**NATURAL SCIENCES TRIPOS** **Part IA** **MB/1**

# MATHEMATICAL BIOLOGY

*You have **three hours** to answer this paper (plus any pre-agreed individual adjustments), following which you have a 30-minute window to upload your answer files.*

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

*Question B4 requires knowledge of the last six lectures in Michaelmas term.*

*All questions carry equal weight. Indicative proportions of marks are given for each question part.*

*You may use an approved calculator and the Mathematical Biology Formula Booklet.*

*Begin each answer on a **separate** page.*

*Answers should be **hand-written,** either electronically on a tablet or on paper and scanned in. Make sure that your handwriting and any diagrams or equations are clear and legible in the submitted pdf files.*

*Where relevant, formulae, equations, diagrams, graphs and sketches should be drawn by hand, not prepared using a computer. Do not copy and paste any figures from other documents.*

*For full credit, calculation steps must be shown where necessary to explain your answers.*

***After you have completed the assessment***

1. *Your answers should be submitted as **separate pdf files**, one for each question attempted.*
2. *Name each file according to the question attempted. For example answers to questions A1 and D7 should have filenames A1.pdf and D7.pdf respectively. Your Blind Grade Number will be recorded automatically by Moodle.*
3. *Follow the guidance on the University Moodle site on how to upload your answer files.*

**A1**

(a) A study reports that the range of body temperatures, measured in degrees Celsius, for individuals infected with disease X is Normally distributed with mean $\mu = 38.5$ and standard deviation $\sigma = 0.65$. What is the probability that a randomly selected infected individual has a temperature between 37.9°C and 39.8°C?

[~20% marks]

(b) A border control agency decides to randomly screen 5 passengers for disease X from each incoming flight.

(i) Using the notation that a particular flight has $N$ passengers of whom $I$ are infected, write down an expression for the probability that the group selected from one flight contains exactly $i$ infected passengers. Include any conditions on your parameters.

(ii) Of two incoming flights, flight F1 has 50 passengers of whom 3 are infected and flight F2 has 80 passengers of whom 6 are infected. Is it more likely that at least one infected passenger will be selected from flight F1 or flight F2? (probabilities must be shown for credit).

[~35% marks]

(c) Any passenger with a temperature 38°C or higher is recorded as potentially infected with disease X. 78% of infected individuals with disease X will have a temperature of at least 38°C. However, 5% of all arriving passengers who do not have the disease will also have a temperature of at least 38°C. Border control keeps a list of the prevalence of the disease in each country. A passenger arriving from Country Z, which has a reported prevalence of disease X of 1.5%, has a temperature of 38.7°C.

(i) What is the probability, given the information available to border control, that the passenger is infected with disease X?

(ii) The passenger is then asked to provide a swab for a lateral flow test. Suppose the lateral flow test has a specificity of 99% and sensitivity of 70%. The passenger tests positive. What is the probability, given all the information available to border control, that the passenger in infected with disease X?

(Recall that specificity is the proportion of people without the disease who have a negative test, and sensitivity is the proportion of people with the disease who have a positive test.)

[~45% marks]

**A2**

(a) Consider the linear geometric transformation of points in the two-dimensional plane wherein points are reflected in the line $y = mx$, with $m$ a scalar constant.

By considering how it transforms the standard basis, construct a 2 x 2 matrix representing this transformation.

[~30% marks]

(b) Explain, by sketching an example or otherwise, why reflection in a line $y = mx + c$, where $c$ is nonzero, is **not** a linear transformation of points in the two-dimensional plane.

[~20% marks]

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

$$M = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{pmatrix}$$

[~50% marks]

**B3**

(a) A population ecologist wants to investigate the factors that favour the extinction of lizard populations. She uses a long-term monitoring dataset, and quantifies how many populations have gone extinct and how many are still extant. She also classifies populations as being viviparous (giving birth to live young) or oviparous (laying eggs). She obtains the following data:

| | Extinct | Surviving |
|---|---|---|
| Viviparous | 16 | 74 |
| Oviparous | 5 | 81 |

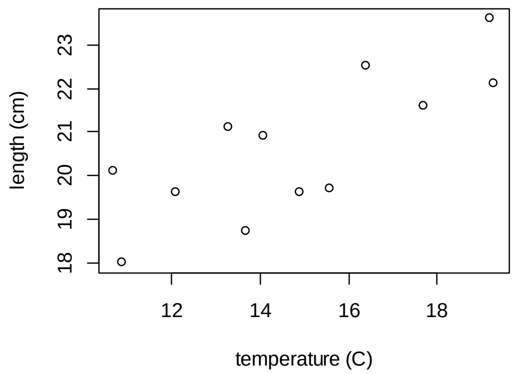
Carry out a test to assess whether reproductive mode affects the probability of extinction of a population.

