

# R\_HW1

Zahra Soukhtedel-98105138

2023-04-28

## Problem 1

In order to see the intensity and weakness of Corona in different months, we have drawn the heat map of the additional deaths. To calculate the excess death rate, we first predict the death rate in the absence of Corona and compare it with the actual death rate and declare it as an excess death.

To predict the amount of additional deaths, we use the data of 5 years before Corona and use the linear regression model, and if the p-value is high, we use the average mortality. (In case of excess death, we declare that the actual death rate has a significant difference of more than 1.96 sigma from the predicted death rate).

libraries to be added:

```
library(knitr)
library(RColorBrewer)
library(data.table)
library(ggplot2)
```

function to calculate p-value:

```
lmp <- function (modelobject) {
  if (class(modelobject) != "lm") stop("Not an object of class 'lm' ")
  f <- summary(modelobject)$fstatistic
  p <- pf(f[1],f[2],f[3],lower.tail=F)
  attributes(p) <- NULL
  return(p)
}
```

load data and prepare it for analyses(covid-data df to save comparable datas, and moral-num to save actual death values):

```
d = fread('C:\\Users\\acer\\Documents\\R_hw1\\iranprovs_mortality_monthly.csv', encoding = 'UTF-8')
d$ym_num = d$y + d$m / 12 - 1/24
ds = d[, .(n = sum(n)), .(y, m, ym_num, prov)]
ym_num_covid = 1398 + 10/12 - 1/24
covid_data = data.frame(matrix(nrow = 31, ncol = 34))
moral_num = data.frame(matrix(nrow = 31, ncol = 34))
rownames(covid_data) = unique(ds[ym_num >= ym_num_covid]$prov)
colnames(covid_data) = sort(unique(ds[ym_num >= ym_num_covid]$ym_num), decreasing = F)
rownames(moral_num) = unique(ds[ym_num >= ym_num_covid]$prov)
colnames(moral_num) = sort(unique(ds[ym_num >= ym_num_covid]$ym_num), decreasing = F)
ym_num_start = ym_num_covid - 6
```

fill the covid-data data frame with computing values of each year-month-province with the model discussed above:

```

for (M in c(1:12)) {
  for (PROV in unique(d$prov)) {
    dsm = ds[prov == PROV & m == M & ym_num > ym_num_start,]
    dsm2fit = dsm[ym_num < ym_num_covid]
    fit = lm(n ~ ym_num, dsm2fit)
    if (lmp(fit) > 0.05){
      dsm$n_predicted = mean(dsm2fit$n)
      dsm$sd = sd(dsm2fit$n)
    }else{
      dsm$n_predicted = predict(fit ,dsm)
      dsm$sd = summary(fit)$sigma
    }

    dsm$moral_limit = dsm$n_predicted + 1.96 * dsm$sd
    dsm$diff = (dsm$n - dsm$moral_limit)/dsm$n
    dsm$diff[dsm$diff < 0] = 0
    dsm$moral_value = (dsm$n - dsm$n_predicted)
    dsm$moral_value[dsm$moral_value < 0] = 0

    for (ym_i in unique(dsm$ym_num[dsm$ym_num >= ym_num_covid])) {
      covid_data[PROV, as.character(ym_i)] = dsm[ym_num == ym_i & prov == PROV]$diff[1]
      moral_num[PROV, as.character(ym_i)] = dsm[ym_num == ym_i & prov == PROV]$moral_value[1]
    }
  }
}

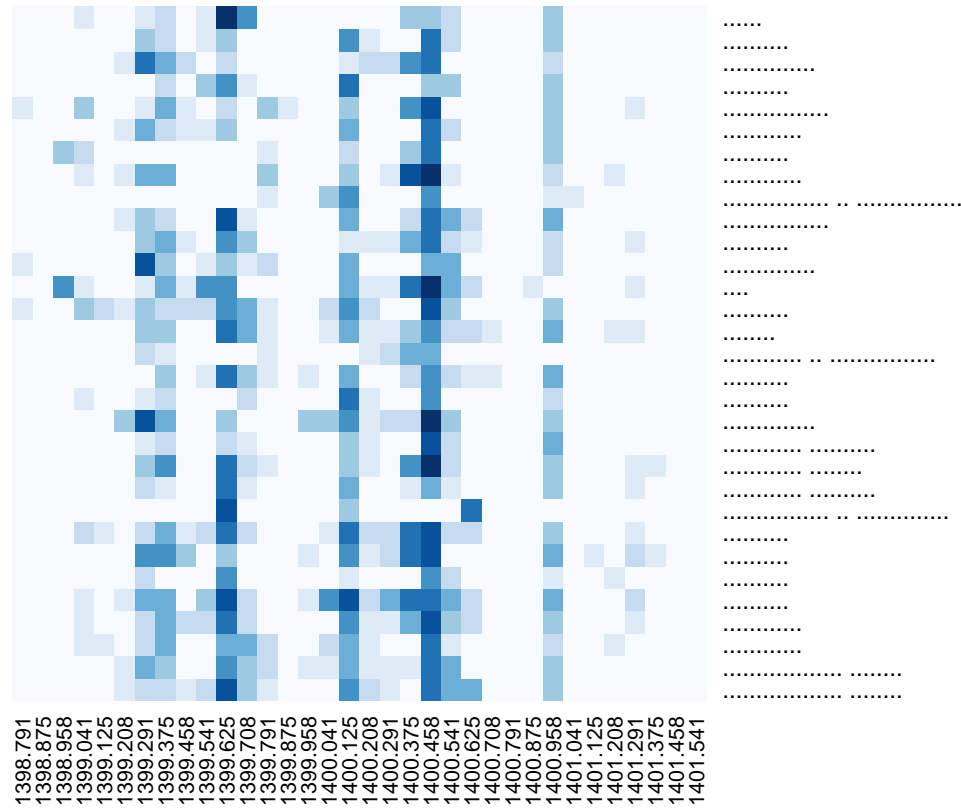
```

