SS6 – Advanced Statistical Modelling
Lab Session

The dataset `Dataset_4species_rep_measures_simulated.csv` will be used in this lab session. The data in this file is simulating a biodiversity and ecosystem function (BEF) experiment. The diversity of four grassland species was manipulated across 50 plots. The first two species are grasses, and the second two legumes. The annual yield of each plot over three growing seasons was recorded.

Analysing repeated measures data

1) Identify the best repeated measures Diversity-Interactions (DI) model fitted to the three years of data. Use the best model in each step for the following step:
   a) Using maximum likelihood (ML), fit the full pairwise repeated measures DI model to all three years of data using the 'unstructured' repeated measures covariance structure, letting theta be set to 1. Compare the two models using AICc.
   b) Using ML, Fit the ID, AV, FG and FULL pairwise interaction DI models. Compare the models using AICc.
   c) Using restricted maximum likelihood (REML), fit the unstructured, compound symmetry and AR(1) covariance structures. Compare the models using AICc.
   d) If the covariance structure in (c) differs from UN, repeat steps (a) and (b) using the identified structure.
   e) Fit the final model using REML and examine the estimates.

2) For the best repeated measures model identified:
   a) Write out the equation of the model and its assumptions.
   b) Prepare three graphs using the `DImodelsVis` package to aid interpretation of the final model.
   c) Create contrasts to compare:
      - The grass 1 identity effect in years 1 and 2
      - The equi-proportional mixture in years 1 and 2, and in year 2 and 3.
      - Any additional comparisons of interest.
   d) Write out 3-5 sentences to describe and interpret the final model.
   e) How would ability to interpret the data change if each year was analysed separately?
   f) Which of these situations is better and how does ability to interpret the data change for each?
      - Each year of data was modelled in a repeated measures model using sown proportions as the predictors.
      - Each year of data was modelled in a repeated measures model using the preceding year's proportions as predictors. I.e., sown for year 1, year 1 realised proportions for year 2, and year 2 realised proportions for year 3.

The LegumeLegacy project has received funding from the European Union's Horizon 2021 Doctoral Network programme under the Marie Skłodowska-Curie grant agreement No. 101072579

R code that may be useful.
When preparing your own script, take care to make good comments throughout.

```
library(DImodels)
library(DImodelsMulti)
library(ggplot2)
library(PieGlyph)
library(DImodelsVis)
library(nlme)
library(contrast)

dataRM <- read.csv("Dataset_4species_rep_measures_simulated.csv", header = TRUE)
dataRM_y1 <- subset(dataRM, yearn == 1)
dataRM_y2 <- subset(dataRM, yearn == 2)
dataRM_y3 <- subset(dataRM, yearn == 3)

dataRM$yearnF <- as.factor(dataRM$yearn)


#### Repeated measures analysis

m1_theta_1 <- DImulti(y = "yield", time = c("yearn", "UN"), unit_IDs = "plot",
          prop = c("G1", "G2", "L1", "L2"), FG = c("G","G","L","L"),
          data = dataRM, DImodel = "FULL",
             method = "ML", estimate_theta = FALSE)
AICc(m1_theta_1)

m1_theta_est <- DImulti(y = "yield", time = c("yearn", "UN"), unit_IDs = "plot",
                prop = c("G1", "G2", "L1", "L2"), FG = c("G","G","L","L"),
                data = dataRM, DImodel = "FULL",
                method = "ML", estimate_theta = TRUE)
AICc(m1_theta_est)
print(m1_theta_est)


m2_ID <- DImulti(y = "yield", time = c("yearn", "UN"), unit_IDs = "plot",
            prop = c("G1", "G2", "L1", "L2"), FG = c("G","G","L","L"),
            data = dataRM, DImodel = "ID",
            method = "ML", estimate_theta = FALSE)
AICc(m2_ID)


m3_FG_UN <- DImulti(y = "yield", time = c("yearn", "UN"), unit_IDs = "plot",
            prop = c("G1", "G2", "L1", "L2"), FG = c("G","G","L","L"),
            data = dataRM, DImodel = "FG",
            method = "REML", estimate_theta = FALSE)
AICc(m3_FG_UN)
```

The LegumeLegacy project has received funding from the European Union's Horizon 2021 Doctoral Network programme under the Marie Skłodowska-Curie grant agreement No. 101072579

```
model_final <- DImulti(y = "yield", time = c("yearn", "UN"), unit_IDs = "plot",
                       prop = c("G1", "G2", "L1", "L2"), FG = c("G","G","L","L"),
                       data = dataRM, DImodel = "FG",
                       method = "REML", estimate_theta = FALSE)
AICc(model_final)
summary(model_final)
print(model_final)

gradient_change(model_final)

visualise_effects(model_final)

grouped_ternary(model_final, FG = c("Gr1","Gr2","L","L"))
```

