

Assessment 2: Multiple Regression (Chapter 4)**Due: 21st March**

Refer to the **Assessment Criteria and Guidelines** and the **Sample Assignment and Solutions** in the Assessment block of the Moodle site.

Five marks will be awarded for the following:

- Clear expression, correct use of terminology and notation.
- Presentation of figures and tables: ensure that you include all relevant R output.
- Clearly and concisely annotated R code: include your R script file as an appendix.

Question**[95 marks]**

Microplastics are found in almost all marine and fresh water environments, where they pose a potential risk to fish and crustaceans. Therefore, the effects of microplastics on aquatic organisms are currently the subject of intense research. Here we have a dataset containing 200 seawater samples collected at different sites around a bay in Sweden. The dataset is saved as `Mplastics.csv`. There are 5 variables included in the study:

- PE = polyethylene microplastics ($\mu g/m^3$)
- PP = polypropylene microplastics ($\mu g/m^3$)
- PS = polystyrene microplastics ($\mu g/m^3$)
- temp = water temperature at each site (Celsius)
- larvae = number of fish larvae of a single species per 100 m^3

- (a) Use the R function `pairs` to plot the data. Summarise the information available from the plot. [15 marks]

NB: save `Rfunctions.R` (available from Topic 2 block) to your working directory.

- (b) Fit a model of the form [10 marks]

$$larvae = \beta_0 + \beta_1 PE + \beta_2 PP + \beta_3 PS + \beta_4 temp + \epsilon$$

Print the table of regression coefficients and write down the least squares regression equation.

- (c) Which variables are significant predictors of fish larvae density in this model, at a 5% level? Write down and test appropriate hypotheses. [15 marks]

Now drop the non-significant terms and refit the model using only the explanatory variables that are significant. This is referred to as the final model.

- (d) Print the table of regression coefficients and write down the least squares regression equation for the final model. [10 marks]
- (e) Find and interpret the 95% confidence intervals for the regression coefficients in the final model. [10 marks]
- (f) Produce the diagnostic plots for the final model and explain what can be understood from the plots. [10 marks]
- (g) Using the final model, estimate *mean fish larvae density* and the 95% CI for the estimate when [15 marks]
- temperature = 17.5, PE = 300, PP = 50, PS = 80.5 and
 - temperature = 20.5, PE = 300, PP = 50, PS = 80.5

NB: You may not need to include values for all 4 predictor variables, depending on your final model.

Comment on the reliability of these predictions.

- (h) Write a concise, informative conclusion based on your analysis and results. [10 marks]