[~15% marks]

(b) Provide a sentence that you could use to report these results in a scientific paper.

[~10% marks]

(c) The population ecologist is also interested in whether local temperature is linked to size of lizards at that location. She collects lizards from 12 locations, and then plots length (cm) against the mean summary temperature (degrees Celsius):




She runs a linear regression to test whether length is predicted by temperature. Complete the following ANOVA table, and test whether the relationship is significant:

| | SS | df | MS | F |
|---|---|---|---|---|
| Model | 17.86 | | | |
| Residual | | | | |
| Total | 29.62 | | | |

[~30% marks]

(d) Provide a sentence to report your results in a paper.

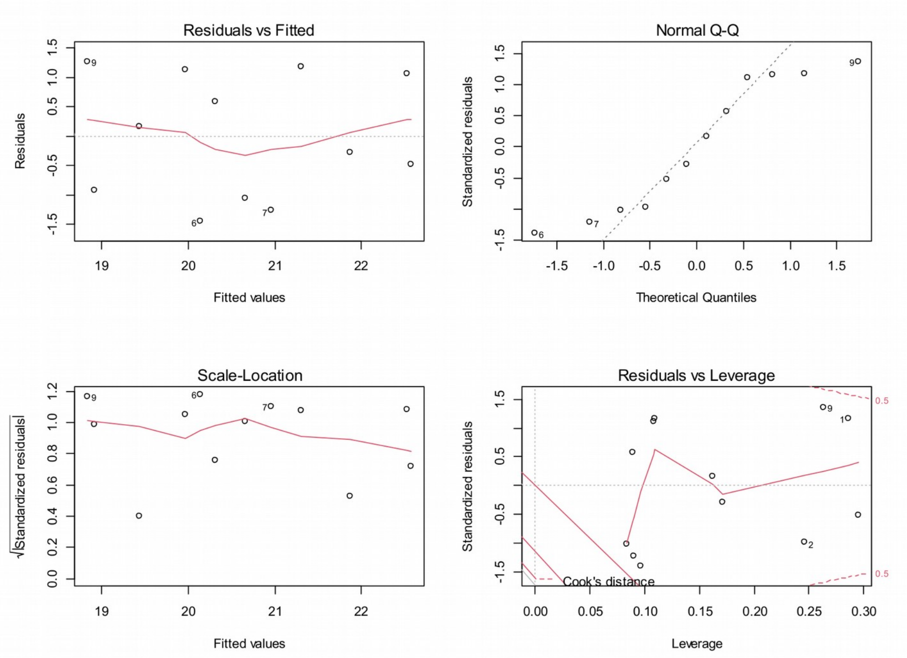
[~10% marks]

**QUESTION CONTINUES ON THE NEXT PAGE**

(e) One lizard is collected from a location with a mean temperature of 15.2 degrees, and another from a location with mean temperature 21.5 degrees. Given that the above analysis gives a line of best fit with an intercept of 14.18 and a slope of 0.43, what would be the predicted lengths of these lizards? Which prediction is likely to be more accurate?

[~15% marks]

(f) Comment on the following diagnostic plots:

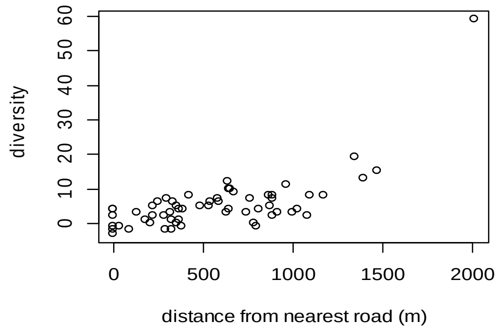



[~20% marks]

**B4**

*This question requires knowledge of the last six lectures in Michaelmas term.*

(a) A conservation biologist wants to test whether distance from a road affects bird biodiversity found at given spot. He runs 60 surveys, and plots bird diversity (number of species observed during a survey) versus distance from the nearest road (m).




