

Exercise 1

Load the `loaloe` data in using the following lines of code

```
library(PrevMap)
data(loaloe)
```

To see the description of the variables enter

```
?loaloe
```

1. Consider the following Binomial model for the number of infected (`NO_INF`) using elevation (`ELEVATION`) as an explanatory variables.

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \beta_0 + \beta_1 d_i + \beta_3 \max\{d_i - c, 0\}$$

where d_i is the elevation variables and c is the value of elevation which corresponds to the change in slope of a linear spline. Use the function `glm`, to fit 3 models each specifying the following three different values for c : 650 (meters), 700 (meters) and 750 (meters). Which of these three models has a higher value of the likelihood and is therefore a better fit to the data? (HINT: Use the function `logLik` to obtain the likelihood value.)

2. Create a scatter plot of the empirical logit against elevation. Add to this plot the fitted linear spline from the best model identified in the previous point.
3. Consider another model, where elevation is introduced as a quadratic effect on the logit-linear scale, i.e.

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \beta_0 + \beta_1 d_i + \beta_3 d_i^2.$$

Fit this model in R and, based on this, provide an estimate of the elevation at which prevalence reaches its maximum.

4. Add the quadratic fit from the previous point to the plot generated in 2. Based on the residual deviance which of the two model is a better fit to the data?
5. Consider the variable mean NDVI (`MEAN9901`). What relationship do you observe between this variable and the empirical logit? Can we assume this to be linear?
6. Consider a Binomial mixed model which uses mean NDVI (`MEAN9901`) as a lineal covariate, i.e.

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \beta_0 + \beta_1 d_i + Z_i$$

where Z_i is a set of independent and identically distributed Gaussian random variables. Using the function `glmer` in the `lme4` package provide an estimate of σ^2 . Is there evidence of overdispersion from this model?

7. Carry out the test for spatial independence based on the empirical variogram. Is there evidence of residual spatial correlation in the data?