Ex1)

1. The idea is to build a LDA or QDA classified depending on the verification of the hypothesis.

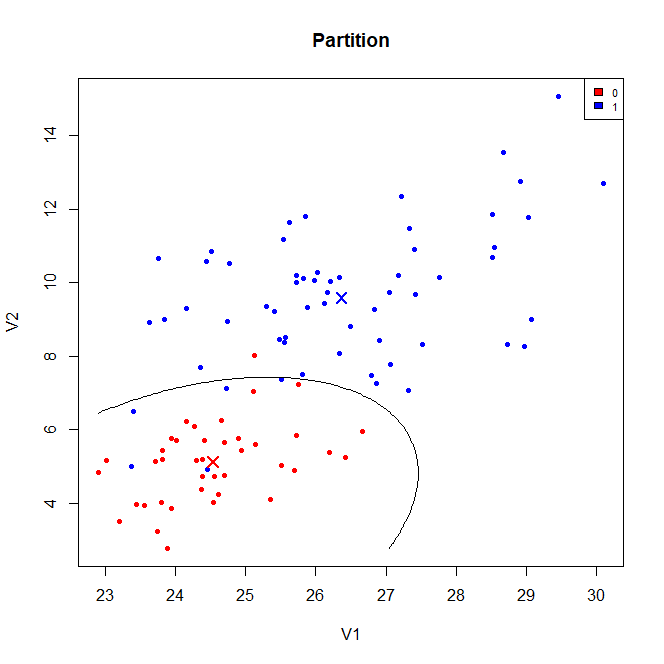
Both the classifier need the assumption of gaussianity in each group so we perform a mcshapiro test obtaining as pvalue 0.8660 and 0.4128 so we can assume gaussianity.

Then check if the 2 groups have same covariance structure (if yes go with LDA otherwise use QDA)

Both qualitatively and visually the homogeneity assumption does not seem to be met

* Build a QDA classifier

Plot of the classification region:



Parameters ?

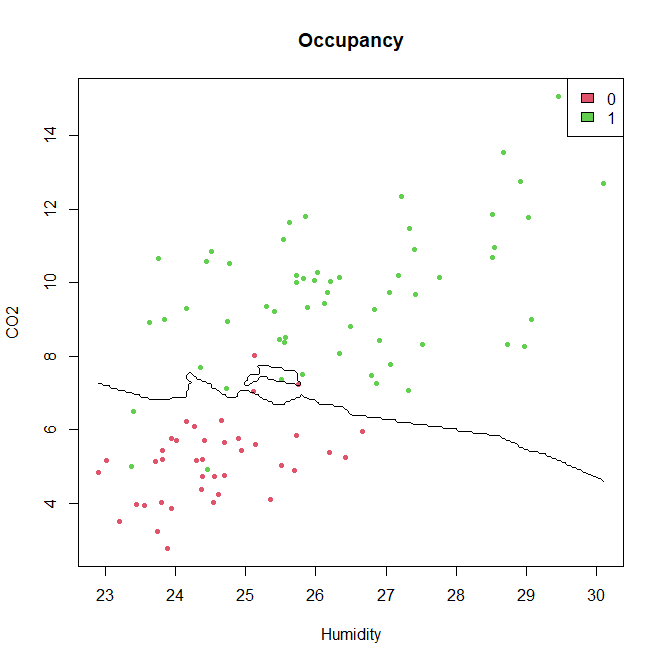
1. APER = 5/60\*(9/24) + 1/40 \* (15/24) = 0.046875
2. The posterior probability are

0 1

[1,] 0.008318664 0.9916813

* Classify as 1

1. KNN



APER = 0.06

So we can conclude that the error rate performed on the training set is a bit lower in the case of QDA so in this case the more flexibility given by KNN does not seem to be needed.

