Covid-19 deaths in Iran

A statics review

Radmehr Karimian
Dr.Mirsadeghi
Alireza Kodivar
May 4, 2023

First, our method is based on linear regression, so we limit ourselves to only using linear regression for this work.

For example, consider this:

We want to show that seasons and deaths aren't independent, so we have:

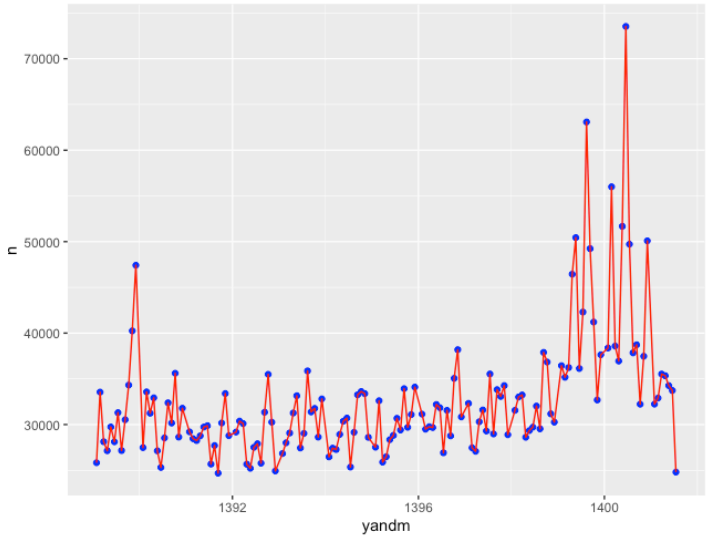

Figure 1: deaths in Iran based on months

As you can see, in the second half of the year, we call it Winter. The deaths are more than in Summer because the Contagious disease is more likely in the winter and cold days, but as we said, we ignore that part.

The second thing is that we shouldn't consider the ancient days and years in our estimation because the parameters and influences changed a lot through the years; for example, in 1389, Swine influenza caused deaths, and it's none of covid business.

So the question is, when should we consider deaths?

Based on our try-and-find method, it's better to consider from 5 years before:

You can watch trends in 5 years by your eye in the plots from Tehran in the other months:

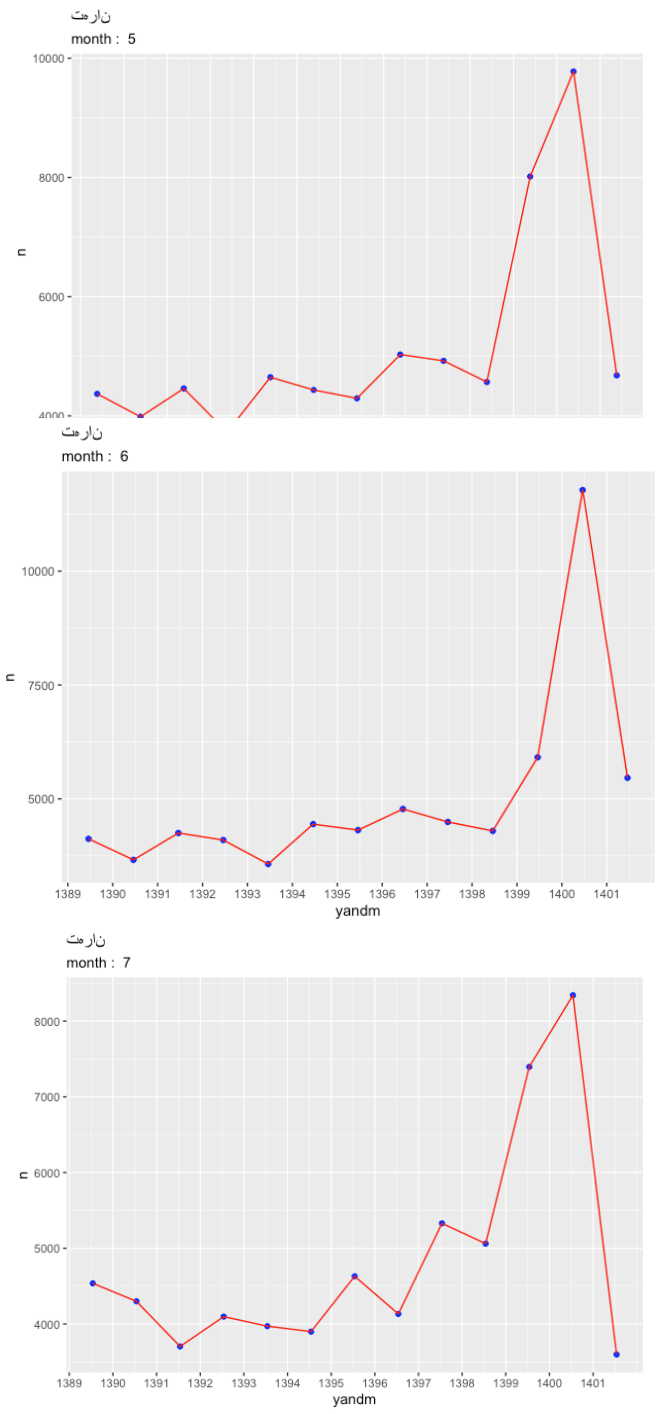

Fig 2: deaths in Tehran in different years based on months

On the other hand, we have to consider our main parameters for linear regression:

We can use Sex as a boolean parameter, but we prefer not to think that. Moreover, we decided not to consider age groups because it doesn't look that

different. We also do our regression month by month because we know that covid has a pick and waves through time.

Moreover, we should set a limit for our p-value. If our p-value is more than that, we consider our regression as the mean of other data.

We set that limit to 0.5:

```
Call:
lm(formula = n ~ yandm, data = data2fit)

Residuals:
     1      2      3      4
 299.5 -740.0  581.5 -141.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15633.3   440505.6   0.035    0.975
yandm           -7.5      315.5  -0.024    0.983

Residual standard error: 705.5 on 2 degrees of freedom
Multiple R-squared:  0.0002825, Adjusted R-squared:  -0.4996
F-statistic: 0.0005651 on 1 and 2 DF,  p-value: 0.9832
```

Fig 3: very high p-value for West Azerbaijan in Ordibehest

```
Call:
lm(formula = n ~ yandm, data = data2fit)

Residuals:
     1      2      3      4      5
-223.6  523.1 -628.2  581.5 -252.8

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -140525.5   269726.4  -0.521    0.638
yandm           104.3      193.2   0.540    0.627

Residual standard error: 611.1 on 3 degrees of freedom
Multiple R-squared:  0.08851,   Adjusted R-squared:  -0.2153
F-statistic: 0.2913 on 1 and 3 DF,  p-value: 0.6269
```

Fig 4: very high p-value for East-Azerbaijan in Ordibehest

So here we consider the mean of other data for that,

We also believe the start of the covid-19 in Iran was in 1398/10.

After all this implication, we find that the total number of deaths in Iran is **269911**.

We also bring deaths in different provinces in prov_ration file.

We draw the heatmap too:

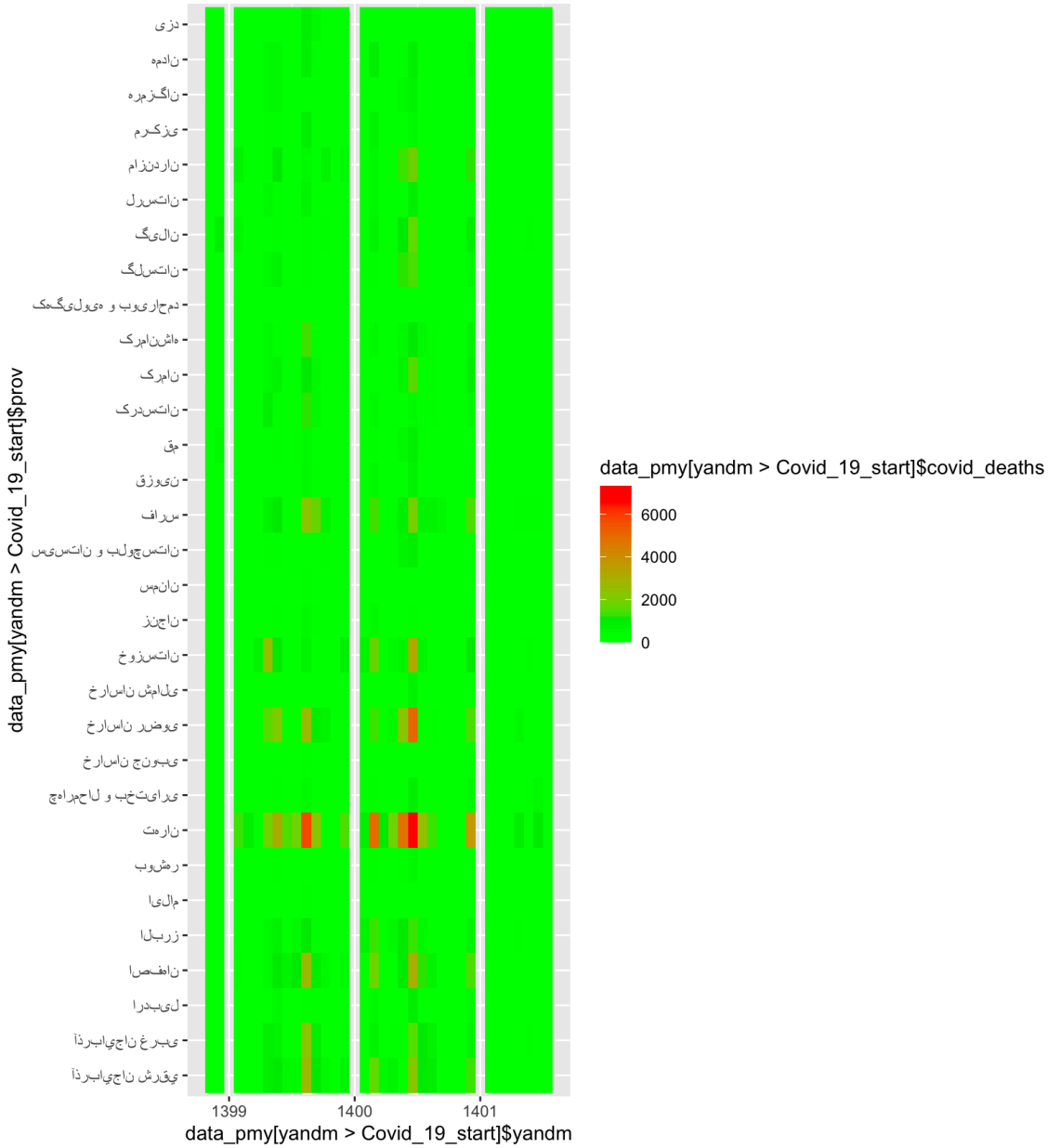

Fig 5: heatmap

Also, we bring this; you can watch others by running the R code:

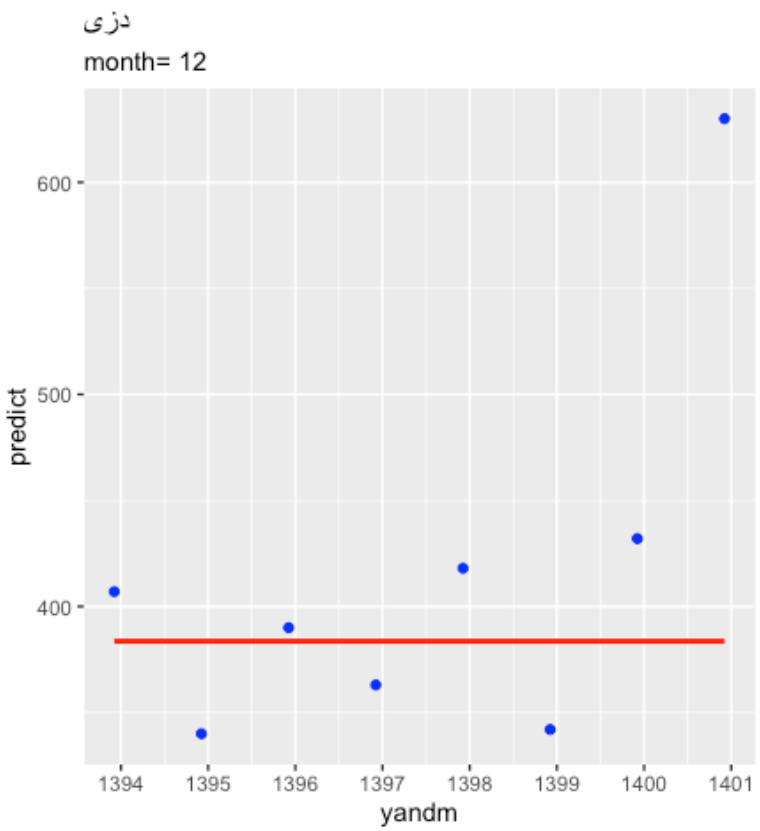

Fig 6: Yazd's death in esfand

For controlling covid, we rate provinces based on their normalized deaths, and you can watch the result in the ratio_death files. Based on our results best provinces are "Sistan & Balochistan," South Khorasan," and "North Khorasan," in order.

The worst are "Charmahal Bakhtiari," Alborz," and "Kurdistan."