

# Student Assignment Brief

---

This document is intended for Coventry University Group students for their own use in completing their assessed work for this module. It must not be passed to third parties or posted on any website. If you require this document in an alternative format, please contact your Module Leader.

---

## Contents:

- [Assignment Information](#)
- [Assignment Task](#)
- [Marking and Feedback](#)
- [Assessed Module Learning Outcomes](#)
- [Assignment Support and Academic Integrity](#)
- [Assessment Marking Criteria](#)

The work you submit for this assignment must be your own independent work, or in the case of a group assignment your own groups' work. More information is available in the '[Assignment Task](#)' section of this assignment brief.

---

## Assignment Information

**Module Name:** Principles of Data Science

**Module Code:** 7144CEM

**Assignment Title:** Individual Coursework

**Assignment Due:** Friday 01/12/2023 at 6 pm UK time

**Assignment Credit:** 10 credits

**Word Count (or equivalent):** 2000 words equivalent (not including reference list or output )

**Assignment Type:** Written

**Percentage Grade** (Applied Core Assessment). You will be provided with an overall grade between 0% and 100%. You have one opportunity to pass the assignment at or above 40%.

---

## Assignment Task

This *Individual Coursework* involves investigating the different properties and operations involving vectors, matrices, and probability, including discrete dynamical systems and polynomial regression. You are encouraged to explore the topic, use your initiative, and show some originality, within the time available. Make sure you read each task through carefully, and answer all parts of each task. Aim to demonstrate your understanding of the topics and the relevant module learning outcomes.

### **Task 1: Population Dynamics in Ecosystems**

A discrete dynamical system refers to a mathematical model used to describe the evolution of a system over discrete time steps. It is often used to analyse how the state of a system changes as time progresses in distinct intervals. In such systems, the state of the system at any given time step depends on its state at the previous time step, and this evolution is defined by a set of equations or rules.

Consider the populations of two animal species in an ecosystem: rabbits and foxes, with the following dynamics:

- Rabbits: The rabbit population increases by 5% through natural reproduction and decreases by 5% of the foxes' population due to predation.
- Foxes: The fox population decreases by 8% due to natural death and increases by 5% of the rabbit population through successful hunting for food.

Therefore, at each discrete time step (e.g., one week), we can update the populations of rabbits and foxes as follows:

- New Rabbit Population = Old Rabbit Population + 0.05 \* Old Rabbit Population - 0.05 \* Old Fox Population
- New Fox Population = Old Fox Population - 0.08 \* Old Fox Population + 0.05 \* Old Rabbit Population

This discrete dynamical system captures how the populations of rabbits and foxes interact over time. The populations in the next time step are determined by the populations in the current time step. By repeatedly applying these equations, we can simulate the evolution of rabbit and fox populations over multiple time steps. The behaviour of the system might exhibit various patterns, including cyclic behaviour, convergence to equilibrium, or chaotic behaviour, depending on the parameters and interactions defined in the model.

- (a) Model this dynamical system using matrix equations by writing the population of rabbits and foxes at the current time step in a  $2 \times 1$  vector  $\mathbf{p}$ . Construct a  $2 \times 2$  matrix  $A$  (as a numpy array) that encapsulates these population changes so that the populations at the next time step is  $A\mathbf{p}$ .

[10 marks]

- (b) Assume the ecosystem begins with 2000 foxes and 8000 rabbits. Use numpy to calculate the expected population of each species at the end of each time step for the next 100 steps using matrix multiplication. Generate a line graph using Python, depicting the relative population trends, where the relative population is the percentage of foxes and rabbits over time. Include a legend and provide a brief analysis of your observations.

[10 marks]

- (c) Use numpy to determine the eigenvalues of the matrix  $A$  from part (a). Explain how the magnitude of these eigenvalues corresponds to the stability of the fox and rabbit populations in the long run. Discuss how one of these eigenvalues are linked to the dominant factor driving the population dynamics.

[10 marks]

- (d) Modify the matrix  $A$ , so that the rabbit population decreases by 200% of the foxes' population due to predation, and the other settings are all the same as part (b). Generate a graphical plot, depicting the population trends, where the  $x$ -axis is the rabbit population and  $y$ -axis is the fox population. Try different starting points of initial population for each species and add the trajectories to the plot. Investigate the behaviour of these trajectories and make clear conclusions based on what you observe.

