Ex1)

1. We assume the 2 population in the 2 cities to be independent.

To perform the test we need the following hyp

# - independent population -> ok

# - gaussian distribution -> perform a mcshapiro test in each population:

The pvalues are quite high so we can assume gaussianity

# - same covariance structure -> comparing qualitately the 2 covariance matrixes we assume same cov structure

We can perform the test:

Test H0: mu1-mu2 == c(0,0) vs H1: mu1-mu2 != c(0,0)

The pvalue is 0.0004434524 -> at level 99% we reject H0 and so we can state that there is difference in the mean evaluations in the 2 cities

1. Bonf intervals for the difference of the mean evaluation given by T1 in the 2 cities and the difference of the mean evaluation given by T2 in the 2 cities

T1 0.2135749 0.9956667 1.777758

T2 -0.5243082 0.2596667 1.043642

We see that the CI related to T1 does not include the 0 which means that the difference of the mean evaluation given by T1 is statistically different from 0.

Instead using Bonf we cannot say the same for T2.

But this does not contradict the test of point a since to reject H0 it is enough to rejet H0 in some direction of R2.

1. We construct new 2 populations accounting for the 2 cities, each of them with only one variable which is the average evaluation of the 30 dishes.

Performing the test

Mu1 = media pop1 (acoruna)

Mu2 = media pop2 (pontevedra)

Test H0: mu1-mu2 <= 0 vs H1: mu1-mu2 >0

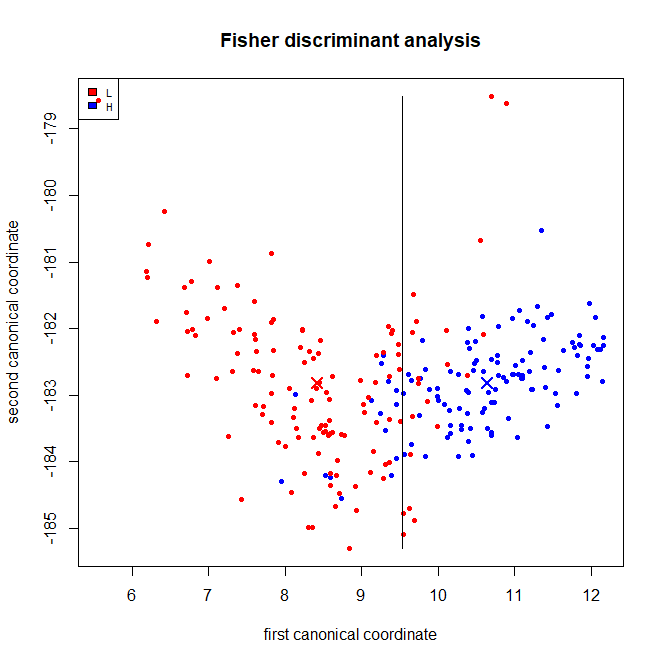
We obtain a pvalue of 0.01312327 so we assume H0 at level 99% and so there is not statistical evidence to state that the average evaluation of dishes in acoruna is higher than the one in Pontevedra.

Ex2)

1. We try to verify the assumption of normality in the 2 groups in order to buil a LDA or QDA classifier but we cannot assume gaussianity since the pvalue of the mcshapiro test in the group L is equal to 0 so at any level we reject the gaussianity assumption.

So we move to the Fisher DA approach, we need to verify the assumption of same covariance structure and comparing them both qualitatively and graphically we deduce than we can assume it.

Plot:



Parameters in the group1 (= group L ???) -> dalla table penso di sì

* Mean

x y

12.27637 39.56299

S1:

x y

x 0.3213866 -0.1357833

y -0.1357833 0.1082292

Parameters in group2 = groupH

* Mean

x y

12.27637 39.56299

S2:

x y

x 0.3213866 -0.1357833

y -0.1357833 0.1082292

Prior stimati empiricamente?

pL = 130/250 = 0.52

pH = 120/250 = 0.48

1. APER = 35/250 = 0.14 -> AER?

Weakness : linear?