plot result heat-map:

```

covid_matrix <- as.matrix(covid_data)
for (i in c(1 : length(colnames(covid_matrix)))) {
  colnames(covid_matrix)[i] = substr(colnames(covid_matrix)[i], start=1, stop=8)
}
col <- brewer.pal(9, "Blues")
heatmap(covid_matrix, col=col, Colv = NA, Rowv = NA, scale="none")

```



As can be seen from the heat-map, it is easy to see the death waves of Corona.

## Problem 2

What was the total number of additional deaths in the entire country?

the number of total number of additional deaths is shown as follows:

```
sum(moral_num)
```

```
## [1] 337909.5
```

## Problem 3

What was the total number of additional deaths in each province?

the number of total number of additional deaths in each province is shown as follows:

```
rowSums(moral_num)
```

```
##
##      20932.181      14943.333      6019.533      23399.757
##
##      14789.200      1792.233      3794.919      59044.638
##
##      2392.610      2551.567      28498.300      3384.467
##
##      19987.090      3844.933      2779.976      7278.867
##
```

##	20248.533	5903.976	6281.305	6656.267
##				
##	11213.481	9362.600	1703.867	8327.600
##				
##	10113.767	6337.967	14142.667	5453.429
##				
##	4722.467	7797.624	4210.367	

## Problem 4

Which provinces have been more successful in controlling the disease? Is there a significant difference between the provinces?

first we compare normalized death sum in provinces, we have:

```
sort(rowSums(covid_data))
```

##				
##	1.404134	1.409505	1.644946	1.889617
##				
##	2.022819	2.023889	2.150018	2.336836
##				
##	2.508789	2.516576	2.700821	2.923400
##				
##	2.936497	2.987185	3.091699	3.416944
##				
##	3.546332	3.588921	3.609844	3.613025
##				
##	3.620324	3.866339	3.996669	4.263901
##				
##	4.272944	4.274724	4.288385	4.366568
##				
##	4.440086	4.735202	5.840383	

from data above we can conclude that most successful provinces are {gilan, south khorasan, chaharmaha, sistanal, now we will compare it with actual death values :

```
sort(rowSums(moral_num))
```

##				
##	1703.867	1792.233	2392.610	2551.567
##				
##	2779.976	3384.467	3794.919	3844.933
##				
##	4210.367	4722.467	5453.429	5903.976
##				
##	6019.533	6281.305	6337.967	6656.267
##				
##	7278.867	7797.624	8327.600	9362.600
##				
##	10113.767	11213.481	14142.667	14789.200
##				
##	14943.333	19987.090	20248.533	20932.181
##				
##	23399.757	28498.300	59044.638	