[20 marks]

## **Task 2: Regression by Matrix Operations**

Firstly, a bit of background is needed. Consider the three points (0,1), (3,4), and (6,5), plotted below. We wish to find the line of best fit through these points, i.e., a line of the form  $y = a + bx$ . Assemble the  $x$ -coordinates in the matrix  $X$  and the  $y$ -coordinates in the vector  $y$  as follows.

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \text{ and } y = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$$

Then

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 9 \\ 9 & 45 \end{bmatrix}$$

giving

$$(X^T X)^{-1} = \begin{bmatrix} 5/6 & -1/6 \\ -1/6 & 1/18 \end{bmatrix}$$

and also

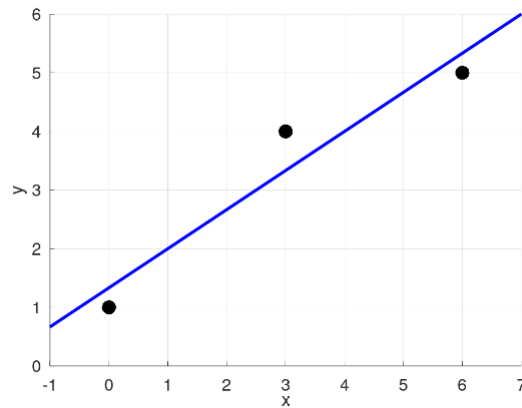
$$X^T y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 42 \end{bmatrix}$$

Finally

$$\beta = (X^T X)^{-1} X^T y = \begin{bmatrix} 5/6 & -1/6 \\ -1/6 & 1/18 \end{bmatrix} \begin{bmatrix} 10 \\ 42 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 2/3 \end{bmatrix}$$

which are the intercept and multiplier, i.e., the line of best fit is

$$y = \frac{4}{3} + \frac{2}{3}x$$



The fitted (predicted) values are

$$X\beta = \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} 4/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 10/3 \\ 16/3 \end{bmatrix}$$

- (a) Implement the matrix procedure described above, using numpy, for the general case of linear regression with one predictor variable and one response variable. Check the results using the example above and use matplotlib to plot a scatterplot and the fitted line. Calculate the  $R^2$  value only using numpy. You may find `numpy.hstack()`, `numpy.ones()`, `numpy.reshape()` and `numpy.corrcoef()` useful.

[10 marks]

- (b) Suppose we wish to find the *quadratic* polynomial of best fit, i.e.,  $y = a + bx + cx^2$ . The procedure described above is simply adapted by extending the  $X$  matrix using columns of 1,  $x$  and  $x^2$  as follows.

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 9 \\ 1 & 6 & 36 \end{bmatrix}$$

Then, using exactly the same matrix equations we obtain

$$\beta = \begin{bmatrix} 1 \\ 4/3 \\ -1/9 \end{bmatrix}$$

so the quadratic of best fit is

$$y = 1 + \frac{4}{3}x - \frac{1}{9}x^2$$

Modify your Python implementation from part (1) to carry out the matrix procedure described above for a general case of “quadratic regression” with one predictor variable and one response variable. Check the results using the example above and use matplotlib to plot a scatterplot and the fitted quadratic curve. Calculate the  $R^2$  value only using numpy.

[5 marks]

(c) Similarly, we can easily extend this matrix-based procedure to polynomials of higher degree by adding further columns to the matrix  $X$  corresponding to higher powers of  $x$ . We wish to apply “polynomial regression” to use the day of the year (day 1 is 1 January 2022 through day 617 which is 10 September 2023) to predict the *length* of that day (minutes between sunrise and sunset) in Coventry, UK. *The only Python libraries you may use in this part are numpy and matplotlib, except that you may use pandas only to load the dataset and convert the relevant columns to numpy arrays. You may use numpy.polyfit() only to check your output but model fitting should be carried out using matrix operations.*

- (1) Assemble this dataset from <https://sunrise-sunset.org/search?location=coventry> using the column “Day length”. You must use at least two dates from each month from January 2022 through to September 2023.

[5 marks]

- (2) Modify your Python matrix-based implementation from part (b) to fit quadratic (degree 2 polynomial) and cubic (degree 3 polynomial) models to this day length dataset. Give the equation of each fitted model, plot each of these fitted models on the same axes, and calculate the  $R^2$  value in each case. Compare and critique your fitted models.

