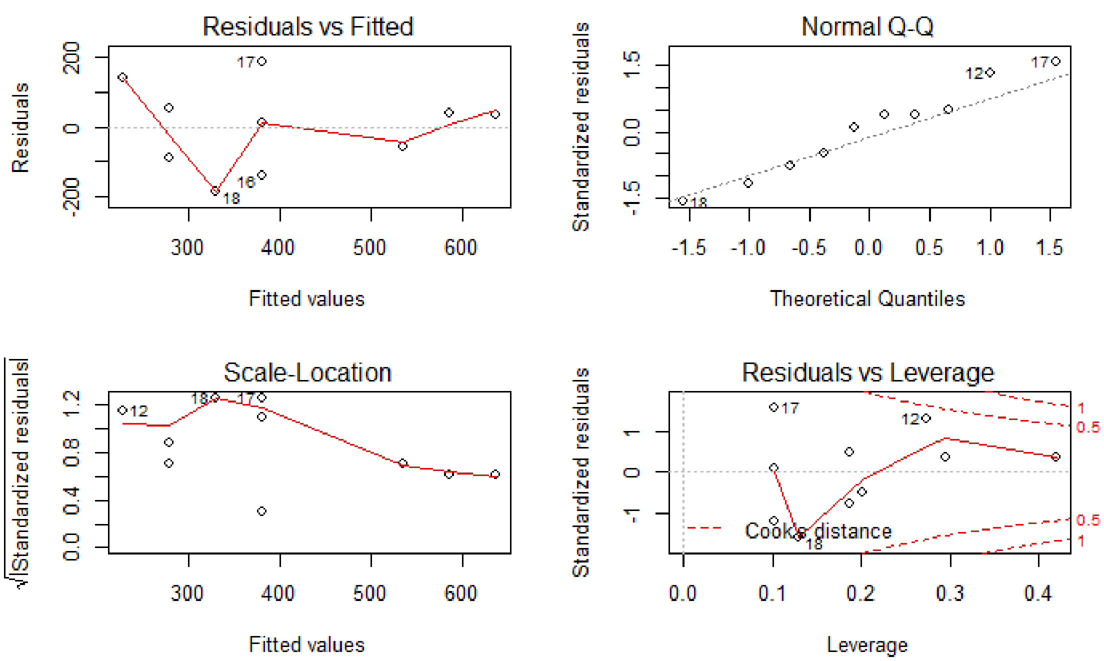
(d) Complete the table above. What can you conclude?

[~20% marks]

(e) Provide a sentence that you could use to report your findings for part ***(d)*** in a scientific paper.

[~15% marks]

He obtains the following diagnostic plots.