He tests whether distance predicts diversity, and gets the following output:

```
Call:
lm(formula = count ~ distance, data = birds)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.430155   1.351134  -1.799   0.0773 .
distance     0.013409   0.001893   7.085 2.11e-09 ***
```

Given that the $SS_{Residual}$ is 2192 and $SS_{Total}$ is 4089, test whether there is a significant effect of distance.

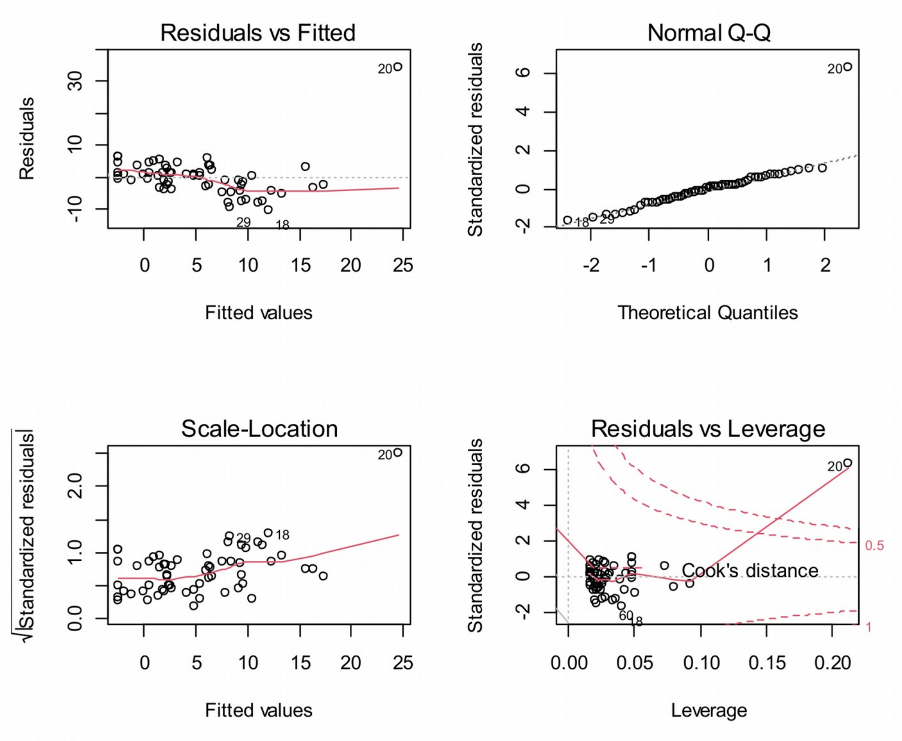
[~30% marks]

(b) Provide a sentence that you could use to report the results in a scientific paper.

[~10% marks]

**QUESTION CONTINUES ON THE NEXT PAGE**

(c) The diagnostic plots associated with this analysis are shown below:




What can you conclude from these plots?

[~10% marks]

(d) He then attempts to fit a General Linear Model with a Poisson Error structure, and gets the following output:

