Ex1)

1. Verify the normality assumption to perform the pest -> the pvalue of the mcshapiro test is around 20% -> we can assume gaussianity

Perform the test:

H0: mu == mu0 vs H1: mu != mu0

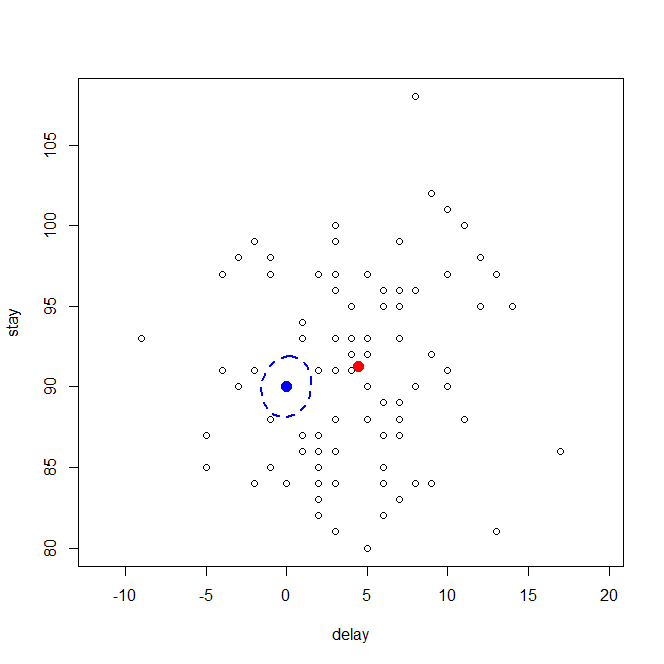
With mu0 = (0,90)

The value of the test statistic is T2 = 80.99549 (computed as

n \* (x.mean-mu0) %\*% x.invcov %\*% (x.mean-mu0), where x.mean is the mean, x.invconv is the inverse of the covariance and n is the number of samples)

At level 99% we reject H0 so we deny that in mean the clients are on time and stay 90’.

We can plot the rejection region



The rejection region is the area outside the blue ellipse and the red dot is the sample mean which is in the rejection region, indeed we reject H0.

The center of the rejection region is mu0=(0,90).

PICCOLA?!

Espressione della rejection region:

{ x \in R^p t.c. n \* (x - m0)' %\*% x.invcov %\*% (x - m0) > cfr.fisher }

1. Pvalue = 4.304335e-13 so we reject at any reasonable level
2. Bonferroni intervals for the components of the mean:

inf center sup

delay 3.004227 4.433333 5.86244

stay 89.491055 91.233333 92.97561

so we can see that 0 is not included in the Bonf intervals for the delay and neither 90 in the one for the stay and so the Bonf intervals are appropriate to support our conclusion (we reject).

1. Performing a test for a linear combination of the mean with a = (0,1) so we extract only the mean of the stay we check if it is = 90 or not

H0: mean(stay) == 90

H1: != 90

Al level 90% we assume H0 so the policy is correct.

ERROR: VUOLE LA SOMMA

Forse serve anche mettere a 0 i delay che sono negativi

Ex2)

1. We have to choose between a LDA and a QDA classifier, to do so verify the assumptions:

Both the models rely of the gaussianity assumption -> performing the mcshapiro test we obtain a pvalue = 0.9328 0.5764 in the 2 groups so we can assume gaussianity.

Then to be allowed to use LDA weh have to verify the hyp of homogeneity in the 2 groups but computing the covariances matrix this assumption does not seem to be met so we choose to rely on a QDA classifier.

Covariance matrix in the group normal

incidence tilt

incidence 109.63118 39.74931

tilt 39.74931 41.38297

Covariance matrix in the group abnormal

incidence tilt

incidence 314.53723 95.15106

tilt 95.15106 103.33664

Prior probabilities of groups:

NO AB

0.65 0.35

Group means:

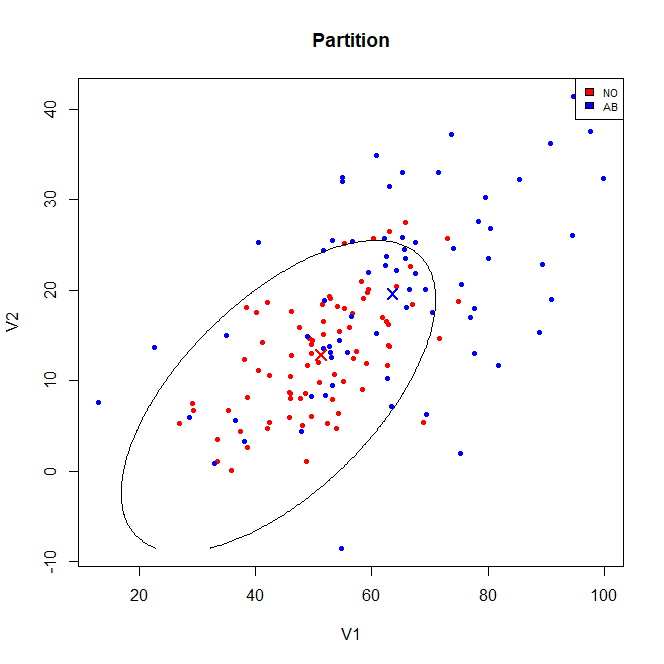
incidence tilt

NO 51.26018 12.89092

AB 63.56038 19.58847

Covariances dei gruppi sputati da qda fatti ora sul template

Plot the partition:



Inside the plotted region we classify normal and outside abnormal

1. 8/80 \* 0.65 + 32/70 \* 0.35 = 0.225

Where I took the value from the table

class.assignedCV

class.true NO AB

NO 72 8

AB 32 38

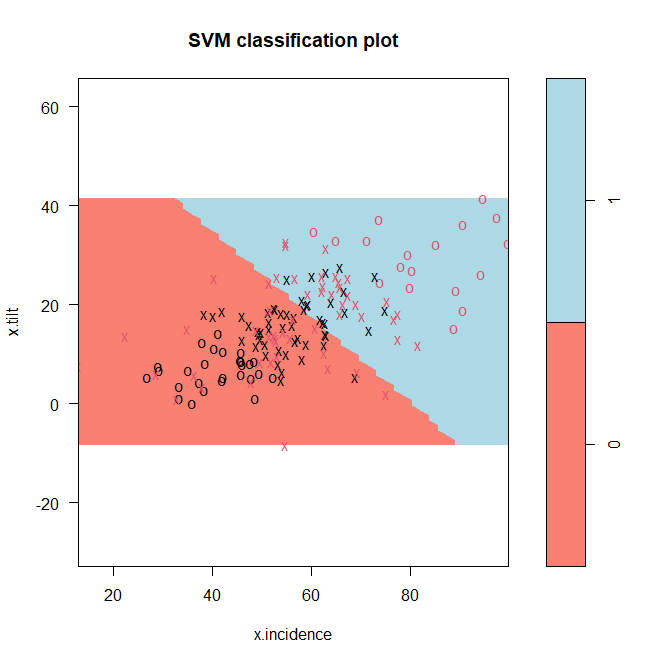
1. Estimated posterior probability of the new observation:

NO AB

[1,] 0.2392177 0.7607823

* Classify as abnormal

1. Use SVM



Where 0 = Normal and 1= Abnormal

The support vectors are the one indicated by the cross and we see that there are some classification error since we are applying the soft version of SVM.

The patient at the previous would be classified as normal with this classifier.

Ex3)

1. To fit the requested model we create 2 dummy variables to encode the variable OS

Dummy\_mac = 1 if OS=mac

Dummy\_linux = 1 if OS=linux

Fir the model with interaction of both the dummy of both the numerical regressors but we don’t include the dummies ‘alone’ so that alpha will not depend on g.

Estimates parameters:

(Intercept) freq cache\_acc

Mac 1345.657 121.21286 -4.238733

Linux 1345.657 34.62254 -2.588699

Windows 1345.657 76.20252 -9.002267

Where the column of intercept indicates alpha, the column of freq beta\_g and the last gamma\_j

Estimate for sigma^2 = 11497.83

1. The assumptions of residual omoschedastic and centered in 0 seems to be met and also the normality since the pvalue of the Shapiro is 0.15 so we can accept gaussianity.

Perform the test checking if all the coefficients related to the factor OS can be put at 0 through a linear hypothesis test

In H0 we have all the coeff about the dummies = 0 and in H1 that there exists at least one !=0:

Pvalue = 1.903e-08 -> reject

* The factor OS has a significant impact

1. Perfom a test to check if we we can remove cache, so putting at 0 all his coeff and we can accept H0 so we remove cache (pvalue 0.36)

* Cache\_acc has not a significant impact on the mean of the price

1. Fitting the model without cache we see that freq:dummy\_linux is not significant so we remove it arriving at the final model:

price ~ freq + dummy\_mac:freq

Estimates of the parameters:

(Intercept) freq

Mac 1273.335 123.75317

Linux 1273.335 53.08275

Windows 1273.335 53.08275

Estimates of sigma^2 = 11762.65

1. CI = [ 1396.313 ,1490.087]

Ex4)

1. We are assuming that the mean is constant, so we don’t have any regressors and we are under the hyp of second- order stationarity

The model estimated for delta is given by:

model psill range

Sph 42.77012 154.6641

The estimated coeff a0 = 12.18036

1. Now introduce as regressor the colour -> the model is non stationary and the mean is not anymore constant

Model estimated for delta:

model psill range

1. Sph 4.951555 1547.267

Coefficients:

red = 9.2311 , yellow = 24.22476, orange = 9.007159

1. The second model seems more appropriate since in the first one the sample variogram seems to be almost constant (so it is like there was no spatial dependence) instead in the second one we have the classical shape of a variogram linear in the origin which then lay on an asympthot.

From the plot to compare the 2 variograms the fit is large so it seems to be significant -> choose the non-stationary model

1. Predictions:

Red\_week = 10.48128

Orange\_week = 10.25734

Yellow\_week = 25.47494