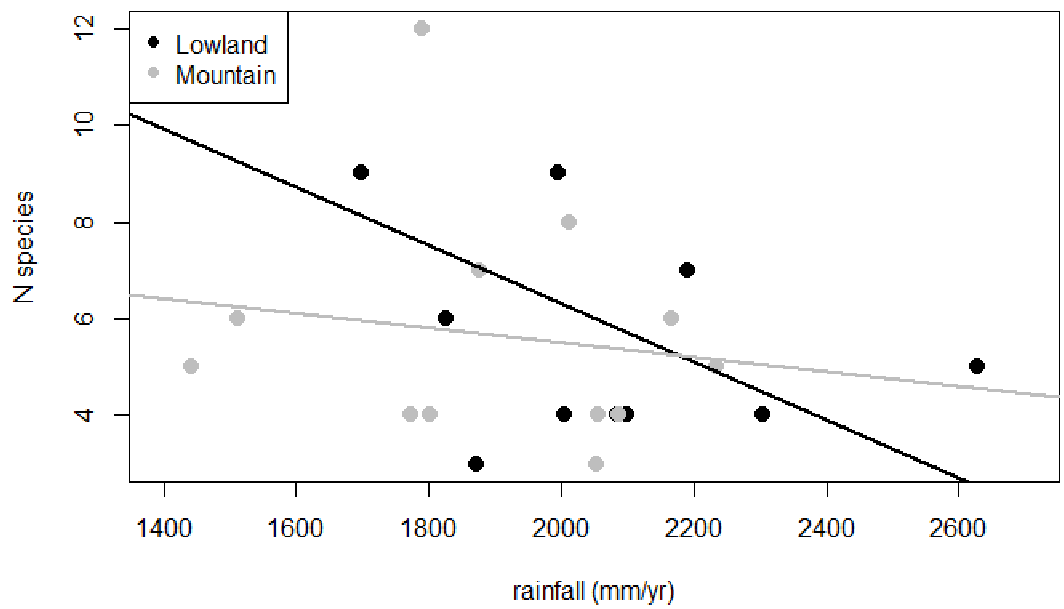


(f) Comment on the validity of this analysis.

[~20% marks]

B4 *

An ecologist wants to investigate whether the biodiversity (number of tree species) in mountain rain forests is different from that in than lowland forests. She also hypothesises that rainfall (in mm/yr) might affect the number of species. She counts the number of species in 12 mountain and 12 lowland plots, and gets:



She fits a linear model with an interaction between forest type and rainfall and obtains the following table:

```
Analysis of Variance Table

Model 1: N.species ~ (Type + Precipitation)
Model 2: N.species ~ (Type + Precipitation + Type:Precipitation)
  Res.Df    RSS   Df Sum of Sq      F
1     21 141.03
2     ?? 132.39    1        ??     ??
```

(a) Complete the table above. What can you conclude? ***[~20% marks]***

After dropping the interaction, she tests the main effects:

```
Analysis of Variance Table

Model 1: N.species ~ Precipitation
Model 2: N.species ~ (Type + Precipitation)
  Res.Df    RSS   Df Sum of Sq      F
1     22     ??
2     21 141.03   ??      5.00     ??
```

B4 * (CONTINUED)

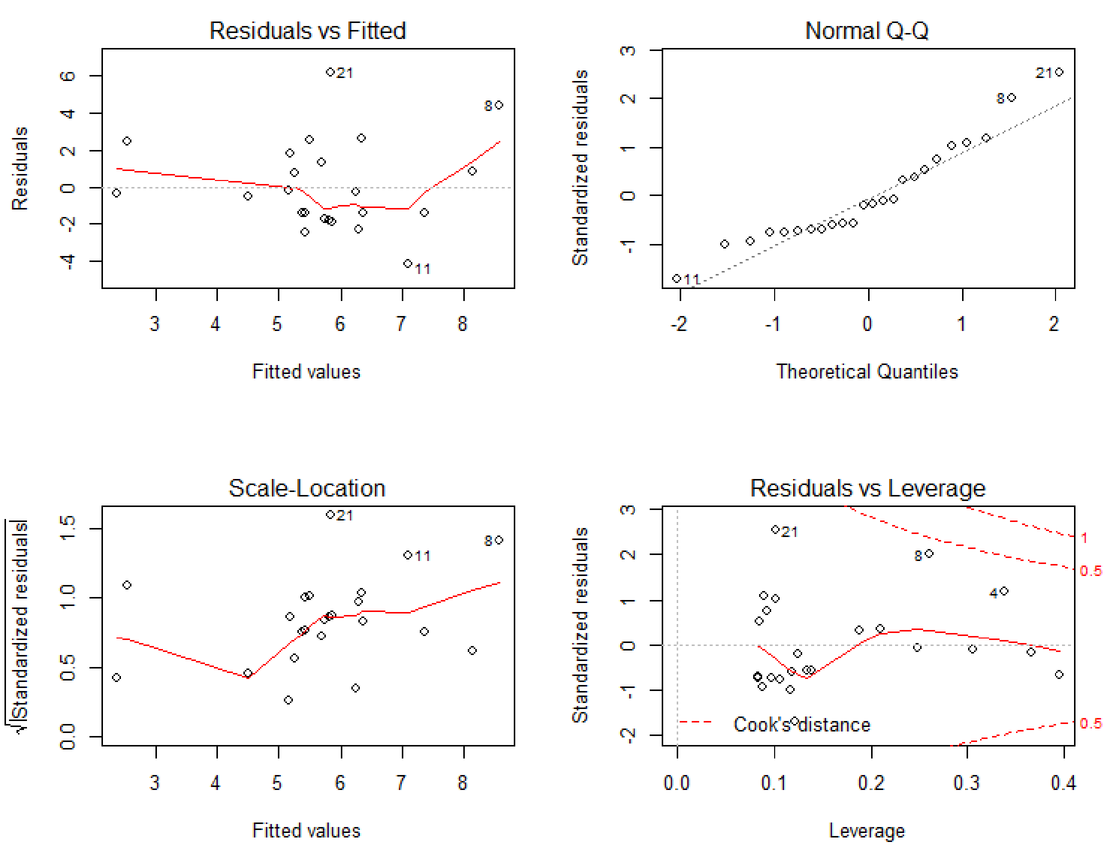
```
Analysis of Variance Table
Model 1: N.species ~ Type
Model 2: N.species ~ (Type + Precipitation)
  Res.Df    RSS   Df Sum of Sq      F
1      22 176.33
2      ??     ??   ??     35.30    ??
```