[20 marks]

- (3) Based on your conclusions from part (2), investigate at least one further polynomial regression model applied to this day length dataset. What model would you recommend to predict day length from day of the year? *Briefly justify your recommendation.*

[10 marks]

## Submission Instructions:

Ensure that your coursework is all your own work and you clearly cite and reference any sources you have used using [APA style](#) referencing. Please include both in-text citations and a list of references for each task (where relevant). **No collaboration with other students is permitted.**

Please submit one report (e.g. as a single Microsoft Word document or a single PDF document) covering all of the tasks above, clearly organised by subtask. Start each task on a new page. Make sure you include your Python code and relevant output/plots directly in the report. You must not submit a zip file. You must not submit a Jupyter notebook (but you can print a Jupyter notebook to a PDF file and submit the PDF file).

Do not use screenshots to include Python code or text output into your report. If your report is a Microsoft Word file, then please use a syntax highlighter (such as <http://hilight.me/>) and copy-and-paste the coloured Python code into your report. Also simply copy-and-paste text output into your report.

Submission is online via Aula using the submission box provided. *Do not leave uploading too late.*

## Marking and Feedback

### How will my assignment be marked?

Your assignment will be marked by the module team.

### How will I receive my grades and feedback?

Provisional marks will be released once internally moderated. Feedback will be provided by the module team alongside grades release. Students will be able to access their feedback via Aula/Turnitin. Your provisional marks and feedback should be available within 2 weeks (11 working days).

### What will I be marked against?

Details of the marking criteria for this task can be found at the [bottom of this assignment brief](#).

---

## Assessed Module Learning Outcomes

The Learning Outcomes for this module align to the [marking criteria](#) which is provided above. Ensure you understand the marking criteria to ensure successful achievement of the assessment task. The following module learning outcomes are assessed in this task:

**MLO 1.** Demonstrate systematic knowledge and critical understanding in topics in linear algebra, probability and statistical models, relevant to data science.

## Assignment Support and Academic Integrity

If you have any questions about this assignment please see the [Student Guidance on Coursework](#) for more information.

### Spelling, Punctuation, and Grammar:

You are expected to use effective, accurate, and appropriate language within this assessment task.

### Academic Integrity:

The work you submit must be your own, or in the case of groupwork, that of your group. All sources of information need to be acknowledged and attributed; therefore, you must provide references for all sources of information and acknowledge any tools used in the production of your work, including Artificial Intelligence (AI). We use detection software and make routine checks for evidence of academic misconduct.

Definitions of academic misconduct, including plagiarism, self-plagiarism, and collusion can be found [on the Student Portal](#). All cases of suspected academic misconduct are referred for investigation, the outcomes of which can have profound consequences to your studies. For more information on academic integrity please visit the [Academic and Research Integrity](#) section of the Student Portal.

## **Support for Students with Disabilities or Additional Needs:**

If you have a disability, long-term health condition, specific learning difference, mental health diagnosis or symptoms and have discussed your support needs with health and wellbeing you may be able to access support that will help with your studies.

If you feel you may benefit from additional support, but have not disclosed a disability to the University, or have disclosed but are yet to discuss your support needs it is important to let us know so we can provide the right support for your circumstances. Visit [the Student Portal](#) to find out more.

## **Unable to Submit on Time?**

The University wants you to do your best. However, we know that sometimes events happen which mean that you cannot submit your assessment by the deadline or sit a scheduled exam. If you think this might be the case, guidance on understanding what counts as an extenuating circumstance, and how to apply is [available on the Student Portal](#).