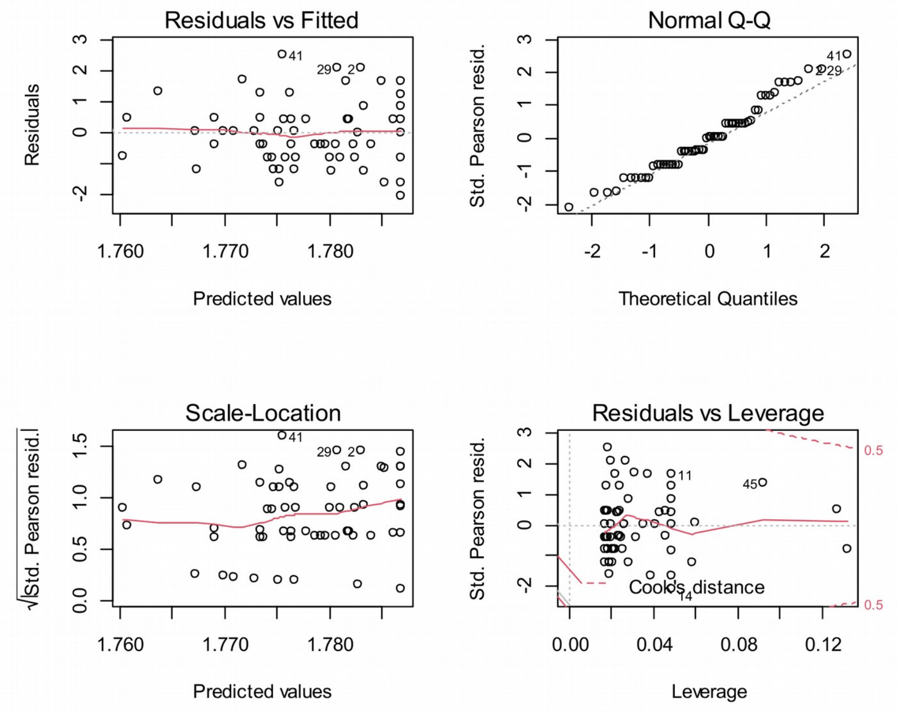
```
Call:
glm(formula = count ~ distance, family = "poisson", data = birds)

Deviance Residuals:
    Min        1Q    Median        3Q       Max
-2.5233  -0.8336   0.0171   0.4465   2.1981

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.787e+00  8.992e-02  19.870   <2e-16 ***
distance    -1.726e-05  1.393e-04  -0.124    0.901
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 64.936  on ??  degrees of freedom
Residual deviance: 64.921  on ??  degrees of freedom
```

Compute the df, and then test whether the relationship is significant.

[~20% marks]

**QUESTION CONTINUES ON THE NEXT PAGE**

(e) The diagnostic plots are shown below:




How do these plots compare to those for the linear model in (c)?

[~10% marks]

(f) Use the relationships estimated in (a) and (d) to predict bird diversity at 1900 meters from the nearest road. Can you explain the discrepancy between these predictions?

[~20% marks]

**C5**

A simple model of coronavirus disease does not account for any reduction in the number of susceptible individuals as the epidemic progresses. It tracks only the number of currently infected individuals $(I)$ as a function of time $(t)$

$$\frac{dI}{dt} = \beta NI - \mu I .$$

The positive parameter $\beta$ sets the rate of new infections, $N$ is the total size of the susceptible population (which is assumed to remain constant over the timescale of the model), and $\mu$ is a further positive constant.

(a) Suggest a plausible interpretation of the meaning of the parameter $\mu$.

[~5% marks]

(b) Solve the model, assuming $I_0$ infections at time $t=0$.

[~5% marks]

This type of model can be used to understand selection for new variants of SARS-CoV-2. The following pair of uncoupled equations track numbers currently infected by each of two variants

$$\frac{dI_1}{dt} = \alpha_1 N I_1 - \mu I_1 ,$$

$$\frac{dI_2}{dt} = \alpha_2 N I_2 - \mu I_2 .$$

The positive parameter $\alpha_i$ sets the rate of new infections by variant $i$. The number of cases currently infected by variant $i$ at $t=0$ is $I_{0i}$. We assume that variant 2 only started to spread relatively recently and so variant 1 is initially much more common than variant 2, i.e. $I_{01} \gg I_{02}$.

The proportion of cases caused by variant 2 is

$$\rho(t) = \frac{I_2(t)}{I_1(t) + I_2(t)} .$$

(c) By differentiating $\rho(t)$ with respect to *t*, show that, if $\gamma = (\alpha_2 - \alpha_1) N$, then

$$\frac{d\rho}{dt} = \gamma \rho (1 - \rho) .$$

[~20% marks]

**QUESTION CONTINUES ON THE NEXT PAGE**

(d) Find the equilibrium values of $\rho$, and assess the stability of each. You will need to consider separately the following cases:

Case 1: $\alpha_2 > \alpha_1$

Case 2: $\alpha_2 < \alpha_1$

Case 3: $\alpha_2 = \alpha_1$

Interpret the results in the context of the model of two variants of SARS-CoV-2.

[~20% marks]

(e) Find an expression for $\rho(t)$, given the initial proportion of current cases caused by variant 2 is $\rho_0$ at $t=0$. Sketch $\rho(t)$ as a function of $t$.