(b) Complete the two tables above. What can you conclude? ***[~20% marks]***

(c) Was she justified in taking this step? What could she conclude from her analysis? ***[~10% marks]***

(d) Provide a sentence that you could use to report your findings for parts ***(a)*** to ***(c)*** in a scientific paper. ***[~10% marks]***

She obtains the following diagnostic plots:



(e) What should she conclude? ***[~20% marks]***

She decides to reanalyse the data by fitting a Generalised Linear Model with quasiPoisson error structure:

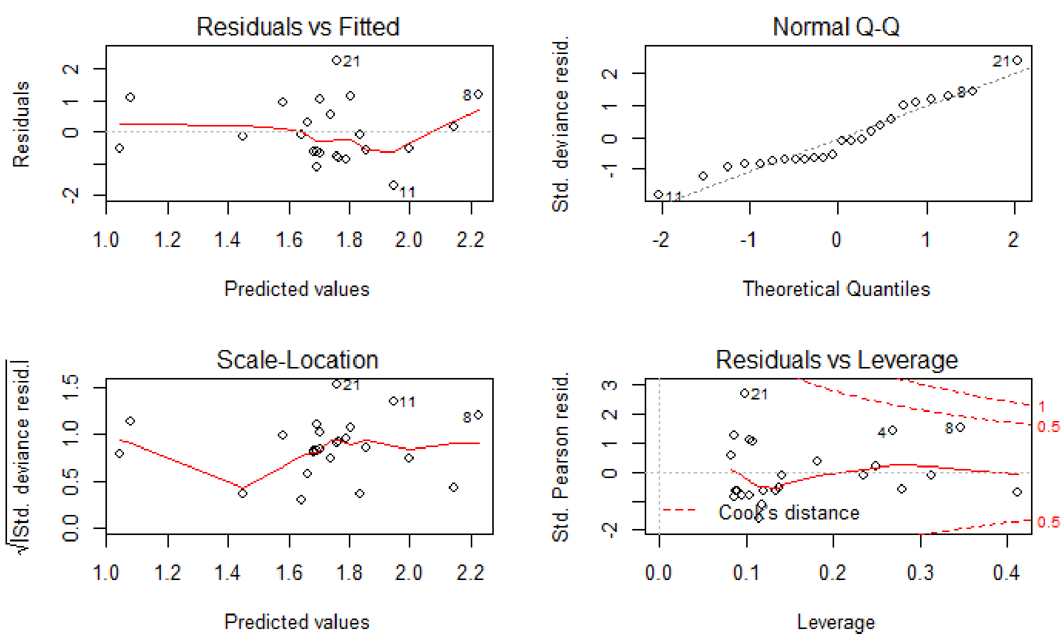
(TURN OVER)

