Tutorial #11: Species community composition

Guy F. Sutton

Centre for Biological Control, Rhodes University

11. Species community composition

In the previous tutorials, we have been analysing a single response variable. However, it is very common for ecologists to collect species community composition data, whereby they record the abundance of multiple species occuring in a community. For example, recording the presence and abundance of the 10 insect species that occur on a plant species, or recording the abundance of ectoparasites (e.g. ticks) on different antelope species. While it is tempting to run a seperate GLM to each species, this would not be appropriate because the data points are not independent. Moreover, you often want to make inference about the community itself - e.g. does insect community composition change between seasons? Do different antelope species host different ectoparasite communities?

To do this, we can use a statistical test called a `multivariate GLM`. You may have heard of a test called Permutational ANOVA (*PERManova*) to compare community composition. Unfortunately, this test is used alot, but is usually not an appropriate fit to the data - it makes assumptions about the variance in the data that usually aren't met.

Below, we are going to work through a dataset that I collected during my PhD. The dataset contains the abundances of six insect species that I recorded on a grass species that I worked on. I collected these data from sites that I sampled in different seasons (`season`) and at sites that had been recently disturbed by mowing or left undisturbed (`disturb_allow_overwinter`). We wanted to know if species community composition differed between seasons and between disturbed versus undisturbed sites, and if so, which of the individual insect species explained any observed shifts in species community composition between seasons or disturbed versus undisturbed sites.

