Generalized linear mixed models

This work should be uploaded on Moodle before **Friday, March 10th, 5 pm**.

Numbers in parentheses indicate the number of points for each question.

Total: 10 points.

Data

The data for this exercise are from the paper:

> Marrotte, R. R., & Bowman, J. (2021). Seven decades of southern range dynamics of Canada lynx. Ecology and Evolution, 11(9), 4644–4655. doi: https://doi.org/10.1002/ece3.7364

We would like to verify if the southern lynx population in Ontario is sensitive to connectivity to the boreal forest and snow conditions. The table spatial_df.csv contains the following columns:

Hex: Sampling unit or Hexagon ID

$x =$ Easting in North America Lambert Conformal Conic of the sampling unit centroid

$y =$ Northing in North America Lambert Conformal Conic of the sampling unit centroid

region: Region of the southern range of the Canada lynx. West, Central, East

OccNb: Number of years where Canada lynx are harvested within a sampling unit

Dist_boreal: Distance to the boreal forest in km

Road_dist_km: Distance to the nearest major road in km.

Snow: Average Winter Severity Index

BobcatP: Number of years where bobcats are harvested within a sampling unit

CoyoteP: Number of years where coyotes are harvested within a sampling unit

```
data = read.csv("../donnees/spatial_df.csv")
str(data)
```

```
## 'data.frame':    65 obs. of  10 variables:
##  $ Hex         : chr  "H100" "H116" "H117" "H118" ...
##  $ x           : num  1299659 1215422 1271580 1327737 1383895 ...
##  $ y           : num  586392 635026 635026 635026 635026 ...
##  $ region      : chr  "East" "East" "East" "East" ...
##  $ OccNb       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Dist_boreal : num  332 244 276 304 339 ...
##  $ Road_dist_km: num  0.305 0.838 1.175 1.093 1.154 ...
##  $ Snow        : num  497 644 624 554 454 ...
##  $ BobcatP     : int  1 1 1 2 6 3 2 8 1 0 ...
##  $ CoyoteP     : int  41 60 49 52 51 42 45 45 43 49 ...
```

1. Variations in the number of years with successfull lynx harvesting

a. Data formatting. Transorm the distance to the boreal forest (*Dist_boreal*) from km to hundreds of km and normalize the Average Winter Severity Index (*Snow*) to have 0 mean and 1 standard deviation. **(1)**

b. We want to know if connectivity to the boreal forest and snow conditions impact the number of years where Canada lynx are harvested within a sampling unit (*OccNb*), controlling for differences among regions. We suppose that the sampling effort is identical in the sampling units and over different years. Which kind of model would you propose and why? **(1)**

c. Data exploration. Plot the *OccNb* as a function of *Dist_boreal* giving different colors to sampling units within the same region. You can use the option *theme(aspect.ratio = 1)*. Then, plot the *OccNb* as a function of *Snow* giving different colors to sampling units within the same region. Finally, group the data by region using *dplyr* and compute the mean and standard deviation of *OccNb*. We can now plot the mean *OccNb* per region and its variability (+ 1 SD) with a bar plot. **(1)**

d. With the full model (function of *Dist_boreal*, *Snow* and their interaction and with the random effect of the region on the intercept), we will test if the excess of zeros may cause some problem. We will use the function *glmmTMB* testing poisson and negative binomial models, with and without zero-inflated versions. We will evaluate the overdispersion of these models and justify the choice of the best model. **(1)**

e. Test the fit of different models according to AIC. Compare the final model of point *d* with other model versions: without random effect, without interaction, with single explanatory variables. Describe the best model. **(1)**

f. Test assumptions. Test for absence of overdispersion in the best model. **(1)**

g. Compute the probability of having zeros. **(1)**

h. According to the fixed effects section of the summary, define the mean number of years where Canada lynx are harvested within a sampling unit and the effect of *Dist_boreal* on this number. Ignore the probability of having zeros in the calculation. **(1)**

i. Calculates confidence intervals for each of the parameters of the mixed model. **(1)**

j. Plot the *OccNb* as a function of the distance to the boreal forest and then of *Snow*. In both plots, add a line showing the fitted values with the best model. Comment the results of the plots. **(1)**