One of the factors that can bias our conclusions is the elderly people in the society, we eliminate their

influence, that is, we remove the data related to elderly people from our data and act as before and the normalized amount of extra deaths in the province We compare different.

```
ds = d[, .(n = sum(n)), .(y, m, ym_num, prov, age_group)]
for (M in c(1:12)) {
  for (PROV in unique(d$prov)) {
    dsm = ds[prov == PROV & m == M & ym_num > ym_num_start & age_group <= 4, ]
    dsm2fit = dsm[ym_num < ym_num_covid]
    fit = lm(n ~ ym_num, dsm2fit)
    if (lmp(fit) > 0.1){
      dsm$n_predicted = mean(dsm2fit$n)
      dsm$sd = sd(dsm2fit$n)
    }else{
      dsm$n_predicted = predict(fit ,dsm)
      dsm$sd = summary(fit)$sigma
    }

    dsm$moral_limit = dsm$n_predicted + 1.96 * dsm$sd
    dsm$diff = (dsm$n - dsm$moral_limit)/dsm$n
    dsm$diff[dsm$diff < 0] = 0
    dsm$moral_value = (dsm$n - dsm$n_predicted)
    dsm$moral_value[dsm$moral_value < 0] = 0

    dsm$ym_num[dsm$ym_num[dsm$ym_num]]

    for (ym_i in unique(dsm$ym_num[dsm$ym_num >= ym_num_covid])) {
      covid_data[PROV, as.character(ym_i)] = dsm[ym_num == ym_i & prov == PROV]$diff[1]
      moral_num[PROV, as.character(ym_i)] = dsm[ym_num == ym_i & prov == PROV]$moral_value[1]
    }
  }
}
```

now in new model we will check the normalized rate of death.

```
sort(rowSums(covid_data))

##
##      0.000000000      0.000000000      0.000000000      0.000000000
##
##      0.000000000      0.000000000      0.000000000      0.000000000
##
##      0.000000000      0.000000000      0.000000000      0.000000000
##
##      0.000000000      0.008712131      0.265675314      0.269194211
##
##      0.271266627      0.309255940      0.362699903      0.423985504
##
##      0.435892889      0.486163710      0.513690126      0.574930545
##
##      0.602466478      0.857631642      0.861239633      1.020823875
##
##      1.131234515      1.362786134      2.576668596
```

here we can see death rate of young people (younger than 40) in different provinces are like above, so setting percent -of young people died over Corona over number of all young dead people- as our base.

if we took elder people as our base we have:

```

ds = d[, .(n = sum(n)), .(y, m, ym_num, prov, age_group)]
for (M in c(1:12)) {
  for (PROV in unique(d$prov)) {
    dsm = ds[prov == PROV & m == M & ym_num > ym_num_start & age_group > 4, ]
    dsm2fit = dsm[ym_num < ym_num_covid]
    fit = lm(n ~ ym_num, dsm2fit)
    if (lmp(fit) > 0.1){
      dsm$n_predicted = mean(dsm2fit$n)
      dsm$sd = sd(dsm2fit$n)
    }else{
      dsm$n_predicted = predict(fit ,dsm)
      dsm$sd = summary(fit)$sigma
    }

    dsm$moral_limit = dsm$n_predicted + 1.96 * dsm$sd
    dsm$diff = (dsm$n - dsm$moral_limit)/dsm$n
    dsm$diff[dsm$diff < 0] = 0
    dsm$moral_value = (dsm$n - dsm$n_predicted)
    dsm$moral_value[dsm$moral_value < 0] = 0

    dsm$ym_num[dsm$ym_num[dsm$ym_num]]

    for (ym_i in unique(dsm$ym_num[dsm$ym_num >= ym_num_covid])) {
      covid_data[PROV, as.character(ym_i)] = dsm[ym_num == ym_i & prov == PROV]$diff[1]
      moral_num[PROV, as.character(ym_i)] = dsm[ym_num == ym_i & prov == PROV]$moral_value[1]
    }
  }
}
sort(rowSums(covid_data))

```

```

##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.0000000
##
##      0.0000000      0.0000000      0.0000000      0.1060979
##
##      0.1363530      0.1515639      0.4860797

```