11.1. Fit the global model - does community composition change?

To fit the `multivariate GLM`, we need to give `R` a species community composition matrix. This is basically just a survey x species matrix of abundances, where each row contains abundances from different surveys, and the columns contain the abundances for the different species. *Note that you cannot have a row total = 0 (i.e. the total abundance across all species = 0). The models will not work. You can simply just add a dummy species with abundance = 1, like I have done below so solve this (sp_dum1).*

```
##      sp_tet1 sp_tet2 sp_bru1 sp_sht1 sp_chl1 sp_eur4 sp_dum1
## [1,]      14       0       2       2       0       0       1
## [2,]       0       0       0       0       0       0       1
## [3,]       0       6       0       4       0       0       1
## [4,]       0       0       0       0       0       0       1
## [5,]      14       2       3       2       0       0       1
## [6,]       0       0       0       0       0       0       1
## [7,]       4       5       4       4       0       0       1
```

```
##    [8,]       9       6       2       1       2       0       1
##    [9,]       4       4       6       4       1       1       1
##   [10,]       3       4      10       3       3       4       1
##   [11,]       0       1      10       0       0       0       1
##   [12,]       0       0       0       0       0       0       1
##   [13,]       0       0       1       0       0       0       1
##   [14,]       2       1       2       0       0       0       1
##   [15,]       1       4       1       0       0       0       1
##   [16,]       1       1       2       0       0       0       1
##   [17,]       1       2       1       0       0       0       1
##   [18,]       0       0       0       0       0       0       1
##   [19,]       0       0       0       0       0       0       1
##   [20,]       0       0       0       0       0       0       1
##   [21,]       0       0       0       0       0       0       1
##   [22,]       0       0       0       0       0       0       1
##   [23,]       0       0       0       0       0       0       1
##   [24,]       0       0       0       0       0       0       1
##   [25,]       0       0       0       0       0       0       1
##   [26,]       0       0       0       0       0       0       1
##   [27,]       0       0       0       0       0       0       1
##   [28,]       0       0       0       0       0       0       1
##   [29,]       1       0       0       4       1       0       1
##   [30,]       1       0       0       0       0       0       1
##   [31,]       0       0       0       0       1       0       1
##   [32,]       0       0       0       0       0       0       1
##   [33,]       0       0       0       4       0       0       1
##   [34,]       0       0       0       0       1       0       1
##   [35,]       3       0       0       0       0       0       1
##   [36,]       1       0       0       5       1       0       1
##   [37,]       0       0       4       0       0       0       1
##   [38,]       4       0       0       0       0       0       1
##   [39,]       5       0       6       0       0       0       1
##   [40,]      12       7       0       0       0       0       1
##   [41,]      12       6       7       0       0       0       1
##   [42,]       3       0       5       0       1       0       1
##   [43,]       6       3       0       0       0       0       1
##   [44,]       0       0       0       1       2       0       1
##   [45,]       0      12       0       3       1       2       1
##   [46,]       0       0       0       0       2       1       1
##   [47,]       5       0       0       3       7       0       1
##   [48,]       0       0       0       0       2       0       1
##   [49,]      12      10       1       2       0       1       1
##   [50,]       2       6       2       2       7       2       1
##   [51,]       6       5       4       2       5       0       1
##   [52,]      10       9       4       3       9       4       1
##   [53,]       7       1       5       1       0       0       1
##   [54,]       3       1       3       1       1       6       1
##   [55,]       0       4      10       1       0       0       1
##   [56,]       2       3       5       0       5       0       1
##   [57,]       1       3       1       2       0       0       1
##   [58,]       4      13       1       0       0       0       1
##   [59,]       8       3       0       0       0       0       1
##   [60,]       0       0       0       0       0       2       1
##   [61,]       0      11       0       0       0       0       1
```

```
## [62,]      7     12      0      0      0      0      1
## [63,]      8      4      0      0      0      2      1
## [64,]      5      4      0      0      0      0      1
## [65,]      6      7      0      0      0      1      1
## [66,]      1      0      0      0      0      0      1
## [67,]      3      5      1      0      0      0      1
## [68,]      0      0      0      0      0      0      1
## [69,]      2      0      4      0      0      0      1
## [70,]      2      3      6      0      0      0      1
## [71,]      1      0      3      0      0      0      1
## [72,]      0      0      0      0      0      0      1
## [73,]      0      0      0      0      0      0      1
## [74,]      0      0      0      0      0      0      1
## [75,]      0      0      0      0      0      0      1
## [76,]      0      0      0      0      0      0      1
## [77,]      0      0      0      0      0      0      1
## [78,]      0      0      0      0      0      0      1
## [79,]      0      0      0      0      0      0      1
## [80,]      0      0      0      0      0      0      1
## [81,]      0      0      0      0      0      0      1
## [82,]      0      0      0      0      0      0      1
## [83,]      0      0      0      0      0      0      1
## [84,]      0      0      0      0      0      0      1
## [85,]      0      0      0      0      0      0      1
## [86,]      0      0      0      1      0      0      1
## [87,]      0      0      0      0      0      0      1
## [88,]      0      0      0      0      1      1      1
## [89,]      0      0      0      3      0      4      1
## [90,]      0      0      0      0      0      0      1
## [91,]      1      5      0      0      0      0      1
## [92,]      0      0      0      0      0      0      1
## [93,]      0      0      0      0      0      0      1
## [94,]      0      0      0      0      2      0      1
## [95,]      0      0      0      0      0      0      1
## [96,]      2      0      0      2      1      0      1
## [97,]      3      0      0      0      1      0      1
## [98,]      7      0      0      0      0      0      1
## [99,]      2      0      0      0      0      0      1
## [100,]     2      0      0      0      6      0      1
## [101,]     0      0      0      0      0      0      1
## [102,]     0      0      0      0      0      0      1
## [103,]     0      0      0      0      0      0      1
## [104,]     0      0      0      0      0      0      1
## [105,]     0      0      0      0      0      0      1
## [106,]     0      0      0      0      0      0      1
## [107,]     0      0      0      0      0      0      1
## [108,]     0      0      0      0      0      0      1
## [109,]     0      0      0      0      0      0      1
## [110,]     0      0      0      0      0      0      1
## [111,]     0      0      0      0      0      0      1
## [112,]     0      0      0      0      0      0      1
## [113,]     0      0      0      0      0      0      1
## [114,]     0      0      0      0      0      0      1
```

To fit a `multivariate GLM`, we will be using the `manyglm` function from the `mvabund` package. Before we fit the model, we need to convert the community matrix into a data format that `mvabund` requires.

```
# Convert species community abundance matrix to 'mvabund' object
insect.spp <- mvabund::mvabund(com_matrix)
```

