Idea:

Patient-level survival and individual characteristics (such as age group, gender)

for a representative sample from each country

Reality:

\*I National/regional summary measurements are most readily available

\*I Note: variables that describe groups of individuals (not the individuals themselves) are often called ecological variables

\*Data on obesity rate, number of confirmed COVID-19 cases and deaths available for 9 countries .（**Reference:data disaggregated by both sex and age on confirmed cases and deaths per 100,000 men and women in the population**

**(Austria by 6.3,England by 5.27,Italy by 5.26,Portugal by 6.1,Mexico by 6.1,Spain by 5.18,Russia by 6.3,Denmark by 6.3,Norway by 6.2)（[https://globalhealth5050.org/covid19/age-and-sex-data/#1589893713714-c1529cca-81cb](https://globalhealth5050.org/covid19/age-and-sex-data/" \l "1589893713714-c1529cca-81cb)）**

\*We calculate the death ratio for each country :

DR = confirmed deaths due to COVID-19 in each age group/ confirmed cases of COVID-19 in each age group

And for gender and age multiple model,we calculate the death ratio for each country:

DR = confirmed deaths due to COVID-19 in each age group(male or feamle)/ confirmed cases of COVID-19 in each age group(male or female)

\*This estimate does not attempt to account for limitations due to data quality

Children(0-18)

Adults: (19-65)

Old(65+)

Introduction

The unidentified virus pneumonia was discovered in December 2019, and the virus was officially named COVID-19 in January 2020. Until now, the outbreak has caused more than 7 million Confirmed cases and 400, 000 Confirmed deaths worldwide.[1]

Then in this poster I will be investigating in whether the death ratio of COVID—19 is related to individual characteristics (such as age group, gender)

Contents

1. HYPOTHESIS MODEL(Logistic regression)
2. TEST THE MODEL

(AIC \& P value \& step function \& confidence interval）

（Logistic Regression)

(To find out which is more associated with mortality)

1. Some graphs(Mosaic plot&ggplot2)

4.LINEAR REGRESSION

Sketching the relationship between death ratio and age

5.CONCLUSION&APPLICATION

6.ANALYZE IMAGES & CONCLUSIONS

1. HYPOTHESIS MODEL

Use the data “COVID19\_open\_line\_list”[2]

Variables: death (death 1, alive 0), age, sex

Set model with General Linear Model.

Model1~glm(death~age) (the relationship between death and age)

Model2~glm(death~sex) (the relationship between death and sex)

Model3~glm(death~age+sex) (the relationship between death and personal characteristics(age and sex), and this is a multiple relationship)

2.TEST THE MODEL(to find out which is more associated with mortality)

Using logistic regression method, AIC value, P value ,Confidence Interval and odds ratio can be obtained .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| mod  Reference data | AIC | P | Confidence Interval(97.5%) | OR | Estimate |
| Model1(age) | 500.43 | 0.128 | 0.003269068 | 0.9887340 | -0.011330 |
| Model2(sexmale) | 502.74 | 0.884 | 0.5303877 | 1.03742574 | 0.0364 |

1. Comparing the Confidence interval of Model1 and Model2 (97.5%), Model1 is very small while Model2 is over 0.5. Meanwhile, comparing the P-value of both, Model1 is too large within the acceptable range, while Model1's AIC is slightly smaller than Model2's AIC. Therefore, it is preliminarily judged that Model1 is better than Model2.

2. According to Model2‘s OR, the mortality rate of male patients is 1.03742574 times that of female patients. Male and female mortality rates are similar, then initial guess that the relationship is very small.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter（Model3)  Reference data | AIC | P | Confidence Interval(97.5%) | OR | Estimate |
| (Intercept) | 502.43 | 1.07e-08 | -1.382845037 | 0.1219892 | -2.103823 |
| sexmale | 0.955 | 0.509320333 | 1.0144218 | 0.014319 |
| age | 0.130 | 0.003316708 | 0.9887573 | -0.011306 |

1. Look at the table in Model3, AIC is relatively large, and the P-value and Confidence Interval (97.5%) of parameter sexmale are too large to be accepted, we can guess again that the relationship between death and sex is very small.

Estimate:Death has little to do with gender, and Model1 is a better model than the other two.

1. Some graphs(Mosaic plot&ggplot2）
2. histograms for three distribitions:

Mosaic plot

Box plot(ggplot2)

Scatter diagram by gplot (ggplot2)

1. Analysis

The first graph shows that the mortality rate between males and females is close, the mortality rate in female adults and older males is slightly lower. The old have higher mortality rates than adults, with children being the highest, but children's residuals are too large to be accepted.

The second and three graphs show that most of the sufferers and the dead are adults, but the mortality rate among adults is the lowest (guess: children and the elderly may be better protected and there are fewer children and the elderly)

1. Additional information：

Graph1:Children(0-17)，Adults: (19-65)，Old(65+)

Death ratio:children14.3%,adults6.6%,old7.69%

4.LINEAR MODEL(between death ratio and age by linear regression)

Sketching the relationship between death ratio and age

Use the data on confirmed cases and deaths per 100,000 men and women in the population.[3]

Using ggplot2,draw a scatter plot and fit the curve

Additional information：DR = \frac{confirmed deathsin each age group}{confirmed cases in each age group}

Using Linear Regression method.**Fitting a linear model**

The residuals graph of linear model shows that the points are randomly distributed around 0.

The boxplot to shows that there are no residuals are outliers.

Most of the points are randomly distributed around 0,this is the evidence to the normal distribution.There is no outlier,and the most of the points are within the 0.04 and -0.06 so the fit is good enough.

## [2]Aadhil Imam.(COVID19\_open\_line\_list.csv

)Available from <https://github.com/aadhil96/covid19_analysis_and_prediction/blob/master/COVID19_open_line_list.csv>[Accessed 23rd March 2020]

### [3]Sliver District. (COVID-19: Data disaggregated by age and sex).Available from [https://globalhealth5050.org/covid19/age-and-sex-data/#1589893713714-c1529cca-81cb](https://globalhealth5050.org/covid19/age-and-sex-data/" \l "1589893713714-c1529cca-81cb)[Accessed 4 June 2020]

[4]From: Alexander Yuryatin

. （COVID-19 case fatality-derived risk of death adjusted by age and gender）. Available from: [https://zenodo.org/record/3787931#.Xt4iWp4zYXo](https://zenodo.org/record/3787931" \l ".Xt4iWp4zYXo)[Accessed 7th May 2020].

[1]Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina and Joe Hasell (2020) .（Coronavirus Pandemic (COVID-19)）. Available from

: 'https://ourworldindata.org/coronavirus'[Accessed 11th June 2020].

<https://globalhealth5050.org/covid19/age-and-sex-data/#1589893214590-52bd08>

[https://globalhealth5050.org/covid19/age-and-sex-data/#1589893214590-52bd08c3-f1e9](https://globalhealth5050.org/covid19/age-and-sex-data/" \l "1589893214590-52bd08c3-f1e9)

The error analysis：

1.In graph it shows that the death ratio in children is not in the range of acceptance,so the children group mortality rate should be questioned.

2.The mortality rate changes with time[3], but the first data group is as of 2/29/2020, and the second data group is as of 6/3/2020.

3.In the first data set, there are many missing data, it is difficult to find whether those patients died or not, more than 12,000 data were deleted, so it is very easy to make errors。