**Task 1: Epidemiology study of sea lice**

**Question two: Effect of environmental variables on sea lice abundance**

* + Using MI Individual Database
  + Looking at the influence of temperature (and its changes throughout the years) and time of the year (month) on sea lice abundance
  + On bay level
  + Mixed-effects linear regression
    - Response variable: sea lice count total mobile
    - Random effect: nested (Farm/Pen)
    - Predictors: temperature, oxygen, salinity, month, year, time in seawater, class of fish (S0/S1/2), company (three levels: Mowi, Bradan, others)

**Structure of the model**

First without random effects

Lepeophtheirus Salmonis Total ~ Bay Name + Temperature + Salinity + Oxygen + Year + Month + Time in Sea + Class + Company

**Distribution**

* **Poisson:** high overdispersion, underfitting of zeros
* **Quasi-Poisson:** underfitting of zeros
* **Negative Binomial:** overdispersion 1.5, zeros are ok
* **Hurdle models:** 
  + **Poisson:** underfitting of zeros
  + **Negative Binomial:** overdispersion 1.22, zeros are ok
* **Zero-inflated models:** 
  + **Poisson:** underfitting of zeros
  + **Negative Binomial:** overdispersion 1.57, underfitting of zeros

From all the distribution the zeros fit well in Negative binomial and Hurdle negative binomial. Overall, the hurdle model shows better fit.

Focusing on Hurdle negative binomial model for further work

**ZANB (Hurdle negative binomial model):**

Detection of Heteroskedasticity

Predicted values by the model are not very accurate, the prediction for high values (more than 30 doesn't work at all)

Detection of multicollinearity (Count part of the model: Bay name, Temperature, Year, Month, Company; Zero part of the model: Bay name, Temperature, Month, Company)

Detection of temporal autocorrelation