The text in this text box is for carrying out contrasts to test the final repeated measures model. Explanation comments are provided for each chunk of code.

```
################################################################
### Contrasts

# We are going to refit our model using gls().
# We can extract the call used by DImulti() as follows (or using summary())

model_final$call$model

summary(model_final$corr$`Repeated Measure`)


# Now, we need to add our interaction columns to our dataset.
# We can easily do this using the DImodels package.

dataRM_gls <- cbind(dataRM, DI_data(prop = c("G1", "G2", "L1", "L2"),
                          FG = c("G","G","L","L"),
                          theta = 1,
                          data = dataRM,
                          what = "FG"))

names(dataRM_gls)[5:8] <- paste0(names(dataRM_gls)[5:8], "_ID")
names(dataRM_gls)[11:13] <- paste0("FG.", names(dataRM_gls)[11:13])


# Next, we fit our gls model in the form the contrast package requires,
# i.e. we drop a species, in this case L2, and expand our interactions ":"
# to include each effect separately "*"


model_final_gls <- gls(model = yield ~ 0
```

The LegumeLegacy project has received funding from the European Union's Horizon 2021 Doctoral Network programme under the Marie Skłodowska-Curie grant agreement No. 101072579

```
                    + yearnF*((G1_ID + G2_ID + L1_ID
                            + FG.bfg_G_L + FG.wfg_G + FG.wfg_L)),
            correlation = corSymm(form = ~ 0 | plot),
            weights = varIdent(form = ~ 0 | yearnF),
            data = dataRM_gls,
            method = "REML")
AICc(model_final_gls)
summary(model_final_gls)
model_final_gls_predictions <- predict(model_final_gls)
dataRM_gls_preds <- dataRM_gls
dataRM_gls_preds$preds <- model_final_gls_predictions

# Compare G1_ID in year 1 versus year 2 (they significantly differ)

contrast(model_final_gls,
      list(yearnF = "1", yearn = 1,
         G1_ID = 1, G2_ID = 0, L1_ID = 0,
         FG.bfg_G_L = 0, FG.wfg_G = 0, FG.wfg_L = 0),
      list(yearnF = "2", yearn = 2,
         G1_ID = 1, G2_ID = 0, L1_ID = 0,
         FG.bfg_G_L = 0, FG.wfg_G = 0, FG.wfg_L = 0))

# Compare G1_ID in year 1 versus year 3 (they do not significantly differ)

contrast(model_final_gls,
      list(yearnF = "1", yearn = 1,
         G1_ID = 1, G2_ID = 0, L1_ID = 0,
         FG.bfg_G_L = 0, FG.wfg_G = 0, FG.wfg_L = 0),
      list(yearnF = "3", yearn = 3,
         G1_ID = 1, G2_ID = 0, L1_ID = 0,
         FG.bfg_G_L = 0, FG.wfg_G = 0, FG.wfg_L = 0))

# Compare G1 monoculture to equi-proportional mixture in year 1
# (they significantly differ)

contrast(model_final_gls,
      list(yearnF = "1", yearn = 1,
         G1_ID = 1, G2_ID = 0, L1_ID = 0,
         FG.bfg_G_L = 0, FG.wfg_G = 0, FG.wfg_L = 0),
      list(yearnF = "1", yearn = 1,
         G1_ID = 0.25, G2_ID = 0.25, L1_ID = 0.25,
         FG.bfg_G_L = 0.25, FG.wfg_G = 0.0625, FG.wfg_L = 0.0625))


# Compare equi-proportional mixture in year 1 and year 2
# (they significantly differ)

contrast(model_final_gls,
      list(yearnF = "1", yearn = 1,
```

The LegumeLegacy project has received funding from the European Union's Horizon 2021 Doctoral Network programme under the Marie Skłodowska-Curie grant agreement No. 101072579

```
            G1_ID = 0.25, G2_ID = 0.25, L1_ID = 0.25,
            FG.bfg_G_L = 0.25, FG.wfg_G = 0.0625, FG.wfg_L = 0.0625),
        list(yearnF = "2", yearn = 2,
            G1_ID = 0.25, G2_ID = 0.25, L1_ID = 0.25,
            FG.bfg_G_L = 0.25, FG.wfg_G = 0.0625, FG.wfg_L = 0.0625))

# Compare equi-proportional mixture in year 1 and year 2
# (they significantly differ)

contrast(model_final_gls,
      list(yearnF = "2", yearn = 2,
            G1_ID = 0.25, G2_ID = 0.25, L1_ID = 0.25,
            FG.bfg_G_L = 0.25, FG.wfg_G = 0.0625, FG.wfg_L = 0.0625),
      list(yearnF = "3", yearn = 3,
            G1_ID = 0.25, G2_ID = 0.25, L1_ID = 0.25,
FG.bfg_G_L = 0.25, FG.wfg_G = 0.0625, FG.wfg_L = 0.0625))
```

The LegumeLegacy project has received funding from the European Union's Horizon 2021 Doctoral Network programme under the Marie Skłodowska-Curie grant agreement No. 101072579