---

## **Administration of Assessment**

**Module Leader Name:** Dr Omid Chatrabgoun

**Module Leader Email:** [ad8337@coventry.ac.uk](mailto:ad8337@coventry.ac.uk)

**Assignment Category:** Written

**Attempt Type:** Standard

**Component Code:** CW





## Assessment Marking Criteria

Mark band	Outcome	Guidelines
90-100% Distinction	Meets learning outcomes	Distinction - Exceptional work with very high degree of rigour, creativity and critical/analytic skills. Mastery of knowledge and subject-specific theories with originality and autonomy. Demonstrates exceptional ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with exceptional ability in the utilisation of research methodologies. Demonstrates, creativity, originality and outstanding problem-solving skills. Work completed with very high degree of accuracy, proficiency and autonomy. Exceptional communication and expression demonstrated throughout. Student evidences the full range of technical and/or artistic skills. Work pushes the boundaries of the discipline and may be strongly considered for external publication/dissemination/presentation.
80-89% Distinction		Distinction - Outstanding work with high degree of rigour, creativity and critical/analytic skills. Near mastery of knowledge and subject-specific theories with originality and autonomy. Demonstrates outstanding ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with outstanding ability in the utilisation of research methodologies. Work consistently demonstrates creativity, originality and outstanding problem-solving skills. Work completed with high degree of accuracy, proficiency and autonomy. Outstanding communication and expression demonstrated throughout. Student demonstrates a very wide range of technical and/or artistic skills. With some amendments, the work may be considered for external publication/dissemination/presentation
70-79% Distinction		Distinction - Excellent work undertaken with rigour, creativity and critical/analytic skills. Excellent degree of knowledge and subject-specific theories with originality and autonomy demonstrated. The work exhibits excellent ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Innovative research with excellent ability in the utilisation of research methodologies. Work demonstrates creativity, originality and excellent problem-solving skills. Work completed with very consistent levels of accuracy, proficiency and autonomy. Excellent communication and expression demonstrated throughout. Student demonstrates a very wide range of technical and/or artistic skills.

60-69%		Merit - Very good work often undertaken with rigour, creativity and critical/analytic skills. Very good degree of knowledge and subject-specific theories with some originality and autonomy demonstrated. The work often exhibits the ability to fully analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Very good research evidence and shows very good ability in the utilisation of research methodologies. Work demonstrates creativity, originality and problem-solving skills. Work completed with very consistent levels of accuracy, proficiency and autonomy. Very good communication and expression demonstrated throughout. Student demonstrates a wide range of technical and/or artistic skills.
50-59%		Pass - Good work undertaken with some creativity and critical/analytic skills. Demonstrates knowledge and subject-specific theories with some originality and autonomy demonstrated. The work exhibits the ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Good research and shows some ability in the utilisation of research methodologies. Work demonstrates problem-solving skills and is completed with some level of accuracy, proficiency and autonomy. Satisfactory communication and expression demonstrated throughout. Student demonstrates some of the technical and/or artistic skills.
40-49%		Pass - Assessment demonstrates some advanced knowledge and understanding of the subject informed by current practice, scholarship and research. Work may be incomplete with some irrelevant material present. Sometimes demonstrates the ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Acceptable research with evidence of basic ability in the utilisation of research methodologies. Demonstrates some originality, creativity and problem-solving skills but often with inconsistencies. Expression and presentation sufficient for accuracy and proficiency. Sufficient communication and expression with professional skill set. Student demonstrates some technical and/or artistic skills.
30-39%	Fails to achieve learning outcomes	Fail - Very limited understanding of relevant theories, concepts and issues with deficiencies in rigour and analysis. Some relevant material may be present but be informed from very limited sources. Fundamental errors and some misunderstanding likely to be present. Demonstrates limited ability to analyse and apply concepts within the complexities and uncertainties of the subject/discipline. Limited research scope and ability in the utilisation of research methodologies. Limited originality, creativity, and struggles with problem-solving skills. Expression and presentation insufficient for accuracy and proficiency. Insufficient communication and expression and with deficiencies in professional skill set. Student demonstrates deficiencies in the range of technical and/or artistic skills.

20-29%		Fail - Clear failure demonstrating little understanding of relevant theories, concepts, issues and only a vague knowledge of the area. Little relevant material may be present and informed from very limited sources. Serious and fundamental errors and virtually no evidence of relevant research. Fundamental errors and misunderstandings likely to be present. Little or no research with no evidence of utilisation of research methodologies. No originality, creativity, and struggles with problem-solving skills. Expression and presentation insufficient for accuracy and proficiency. Insufficient communication and expression and with serious deficiencies in professional skill set. Student has clear deficiencies in range of technical and/or artistic skills.
0-19%		Fail - Clear failure demonstrating no understanding of relevant theories, concepts, issues and no understanding of area. Little or no relevant material may be present and informed from minimal sources. No evidence of ability in the utilisation of research methodologies. No evidence of originality, creativity, and problem-solving skills. Expression and presentation deficient for accuracy and proficiency. Insufficient communication and expression and with deficiencies in professional skill set. Student has clear deficiencies in range of technical and/or artistic skills.