1. ???

Ex3)

1. Parameters:

Alpha beta gamma

handmade 11.81774 15.737418 3.553972

machine 16.68362 8.487567 2.561518

sigma^2 = 68.20368

1. Assumption : Eps ~ N(0, sigma^2)

* Plotting the residuals against the fitted values we can assume that they are centerd in 0 and homoschedastic
* Performing a Shapiro test on the residual we obtain a pvalue of 0.2608 so we can assume normality

Write the model in the following way

# y = beta0 + beta1\*dimension + beta2\*ncolors +

# + beta3\*dummy + beta4\*dummy\*dimension + beta5\*dummy\*ncolor + eps

Dummy = 1 if methos = handmade, 0 otherwise

So the test can be formulated as:

H0: (beta3, beta4, beta5) == (0, 0,0) vs H1: (beta3, beta4, beta5) != (0,0, 0)

We obtain a pvalue equal to the numerical 0 -> we reject H0 and we can state that the method as a significant impact on the price of the tattoo

1. H0: (beta2, beta5) == (0, 0) vs H1: (beta2, beta5) != (0, 0)

We obtain a pvalue equal to 2.916e-11 -> we reject H0 and we can state that the ncolors has a significant impact on the price of the tattoo

1. Remove the dummy since it does not seem to be significant and fit a new model. Remove the interaction between the dummy and the ncolors since it is not significant and reach the final model

# y = beta0 + beta1\*dimension + beta2\*ncolors +

# + beta4\*dummy\*dimension + eps

Parameters:

* Alpha beta gamma

handmade 13.89011 15.633728 3.075839

machine 13.89011 8.621341 3.075839

* Sigma^ 2 = 68.02003

1. Recalling that using the last model alphag = alpha , in order to obtain a global level of 95% we include the bonf correction computing the CI replacing alpha with alpha/2

* Mean fixed cost for a tattoo = 4.736621 ;23.04359
* CI for the new obs = 116.0581 121.1122

Ex4)