[~20% marks]

(f) Verify that the solution you found in part (e) is consistent with that obtained by using your solution to part (b) to find $I_1(t)$ and $I_2(t)$ independently and substituting these expressions into the definition of $\rho(t)$.

[~20% marks]

(g) Outline how results from this type of model could, in principle, be used to interpret surveillance data in order to obtain an estimate of the relative infectivity of a new variant of SARS-CoV-2.

[~10% marks]

**C6**

Certain species are valued by "hobby collectors". Specimens taken from rare species are often particularly prized and so demand a high market price. This can cause hunters to put more effort into capturing individuals from a given species when it is rarer.

The model of a population of great auk (*Pinguinus impennis*) tracks the density of individuals $N$ at time $t$, with

$$\frac{dN}{dt} = f(N) - h(N).$$

The net population growth rate without hunting is modelled as

$$f(N) = \beta N(1-N),$$

in which the population density is measured in units such that $N=1$ corresponds to the carrying capacity without hunting. The net rate of hunting is

$$h(N) = \frac{\gamma N}{1+\left(\frac{N}{\alpha}\right)}.$$

The parameters $\alpha$, $\beta$ and $\gamma$ are all positive, and you should assume $\gamma > \beta$.

(a) What does $h(N)/N$ represent? Sketch a labelled graph of $h(N)/N$ as a function of $N$. Explain what the sketch indicates about the behaviour of hunters.

[~15% marks]

(b) Comment on the biological meaning of the condition $\gamma > \beta$.

[~5% marks]

(c) Show the model has an equilibrium when $N=0$, as well as up to two further equilibria given by solutions to the quadratic equation

$$N^2 - (1-\alpha)N + \alpha\left(\frac{\gamma}{\beta} - 1\right) = 0.$$

[~10% marks]

(d) Give conditions that ensure this quadratic has two positive (real-valued) solutions.

[~15% marks]

(e) Using a method based on differentiating the right-hand side of the differential equation for $\frac{dN}{dt}$ as a function of $N$, prove that the equilibrium at $N=0$ is stable if $\gamma > \beta$.

[~20% marks]

**QUESTION CONTINUES ON THE NEXT PAGE**

You may assume throughout the remainder of the question that the values of the parameters $\alpha$, $\beta$ and $\gamma$ are such that the model has two distinct equilibria with positive values of $N$, as well as an equilibrium when $N=0$.

(f) Show that it is possible to overlay sketches of $f(N)$ and $h(N)$, both as functions of $N$, in a way that is consistent with there being two equilibrium values with $N>0$.

[~10% marks]

(g) Sketch a graph of $\frac{dN}{dt}$ as a function $N$. Use it to determine the stability of each of the positive equilibrium values.

[~10% marks]

(h) Sketch graphs of $N$ as a function of $t$, showing the behaviour for different initial population sizes. Interpret your sketches in the context of the great auk population.

[~15% marks]

**D7**

For the nonlinear system of equations given by:

$$\frac{dx}{dt} = xy + 3$$

$$\frac{dy}{dt} = xy - x^2 - 4x$$

(a) Find and classify the equilibria.

[~30% marks]

(b) Sketch the phase plane, marking on clearly the nullclines, equilibrium points, direction field and sufficient trajectories to illustrate the behaviour of the system.

[~30% marks]

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

[~25% marks]

(d) Mark on the phase plane a new trajectory for $x(0) = -2, y(0) = 1$. Are the long term solutions for x(t) and y(t) the same as in part (c)? Make use of the direction field and location of nullclines to justify your answer.

[~15% marks]

**D8**

Two strains of *E. coli* bacteria, labelled *X* and *Y*, are grown together in liquid broth. The differential equations below represent the dynamics of the densities of *X* and *Y* respectively:

$$\frac{dX}{dt}=\beta X\left(1-\frac{X+\alpha Y}{K}\right)$$

$$\frac{dY}{dt}=\beta Y\left(1-\frac{X+Y}{K}\right)$$

(a) Give an interpretation for each of the three parameters β, α and K, assuming they can only take positive values.

[~15% marks]

(b) Calculate and sketch the nullclines, stationary points and direction field in the (X,Y) phase plane, considering two alternative cases:

Case 1: $\alpha < 1$;

Case 2: $\alpha > 1$.

[~50% marks]

(c) Determine the stability of the stationary points in Case 1.