Ex2)

1. Non fatto
2. Using a BH correction we find we find that the following purity parameters:

2 , 3 , 5 , 7, 9

1. Here we want to bound the FWER so we use a Bonf correction with alpha = 1% we find only the purtity parameter -> guarda codice

Ex3)

1. Model:

# tox = beta\_0 + beta\_1 \* C1 + beta\_2 \* C2 + beta\_3 \* C3 + beta\_4 \* C4 + beta\_5 \* C5 +

beta\_6 \* C6 + Eps

# with Eps ~ N(0, sigma^2)

Estimates of the parameters:

* Beta:

B0 30.15705

b1 0.14915

b2 0.65670

b3 0.72840

b4 -7.04633

b5 0.47276

b6 5.9804

* Sigma^2 = 12.87133

Verify assumption:

* Residuals homoschedastic and centered in 0:

From the plot we can assume this hypothesis

* Residuals normal:

Shapiro test with pvalue = 0.186 -> ok

1. Pointiwise estimate = 37.40647

PI = 29.93171; 44.88122

1. The optimal lambda according to the CV is lambda = 0.2056512

And the significant coefficients are:

B0 = 30.5413458

B1 = 0.1382231

B3 = 0.6166355

B4 = -6.5115196

B6 = 5.6695527

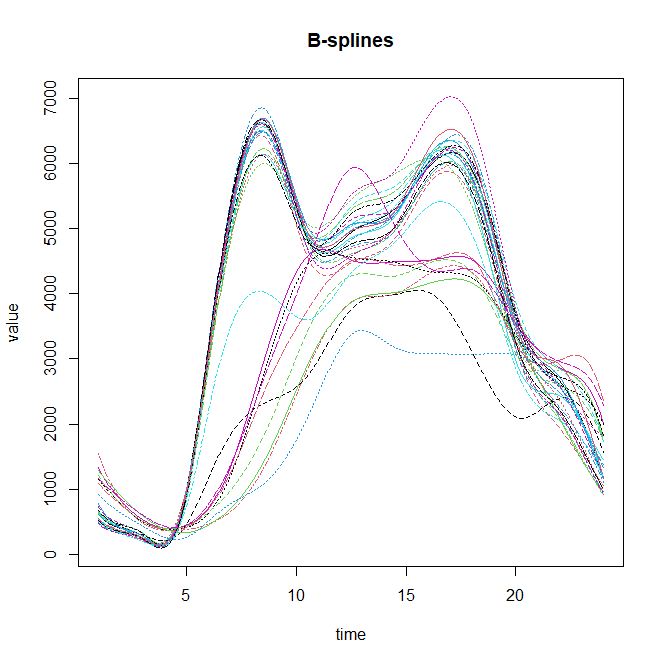
Instead the others are put to 0

1. Pointiwise prediction = 36.63287

PI = ?

Ex4)

1. Smoothed data



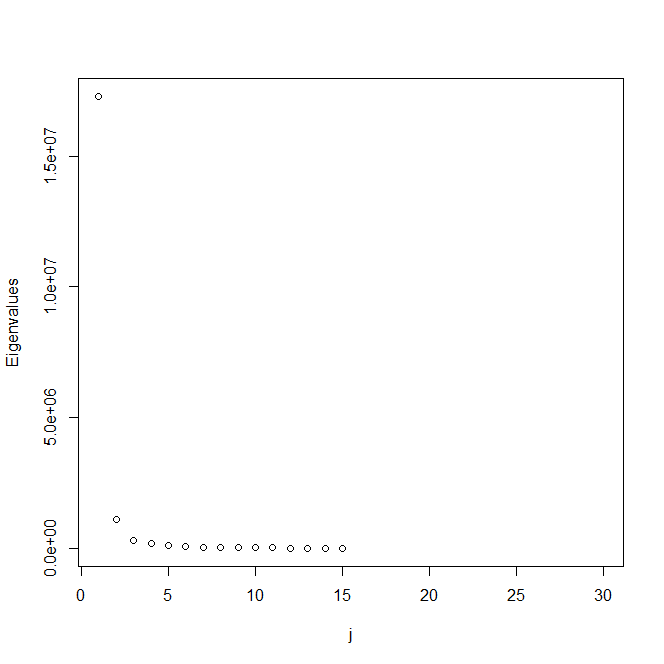
First 3 coeff for Day1: 616.5454 ; 354.9235 ; 463.4883

