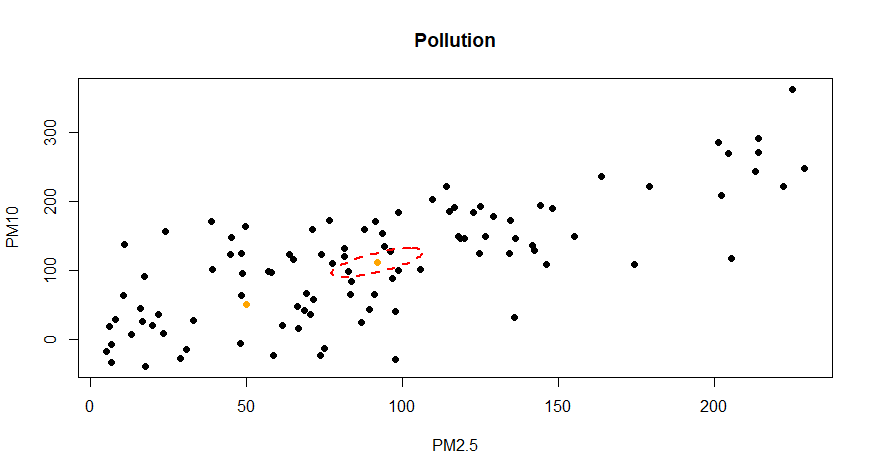
**Exam of Applied Statistics 15/06/2020**

**Exercise 1**

First of all we check that the datas belong to a bivariate normal distribution, to do so we perform a Shapiro test on the whole data obtaining a p-value of 0.2284, so we decide to accept the null hypothesis that datas come from a bivariate normal distribution. We proceed to perform the test calculating the T2 statistic and the quantile of the Fisher distribution with alpha = 0.05 multiplied by (n-1)p/(n-p), which was equal to 6.2414, it was way lower than the T2 (59.68), and the pvalue is very very close to 0 (9.131362e-11):

we cannot accept the hypothesis that the mean is (50,50).



Here we can see the plot of the data, in particular we have ad orange point at (50, 50) and another (with the ellipse) at the sample mean of (92.1, 111.4). We can see the confidence ellipse of level 95% for the mean in red: its radius is 2.49829, the two semi-axes have length of 24.268852 and 8.575489. The two directions are respectively (0.5256937, 0.8506739) and (-0.8506739, 0.5256937). The point (50,50) is clearly outside the confidence ellipse for the mean. This leads us to conclude again that it is not the true mean.

The confidence region is defined as:

E = { m in R^2 | n ( x\_ - m ) ‘ S^(-1) (x\_ - m ) < F (0.95, 2, 98) \* (n-1) \* p / (n – p )}

Where x\_ it the sample mean, S the sample covariance matrix, F is the quantile of the Fisher with confidence of 0.95 and the degrees of freedom p = 2 and n – p = 98 (n = 100 number of observations, p = 2 number of variables).

Keeping the 95% of global confidence we proceed to compute the simultaneous T2 intervals for the two means:

Lower Center Upper

PM2.5 : 77.37913 92.07547 106.7718

L C U

PM10 : 90.31843 111.4498 132.5811

50 is not even in any of the confidence intervals for the means of the two values, since the have lower bounds respectively of 77.37913 and 90.31843. This is an additional confirmation that the true mean is not (50,50).