Multiple Linear regression

In this chapter:

- Multiple Linear Regression

Elsewhere in the Book

- Simple Linear Regression
- Polynomial Regression
- Generalised Linear Models
- Nonlinear Regression
- Generalised Additive Models

In **?@sec-simple-linear-regression** we have seen how to model the relationship between two variables using simple linear regression (SLR). However, in ecosystems, the relationship between the response variable and the explanatory variables is more complex and cannot be adequately captured by a single driver (i.e. influential or predictor variable). In such cases, we can use multiple linear regression to model the relationship between the response variable and multiple explanatory variables.

Multiple Linear Regression

Multiple linear regression helps us answer questions such as:

- How do various environmental factors influence the population size of a species? Factors like average temperature, precipitation levels, and habitat area can be used to predict the population size of a species in a given region. Which of these factors are most important in determining the population size?
- What are the determinants of plant growth in different ecosystems? Variables such as soil nutrient content, water availability, and light exposure can help predict the growth rate of plants in various ecosystems. How do these factors interact to influence plant growth?
- How do genetic and environmental factors affect the spread of a disease in a population? The incidence of a disease might depend on factors like genetic susceptibility, exposure to pathogens, and environmental conditions (e.g., humidity and temperature). What is the relative importance of these factors in determining the spread of the disease?

Multiple linear regression extends the simple linear regression model to include multiple independent variables. The model is expressed as:

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \epsilon_i \tag{1}$$

Where:

- Y_i is the response variable for the i-th observation,
- $X_{i1}, X_{i2}, \dots, X_{ik}$ are the k predictor variables for the i-th observation,
- α is the intercept,
- $\beta_1, \beta_2, \dots, \beta_k$ are the coefficients for the k predictor variables, and
- ϵ_i is the error term for the i-th observation (the residuals).

Assumptions of Multiple Linear Regression

As discussed in SLR.

Multicollienarity

- Definition and problems caused by multicollinearity.
- Detection methods (Variance Inflation Factor (VIF), correlation matrix).
- Remedies for multicollinearity (dropping variables, combining variables, ridge regression).

Model Fitting and Interpretation

- How to fit an MLR model in R.
- Interpretation of coefficients (0, 1, …, k).
- Assessing model fit: R-squared, adjusted R-squared, F-test.

Model Diagnostics

As discussed in SLR.

Variable Selection Methods

- Forward selection.
- Backward elimination.
- Stepwise regression.
- Criteria for selection (AIC, BIC, adjusted R-squared).

Interaction Terms