Exam: 2021/06/18

Marco Scarpelli

06 giugno, 2024

Dataset exploration

```
##     PM2.5    PM10
## 1 218.553 212.320
## 2 263.142 260.849
## 3 289.445 157.020
## 4 132.188 163.843
## 5 142.050 136.461
## 6 167.304 103.708
```

Point A

We verify that the data is multivariate Gaussian:

```
##            Test        HZ  p value MVN
## 1 Henze-Zirkler 0.3895568 0.832463 YES
```

The data is Gaussian. Let verify the hypothesis:

```
##       [,1]
## [1,] FALSE
```

```
##      [,1]
## [1,]    0
```

This means that the mean is significantly different from (50, 50).

Point B

The expression for this ellipse is:

$$\left\{ m \in R^2 \middle| n \left(\overline{X} - m\right)^T \mathcal{S}^{-1} \left(\overline{X} - m\right) < F^* \right\}$$

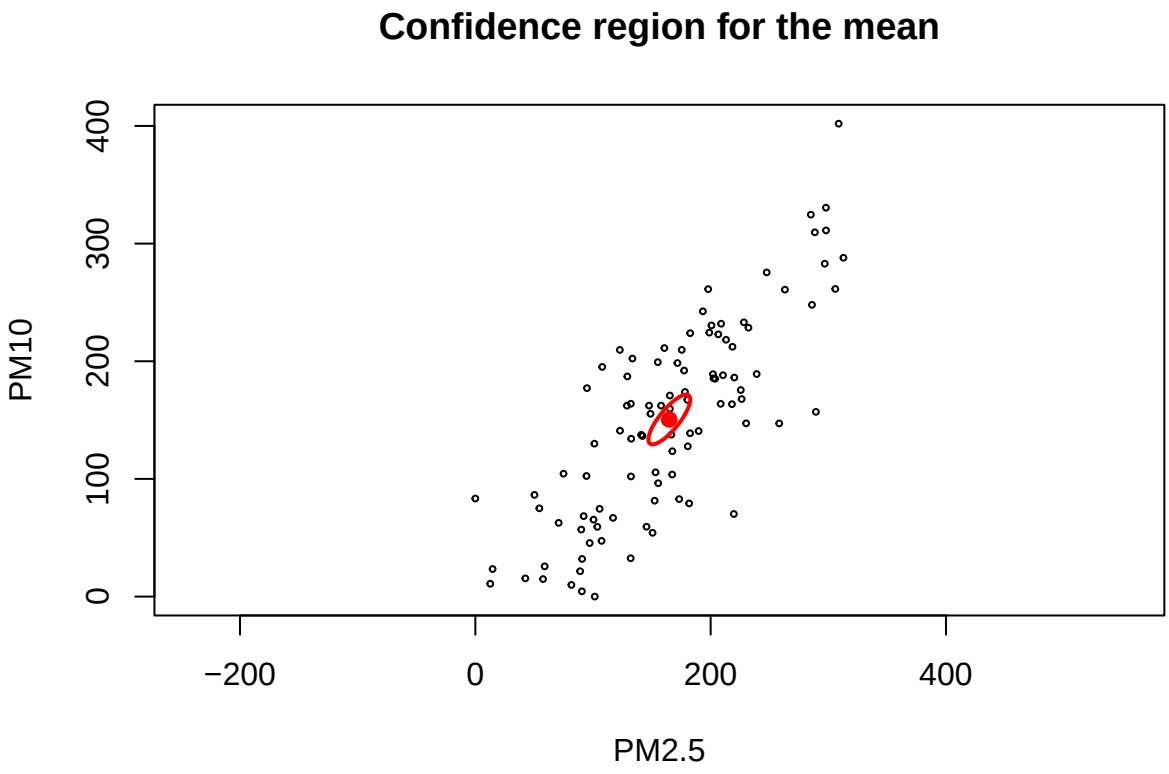
with F^* equal to:

$$\frac{(n-1)p}{n-p} F(1-\alpha, p, n-p)$$

We all requested data:

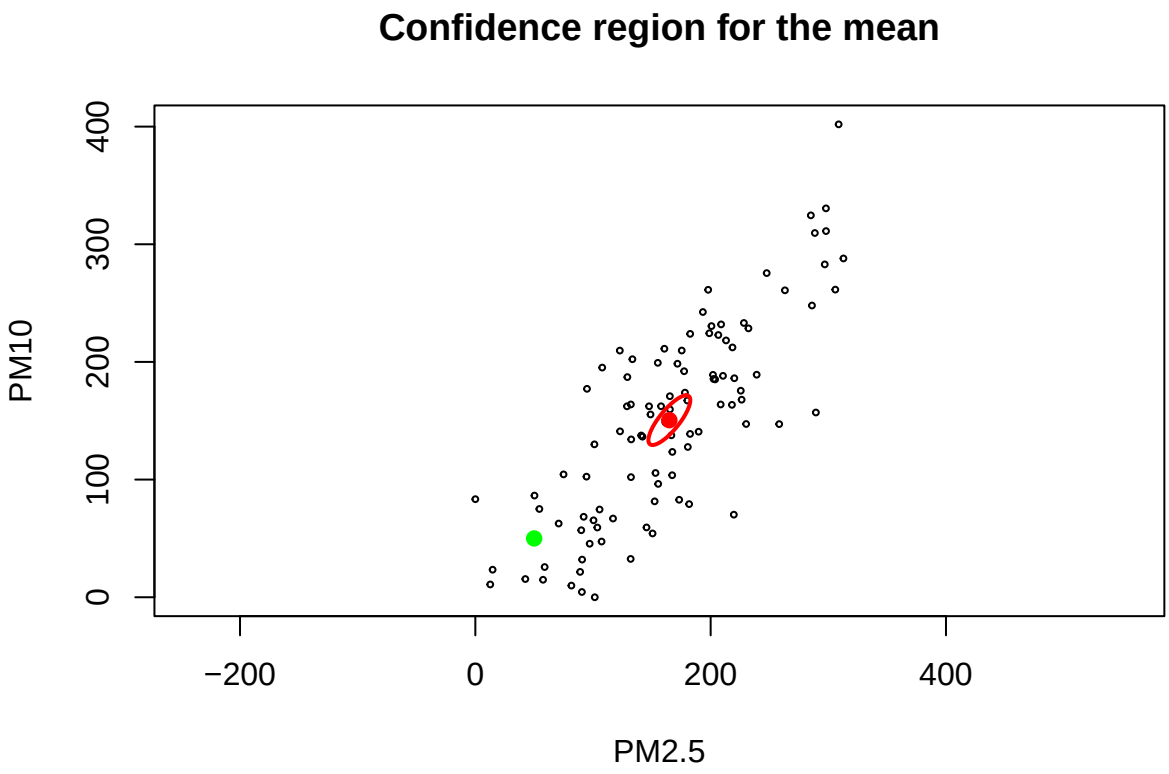
```
##        mean.PM2.5         mean.PM10 semi.axes.length1 semi.axes.length2
##        164.900470        150.484780         26.295269          8.525918
##            radius
##          2.498290
```

We also plot the confidence region:



Plot C

We plot the confidence region and highlight the point at (50, 50) in green:

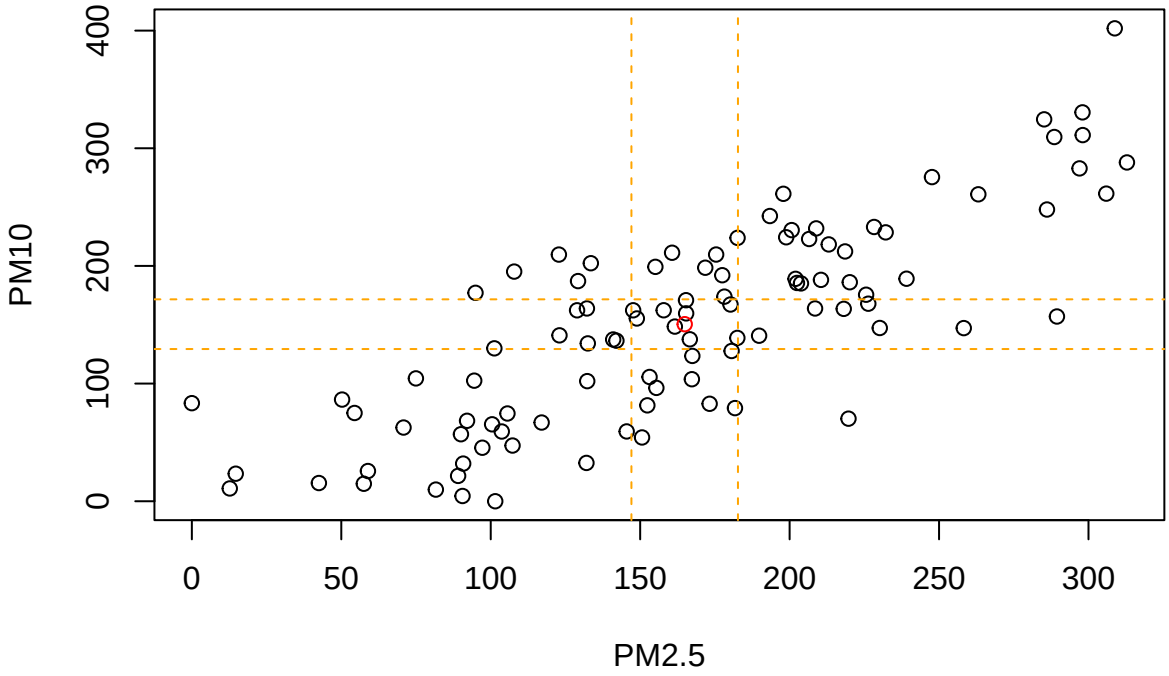


We can see that the point lies well outside the confidence region, in accordance with our statistical test.

Plot D

Confidence intervals, with plot:

```
##            inf   center      sup
## PM2.5 147.0792 164.9005 182.7218
## PM10  129.3534 150.4848 171.6161
```


PM10
0
100
200
300
400
0
50
100
150
200
250
300
PM2.5