B4 * (CONTINUED)

```
Call:

glm(formula = N.species ~ (Type + Precipitation + Type:Precipitation),
family = "quasipoisson",
    data = biodiv)
Deviance Residuals:
    Min       1Q    Median       3Q      Max
-1.7063  -0.6485  -0.3256   0.6254   2.2355

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 4.0871022  0.8771750   4.659 0.000151 ***
TypeMountain               -1.8518778  1.3133086  -1.410 0.173877
Precipitation              -0.0011447  0.0004384  -2.611 0.016718 *
TypeMountain:Precipitation  0.0008802  0.0006758   1.303 0.207543
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.047568)

Null deviance: 27.869  on 23  degrees of freedom
Residual deviance: 19.804  on 20  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

(f) What could be the rationale for preferring this approach to a linear model? Is the quasiPoisson error structure justified, or would a Poisson model be sufficient in this case? ***[~10% marks]***

She obtains the following diagnostic plots from the GLM:



(g) What can she conclude? ***[~10% marks]***

SECTION C

C5

A model for the dynamics of a population of plants is

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\beta N}{1 + (N/\alpha)^2},$$

in which N is the population size at time t, and α and β are positive parameters.

(a) Sketch the relative growth rate as a function of N.

[~10% marks]

(b) Use your answer to ***(a)*** to outline briefly the biology captured by the model, explaining the meanings of the parameters α and β.

[~10% marks]

(c) At which population size would the number of plants increase most quickly?

[~15% marks]

(d) Sketch a graph of the absolute growth rate as a function of the population size.

[~10% marks]

The model above does not account for herbivory by animals. An updated model is proposed,

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\beta N}{1 + (N/\alpha)^2} - \rho \mathrm{N},$$

in which ρ is an additional positive parameter.

(e) Explain this updated model, paying particular attention to the form of the function used to represent herbivory and the meaning of the parameter ρ.

[~5% marks]

For the rest of this question, consider cases in which $\beta > \rho$ and $\beta \leq \rho$ separately.

(f) Find the equilibrium value(values) of the model, and assess its(their) stability.

[~25% marks]

(g) Sketch the solution of the model (i.e. the value of N as a function of t) for a range of initial conditions.

[~15% marks]

(h) Explain how the predictions of the model are consistent with the biology that the model attempts to capture.

[~10% marks]

(TURN OVER)

C6

For the differential equation model

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = F(Y),$$

(a) give a condition which ensures that $Y = \bar{Y}$ is an equilibrium value of the model;

[~5% marks]

(b) state and prove a result relating the stability of the equilibrium at $Y = \bar{Y}$ to the numeric value of $\mathrm{d}F/\mathrm{d}Y$ evaluated at that point.

[~25% marks]

A model for the amount of wheat tissue infected by a fungal pathogen is

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\beta N}{4}\left(1-\left(\frac{N}{\kappa}\right)^4\right),$$

in which β and κ are positive constants.

(c) What are the equilibria of the wheat tissue model?

[~5% marks]

(d) Use the method developed in part ***(b)*** to assess the stability of the equilibria.

[~15% marks]

You are reminded of the general result that a Bernoulli differential equation of the form

$$\frac{\mathrm{d}N}{\mathrm{d}t} + bN = cN^a$$

can be transformed into a linear differential equation using the substitution $Z = N^{1-a}$.

(e) Show that the wheat tissue model can be transformed into the form

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta\left(\frac{1}{\kappa^4} - Z\right).$$

[~25% marks]

(f) Hence, solve the model to find the amount of wheat tissue that is infected at time t, given that the initial amount of infected tissue is $N = N_0$.

[~25% marks]

SECTION D

D7

Consider the system represented by the following pair of simultaneous first-order non-linear differential equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x^2 - xy - 5x$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^2 + x^2 - 2x - 9$$

(a) Find and classify the equilibrium points.

[~50% marks]

(b) Sketch the null clines and then the phase plane for this system, showing the behaviour around the equilibrium points.

[~35% marks]

(c) For the initial condition $x(0) = 5$ and $y(0) = 1$, mark clearly on the phase plane the trajectory associated with this initial condition, and sketch the corresponding graphs for $x(t)$ and $y(t)$.

[~15% marks]

(TURN OVER)

D8

The following system of differential equations describes the spread of a deadly virus in a population of seals, where $S(t)$ is the density of susceptible animals and $I(t)$ the density of infected animals:

$$\frac{dS}{dt} = rS\left(1 - \frac{S}{k}\right) - \beta SI$$

$$\frac{dI}{dt} = \beta SI - \alpha I$$

You can assume that there is no recovery from infection with this lethal virus.

(a) Give an interpretation for each of the four parameters in the model, namely r, k, α and β.

[~25% marks]

(b) What additional assumption was made with respect to the reproduction of infected seals?

[~5% marks]

(c) Find the equations of the null clines.

[~20% marks]

(d) Calculate the stationary points of the model.

[~20% marks]

(e) It can be shown that the disease will persist in the population if its basic reproductive ratio R_0 is greater than one. Use this information to propose an expression for R_0 using the model parameters.

[~5% marks]

(f) Determine the stability of the relevant stationary points when $R_0 < 1$ and sketch the corresponding phase plot with the direction field (with S on the horizontal axis and I on the vertical axis).

[~25% marks]

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

SECTION E

E9

(a) The following dynamic programming sequence alignment matrix was completed using scores from an amino acid substitution matrix; a fixed penalty was used for gaps.

	-	**S**	**A**	**W**	**V**	**C**	**K**	**N**	**H**	**V**	**E**
-	0	0	0	0	0	0	0	0	0	0	0
T	0	1	0	0	0	0	0	0	0	0	0
W	0	0	0	11	4	0	0	0	0	0	0
A	0	1	4	4	11	4	0	0	0	0	0
T	0	1	1	2	4	10	3	0	0	0	0
C	0	0	1	0	1	13	7	0	0	0	0
R	0	0	0	0	0	6	15	8	1	0	0
N	0	1	0	0	0	0	8	21	14	7	0
G	0	0	1	0	0	0	1	14	19	12	5

(i) Was this matrix completed for finding a global or a local alignment? Explain how you can tell.

[~5% marks]

(ii) What is the gap penalty?

[~5% marks]

(iii) Show the highest-scoring local alignment between the two sequences and the score at each position in the alignment. If more than one equally good alignments exist, show them all.

[~10% marks]

Hint: Follow the format of the following example to illustrate an alignment and the score at each position:

```
         P  P  E  G  R  H
         |  |           |
         P  P  D  -  K  H
scores:  7 14 16  9 11 19
```

(TURN OVER)

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E9 (continued)

(b)

(i) Copy and complete the dynamic programming sequence alignment matrix given below for finding an optimal global alignment of the sequences GCAT and GACA.

Use the following scoring scheme: nucleotide match = +3, nucleotide mismatch = −1, gap penalty = −1. Use arrows to show the potential trace back options for all 25 cells.

	-	**G**	**C**	**A**	**T**
-	**0**				
G					
A					
C					
A					

[~25% marks]

(ii) What is the score of the optimal global alignment?

[~5% marks]

(iii) Mark (e.g. with circles) the optimal global alignment path on the matrix.

[~5% marks]

(iv) Show the optimal global alignment between the two sequences and the score at each position in the alignment.

[~10% marks]

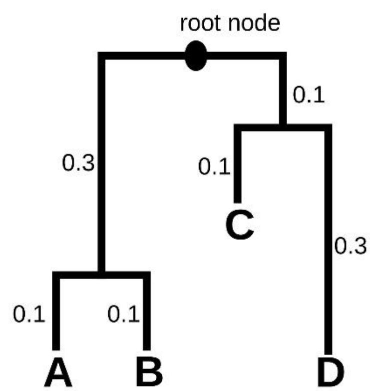
Hint: Follow the format of the following example to illustrate an alignment and the score at each position:

```
         C G G U A G C
         |   |   |   |
         C A G - A U C
scores:  3 2 5 4 7 6 9
```

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E9 (continued)

(c) In the rooted phylogenetic tree below, vertical branch lengths (with numbers) correspond to evolutionary distance. The root represents an ancestral sequence, whereas A, B, C and D represent four present-day sequences.



(i) Does this tree have a constant molecular clock? How can you tell?

[~5% marks]

(ii) Calculate the values of ***w***, ***x***, ***y*** and ***z*** in the distance matrix below, in which the distance measure is evolutionary distance.

	A	B	C	D
A	**0**			
B	***w***	**0**		
C	***x***	**0.6**	**0**	
D	***y***	***z***	**0.4**	**0**

[~5% marks]

(iii) The UPGMA algorithm for constructing a phylogenetic tree assumes that the tree has a constant molecular clock. However, it is often still possible to apply the UPGMA algorithm even when this condition is not satisfied, although the resulting tree may be different from the true evolutionary tree. Apply the UPGMA algorithm to the distance matrix from part ***(ii)***. Show the tree produced by the UPGMA algorithm and annotate the branch lengths according to the algorithm. Note that the resulting tree branch lengths may disagree with the distance matrix from part ***(ii)***.

[~20% marks]

(TURN OVER)

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E10

After consuming an alcoholic drink, ethanol leaves the stomach and enters the bloodstream, causing [S], the concentration of ethanol in the stomach to decrease via a first-order process, according to the rate law $-k_S[S]$, where k_S is a positive rate constant. The concentration of ethanol in the bloodstream increases as alcohol leaves the stomach, but decreases in a zero-order process governed by the positive rate constant k_L, due to the conversion of ethanol to acetaldehyde by the enzyme alcohol dehydrogenase in the liver.

(a) If [B] and [A] represent the concentrations of ethanol and acetaldehyde, respectively, in the bloodstream, write down three differential equations that describe the rate of change of [S], [B] and [A] as a function of time, *t*.

[~10% marks]

(b) If the initial values of [S], [B] and [A] are $[S]_0$, 0 and 0, respectively, determine exact solutions for the time dependence of the concentration of ethanol in the stomach, ethanol in the bloodstream and acetaldehyde in the bloodstream.

[~40% marks]

(c) On the same axis system, draw rough sketches of [S], [B] and [A] as a function of time. Explain why the predicted values of [B] and [A] may become unrealistic as the value of *t* increases. What unstated assumption has been made?

[~25% marks]

(d) If $[S]_0 = 1\ \text{mol L}^{-1}$, $k_S = 3 \times 10^{-3}\ \text{s}^{-1}$ and $k_L = 4 \times 10^{-5}\ \text{mol L}^{-1}\ \text{s}^{-1}$, determine the maximum value of [B] and the time (rounded to the nearest minute) at which this will occur. Use an approximation to work out the time (rounded to the nearest hour) at which [B] will have dropped to 5 % of its maximum value.

[~25% marks]

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E11

Two wolves have captured a single prey together, and each must simultaneously decide whether or not to fight for it. If one wolf fights while the other does not, the former easily obtains the prey, gaining v units of energy (where $v > 0$), while the latter gains nothing. If both wolves fight, an escalated battle results; both wolves expend c units of energy (where $c > 0$), and one (chosen at random) wins the prey, gaining v units of energy, while the other loses and gains nothing. If neither wolf fights, they divide the prey equally (and peacefully), each obtaining $v/2$ units of energy.

The payoff matrix below gives the expected net energetic gain for a wolf, depending on its own action (listed down the left-hand side of the matrix), and that of its rival (listed along the top of the matrix).

	Fight	Don't fight
Fight	$(v/2) - c$	v
Don't fight	0	$v/2$

(a) Under what conditions is each of the two pure strategies ('Fight' and 'Do not fight') in this game evolutionarily stable?

[~20% marks]

(b) Under what conditions does the game yield an evolutionarily stable mixed strategy? Derive an expression for the evolutionarily stable probability of choosing to fight under these conditions.

[~10% marks]

(c) Derive an expression for an individual's expected payoff in a population that adopts an evolutionarily stable mixed strategy. For what value of v is this maximised? Explain in words why payoffs may sometimes be lower when prey are more valuable (i.e. when v is greater).

[~15% marks]

(TURN OVER)

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E11 (continued)

(d) Suppose that one wolf is stronger than the other, and that if both choose to fight, the stronger individual will win with some probability $p > 0.5$ (while the smaller will win with probability $1 - p$). Write out a new payoff matrix for the game, and determine the conditions under which each of the strategies 'Fight if stronger, but not if weaker' and 'Fight if weaker, but not if stronger' are evolutionarily stable. What does the model suggest is the more likely outcome of evolution?

[~40% marks]

(e) Suggest three ways in which you might modify the assumptions of the model to create a more realistic analysis of wolf fighting behaviour.

[~15% marks]

(END OF PAPER)