**Modern Regression Methods: homework 3**

Data for your homework #3 are in the *Lapwing* sheet of the *data\_mrm* Excel file. They represent results from 45 sites where the nesting of the wading bird called lapwing was recorded. Lapwings tend to nest in small nesting colonies to reduce predator impact. Each locality was characterised by the following three parameters:

(1) nesting presence, which interests us most, with levels "present" and "absent",

(2) presence of cereal fields in the grassland neighbourhood (none, winter or spring crop), and

(3) presence of shrubs on the grassland perimeter.

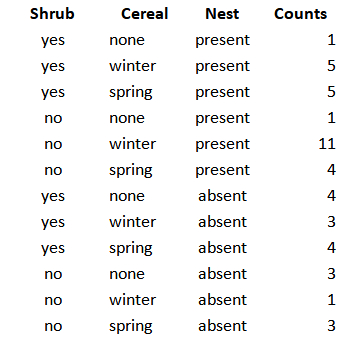
The research question we try to address is "does the nesting of this species depend on the presence of shrubs and/or the presence or type of crop field in the neighbourhood?"

This is, of course, a task suitable for the analysis of contingency tables with log-linear models, as I am happy to tell. Also, I note that you cannot use the table in its present layout, you must change it to be able to use it to fit a log-linear model.

Copy the commands you use to achieve the task in R, as well as their output. Present your results and conclusions drawn from them in a brief text, using full sentences ;-) This homework must be submitted before Friday midnight (23:59) to be scored.

**Does the number of nests depends on the microhabitat or the type of crop field ?**

****



*Figure 1: Table of contingency modified and ready to be used in GLM*

Since the data are counts data, we assume they are following a Poisson distribution. Indeed, in the plot below we can see that the distribution is right skewed.

**

A graph of a graph

Description automatically generated

*Figure 2: Distribution of the counts data*

We can check the assumption of the Poisson distribution which is that Dispersion should be equals to around 1.

A screenshot of a computer code

Description automatically generated

We checked and the ratio var/mean gave us the result of 1.945455. So the assumptions that Dispersion ~ 1 is met here.

We plotted the relationships between Counts and the different variables to have a first vizualisation of the data.

A close-up of text

Description automatically generated

A graph of different types of cereal

Description automatically generated with medium confidence

A graph with a number of lines

Description automatically generated with medium confidence

A graph of a graph with text

Description automatically generated with medium confidence

*Figure 3: Plots of Counts VS predictors*

**MODELISATION**

We start first with the null model : 

Then we try a model with first order interaction, adding all the variables separately.



A screenshot of a computer code

Description automatically generated

Here Nest:Cereal has a significant effect on the number of nests. The model quality of the null model is enhanced by adding the term corresponding to the type of crop field.

Then we decide to run a model with only Nest:Cereal since it is significant.

A close up of a text

Description automatically generated

Then we run the final model where we add the second order interactions:

A text on a white background

Description automatically generated

A screenshot of a computer

Description automatically generated

The Nest:Cereal has a significant effect on the number of nest counted with a p-value = 0.01062 \*.

**CONCLUSION**

The number of nests depends on the type of crop field: if there is winter crop field this is where the higher number of nests will be found, when nests are present.