Fit the model   Much like the previous GLM's, we can specify the statistical distribution of the model and evaluate the fit of different statistical distributions. Below, we are going to fit a Poisson and then a negative binomial distribution to our data and see if either provide a good fit.

```
# Run global model with poisson distribution
mod_p <- manyglm(insect.spp ~
                # Fixed effects - allow disturbance effect to vary with season
                com_data_lts$disturb_allow_overwinter *
                com_data_lts$season,
             # Specify the distribution
             family = "poisson")

# Check model assumptions
plot(mod_p)
```

```
## Warning in default.plot.manyglm(x, res.type = res.type, which = which, caption =
## caption, : Only the first 7 colors will be used for plotting.
```
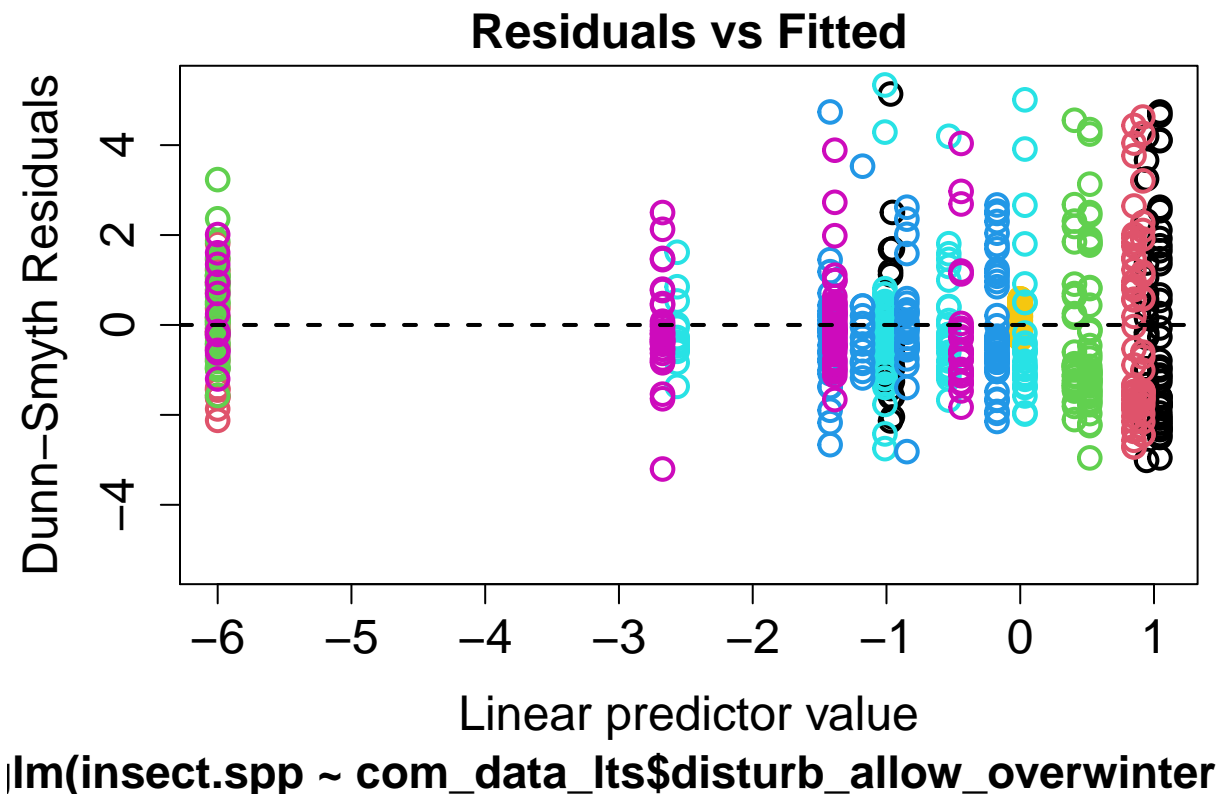
Poisson does not appear to provide a good fit - the range in the residuals gets wider as x increases. This tells us that there is unequal variances - as abundances get higher, so does variance.

```
# Run global model with negative binomial distribution
mod_nb <- manyglm(insect.spp ~
                  # Fixed effects - allow disturbance effect to vary with season
                  com_data_lts$disturb_allow_overwinter *
                  com_data_lts$season,
              # Specify the distribution
              family="negative_binomial")

# Check model assumptions
plot(mod_nb)
```

```
## Warning in default.plot.manyglm(x, res.type = res.type, which = which, caption =
## caption, : Only the first 7 colors will be used for plotting.
```
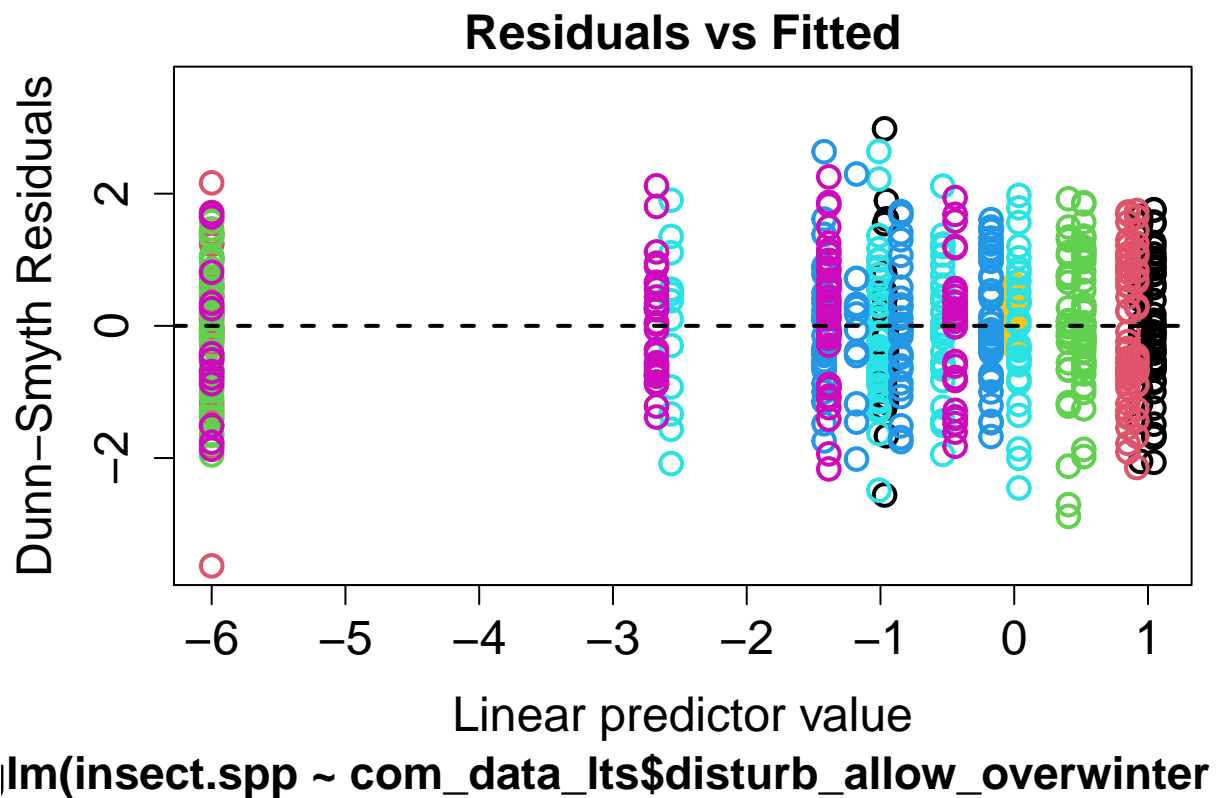