1. Station1: 467.62612432 , -95.75818786 , - 45.39373299

Station2: 493.09532322, -90.15895244, - 40.70577599

1. Variance explained by the first 5 PC:

PC1: 0.832743483

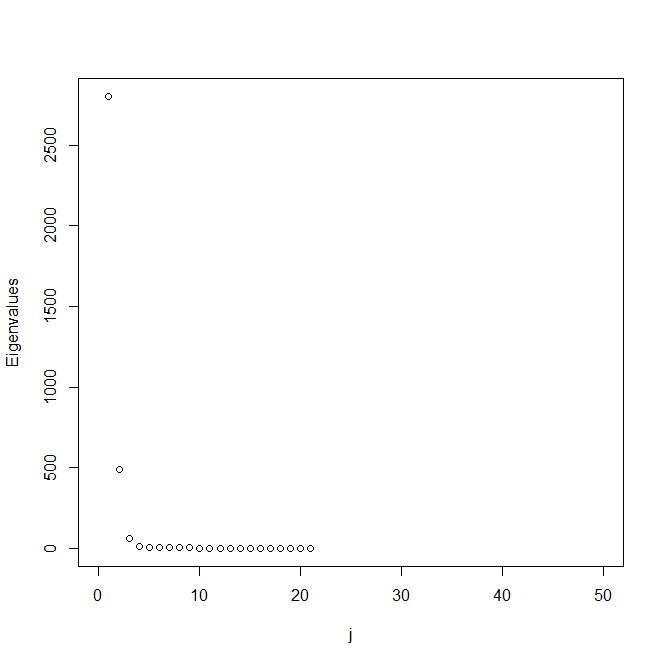
PC2: 0.144954229

PC3: 0.018432630

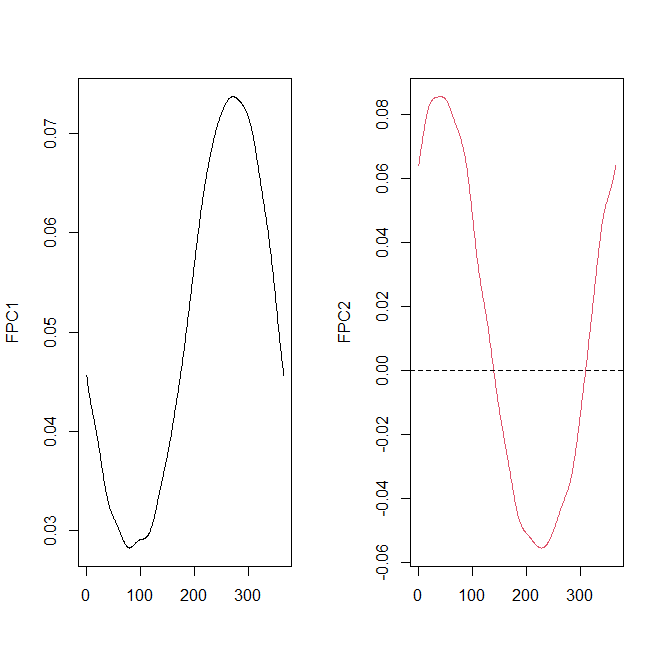
PC4: 0.002125033

PC5: 0.001028362

Scree-plot:



3 eigenfunctions:



Plotting the PC as perturbation of the mean we can inetrprete them:

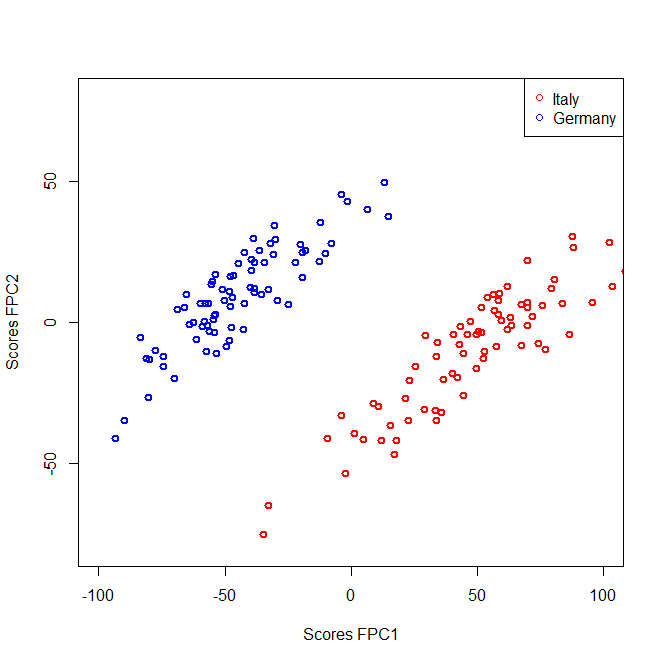
FPC1 shows a difference in amplitude between function with positive and negative values along FPC1, in particular the ones with positive values have a higher temperature.

* High Pc1 means higher temperature

FPC2 shows a contrast between the central part of the year so spring and summer and the extreme part so winter and autumn, functions with low value of pc2 are characterized by a more rigid winter and autumn but a warmer spring and summer, instead for functions with high value the winter and the autumn are less rigid and the spring and summer less warm. So more mith

FPC3 shows a little delay especially in spring for functions withhigh scores, instead the one with low scoresare a bit earlier.

1. Scatterplot



We can observe that the measurements in Italy have almost all a FPC1 score >0 which means that they have higher temperature than the mean and so than the measurements in Germany which are instead characterized by negative values for FPC1 And so lower temperature.

1. Looking at the explained proportion of variance we see that with 2 FPC we are already explaining among the 95% of the variability and adding also the third one we go beyond 99%.

Since also the interpretation of FPC3 is meaningful we can keep the first 2 FPC in order to perform dimensionality reduction.

RESULTS? -> forse con 2 cluster ben distinti tenendo 2 FPC significa che possiamo tenere 2 FPC per la dim reduction