1. Variance explained by the first FPC : 0.90717863

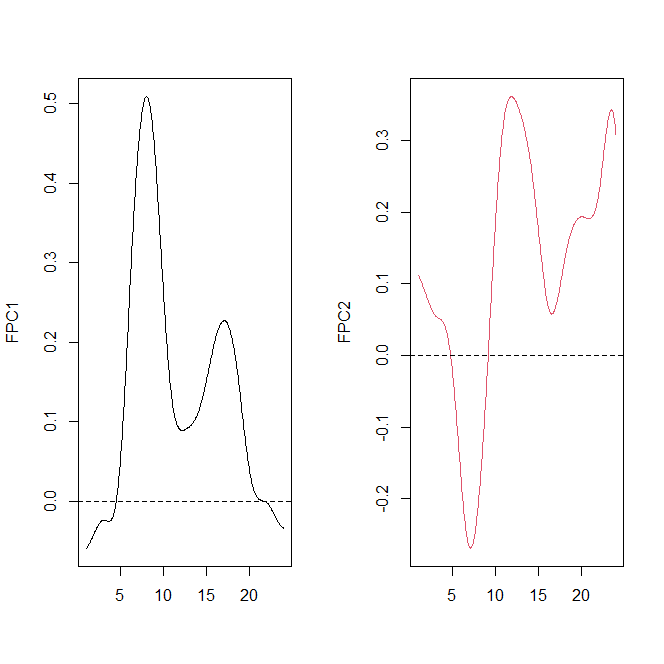
Variance explained by the second FPC: 0.05751325

Variance explained by the third FPC: 0.01483253

Scree-plot



Plot of the first 2 FPC = first 2 eigenfucntions



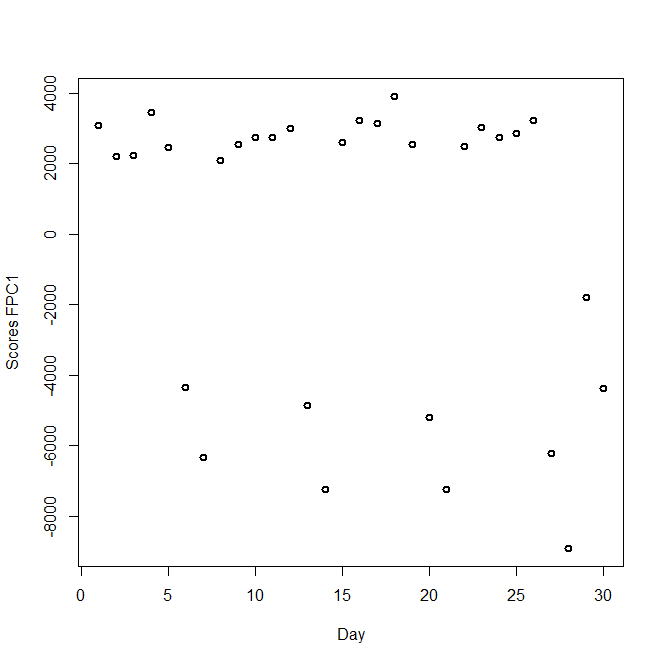
1. Since with the first 2 FPC we are already explaining more than 95% of the variability we can perform dimensionality reduction keeping only those 2.

Plotting them as perturbation of the mean we see that:

* FPC 1 shows a difference in amplitude between those function with positive scores and those with negative ones , in particular the ones with positive scores have a larger value of toxicity wrt the mean -> functions with high value of FPC1 have higher toxicity during the day instead lower level during the night
* FPC 2 shows a contrast? Non mi convince perchè non è vero che si alternano

Days with low value of FPC2 have 2 picks of toxicity during the day instead days with high value of FPC2 have a more or less monotone behaviour

1. Plot



From the plot we can observe that in some days we have positive score for the FPC1 and in others negative meaning that in some days the value of toxicity is higher then the mean one and in others it is lower