The negative binomial looks to be a much better fit. The range of the residuals is pretty consistent across x.

Perform global hypothesis test Here, we can test for statistical significance of `season` and `disturb_allow_overwinter`, and their interaction (does disturbance effect the insect community differently in summer vs winter). We again will use Likelihood Ratio Tests (LRT's) to test parameter significance.

```
# Test multivariate global hypothesis tests
anova(mod_nb,
      nBoot = 999,
      test = "LR")
```

```
## Time elapsed: 0 hr 1 min 6 sec
```

```
## Analysis of Deviance Table
##
## Model: insect.spp ~ com_data_lts$disturb_allow_overwinter * com_data_lts$season
##
## Multivariate test:
##                                                         Res.Df Df.diff    Dev
## (Intercept)                                                113
## com_data_lts$disturb_allow_overwinter                      112       1 104.84
## com_data_lts$season                                        111       1   7.79
## com_data_lts$disturb_allow_overwinter:com_data_lts$season  110       1   1.17
##                                                             Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                          0.001 ***
## com_data_lts$season                                            0.201
## com_data_lts$disturb_allow_overwinter:com_data_lts$season      0.901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Arguments:
##  Test statistics calculated assuming uncorrelated response (for faster computation)
##  P-value calculated using 999 iterations via PIT-trap resampling.
```

Insect community composition was significantly different disturbed vs undisturbed sites as indicated by the significant p-value for `disturb_allow_overwinter` ($X2 = 104.84$, $P < 0.001$). Neither `season` nor the interaction between season and disturbance had a significant effect on insect community composition.

Perform univariate hypothesis test Below, we can evaluate which of the individual insect species explain the shift in community composition between disturbed vs undisturbed sites.

```
anova(mod_nb,
      nBoot = 999,
      p.uni = "adjusted")
```

```
## Time elapsed: 0 hr 1 min 24 sec
```

```
## Analysis of Deviance Table
##
## Model: insect.spp ~ com_data_lts$disturb_allow_overwinter * com_data_lts$season
##
## Multivariate test:
##                                                         Res.Df Df.diff    Dev
## (Intercept)                                                113
## com_data_lts$disturb_allow_overwinter                      112       1 104.84
## com_data_lts$season                                        111       1   7.79
## com_data_lts$disturb_allow_overwinter:com_data_lts$season  110       1   1.17
```

```
##                                                            Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                       0.001 ***
## com_data_lts$season                                         0.234
## com_data_lts$disturb_allow_overwinter:com_data_lts$season   0.892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Univariate Tests:
##                                                           sp_tet1
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                      18.519    0.001
## com_data_lts$season                                         0.045    0.996
## com_data_lts$disturb_allow_overwinter:com_data_lts$season   0.009    0.919
##                                                           sp_tet2
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                      42.442    0.001
## com_data_lts$season                                         0.022    0.996
## com_data_lts$disturb_allow_overwinter:com_data_lts$season       0    1.000
##                                                           sp_bru1
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                      35.159    0.001
## com_data_lts$season                                         0.058    0.996
## com_data_lts$disturb_allow_overwinter:com_data_lts$season       0    0.929
##                                                           sp_sht1
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                       2.872    0.127
## com_data_lts$season                                         1.048    0.753
## com_data_lts$disturb_allow_overwinter:com_data_lts$season   0.141    0.919
##                                                           sp_chl1
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                       0.407    0.496
## com_data_lts$season                                         4.443    0.164
## com_data_lts$disturb_allow_overwinter:com_data_lts$season    0.47    0.916
##                                                           sp_eur4
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                       5.444    0.049
## com_data_lts$season                                         2.171    0.506
## com_data_lts$disturb_allow_overwinter:com_data_lts$season   0.548    0.916
##                                                           sp_dum1
##                                                               Dev Pr(>Dev)
## (Intercept)
## com_data_lts$disturb_allow_overwinter                           0    1.000
## com_data_lts$season                                             0    1.000
## com_data_lts$disturb_allow_overwinter:com_data_lts$season       0    1.000
## Arguments:
##  Test statistics calculated assuming uncorrelated response (for faster computation)
## P-value calculated using 999 iterations via PIT-trap resampling.
```

We see that sp_tet1 (X2 = 18.51, P = 0.001), sp_tet2 (X2 = 42.44, P = 0.001), sp_bru1 (X2 = 35.16, P = 0.001) and sp_eur4 (X2 = 5.44, P = 0.042) were significantly contributors to the observed shift in community composition between disturbed vs undisturbed sites. I.e. The abundances of these species were significantly different between disturbed vs undisturbed sites.