[~35% marks]

**E9**

The following scheme describes the mechanism of a fully reversible enzyme catalysed reaction, in which the enzyme, E, facilitates the conversion of a single substrate, S, into a product, P, via the formation of an enzyme-substrate complex, ES.

$$\mathrm{E + S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathrm{E + P}$$

(a) Construct a mass balance equation for the total concentration of enzyme that is present at time zero, $[E]_0$.

[~5% marks]

(b) Write down two differential equations that describe how the rates of change of [ES] and [P] vary as a function of time, *t*.

[~10% marks]

(c) Using your results from part (b), apply the steady-state approximation to the enzyme-substrate complex and construct an expression for [ES] in terms of rate constants, the current concentration of free enzyme, [E], the current concentration of substrate, [S], and the current concentration of product, [P].

[~10% marks]

(d) Using your results from part (c), construct an expression for [ES] in terms of rate constants, the total concentration of enzyme present at time zero, $[E]_0$, the current concentration of substrate, [S], and the current concentration of product, [P].

[~10% marks]

(e) Construct an expression for the rate of change of [P] in terms of rate constants, the total concentration of enzyme present at time zero, $[E]_0$, the current concentration of the enzyme-substrate complex, [ES], and the current concentration of product, [P].

[~5% marks]

**QUESTION CONTINUES ON THE NEXT PAGE**

(f) Using your results from parts (d) and (e), show that $V$, the rate of formation of product, obeys the following rate law equation:

$$V=\frac{\frac{V_{MAX}^{f}}{K_{M}^{f}}[S]-\frac{V_{MAX}^{b}}{K_{M}^{b}}[P]}{1+\frac{[S]}{K_{M}^{f}}+\frac{[P]}{K_{M}^{b}}}$$

where $V_{MAX}^{f}$ and $V_{MAX}^{b}$ are the maximal rates of reaction in the forward and the backward directions respectively, and $K_{M}^{f}$ and $K_{M}^{b}$ are the Michaelis constants for the forward and the backward reactions respectively. Calculate explicitly how $V_{MAX}^{f}$ , $V_{MAX}^{b}$ , $K_{M}^{f}$ and $K_{M}^{b}$ depend on the values of rate constants $k_1$, $k_2$, $k_3$ and $k_4$.

[~30% marks]

(g) Discuss the effect on $V$ of conducting experiments under initial rate conditions when none of the product is present.

[~10% marks]

(h) Show that if this reversible enzyme catalysed reaction is allowed to run until it reaches equilibrium, the equilibrium constant, $K_{eq}$, will be equal to the ratio of the catalytic efficiencies of the forward and the backward reactions.

[~20% marks]

**E10**

A fish population of size $N$ exhibits logistic growth, with an intrinsic growth rate of $r$ and a carrying capacity of $k$. A fishing authority determines the level of effort $E$ expended on exploiting the population. Assuming that the rate at which fish are caught is proportional to the product of fishing effort and population size, the growth of the population is described by the differential equation

$$\frac{dN}{dt}=rN\left(1-\frac{N}{k}\right)-qEN$$

where the parameter $q$ specifies fishing efficiency.

(a) Determine the non-zero equilibrium population size $N_{eq}$ and corresponding equilibrium catch rate $C_{eq} = qEN_{eq}$.

[~20% marks]

(b) Using your result from (a), obtain an expression for the optimal catching effort $E_{opt}$ that maximises the equilibrium catch rate.

[~20% marks]

(c) Now suppose that the fish population is exploited by two distinct fishing authorities, with fishing efforts of $E_1$ and $E_2$, so that its growth is given by

$$\frac{dN}{dt}=rN\left(1-\frac{N}{k}\right)-q\left(E_1+E_2\right)N$$

Re-derive the non-zero equilibrium population size $N_{eq}$ in this case. From this, determine the optimal catching effort for the first authority, $E_{1opt}(E_2)$, which maximises its own equilibrium catch rate $C_{1eq} = qE_1N_{eq}$ given $E_2$. Briefly describe in words how the first authority's optimal catching effort changes with the second authority's effort.

[~40% marks]

(d) Derive an expression for the stable level of exploitation by each authority in response to its own fishing effort (i.e. derive $E^*$ for which $E_{1opt}(E^*) = E^*$).

[~20% marks]