**Motivation（Introduction）：**

This year there has been a global outbreak of COVID-19,until now, Covid-19 has caused more than 7 million Confirmed cases and 400, 000 Confirmed deaths all over the world.[1]

In order to reduce mortality，I will be investigating in whether the death ratio of COVID—19 is related to individual characteristics (which includes age group and gender)in this poster.

**~~What you did (explain main points of your project) Contents~~**

I Divide the contents into 6 parts

1. HYPOTHESIS MODEL

I use the first data set [2] which contains main variables (death, age, sex.) to set model with General Linear Model.

And after processing, I marked death as 1, alive as 0, and deleted rows with Spaces.

Then we can set these 3 models.

Model1~glm(death~age)

Model2~glm(death~sex)

Model3~glm(death~age+sex),which is a multiple relationship

1. TEST THE MODEL

I use the Logistic Regression to find out the Reference data in these three models

Look at these two tables.

Comparing the Confidence interval of Model1 and Model2, Model1 is very small while Model2 is over 0.5.

Meanwhile, comparing the P-value of both, Model2 is too large within the acceptable range, while Model1's AIC is slightly smaller than Model2's AIC.

Then I can find that Model1 is better than Model2.

\\2.According to Model2‘s Odds ratio, the mortality rate of male patients is 1.037 times that of female patients.

That means The risk of death in male is slightly higher than that in female.

In Model3, AIC is relatively large, the P-value and Confidence Interval of parameter sexmale are too large to be accepted, we can guess that the model1 is the best one.

To confirm the accuracy of this guess, I set 2 models:mod.full and mod.none with full and none parameter, use the step function,and confirm that model1 is the best one.

~~3.ANALYSIS OF SOME GRAPHICS~~

In the third part,

I define Children means0-18,Adults means19-65),Old means 65+)

I draw 3 graphs, Mosaic plot,box plot and Scatter diagram

which show that the mortality rate is slightly lower in female adults . The old have higher mortality rates than adults, with

children being the highest, but its residuals are too large to be accepted,and most of the sufferers and dead are adults, but the mortality rate among adults is the lowest .

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

Then I use the data2 to draw a scatter graph of death ratio against age and fit the curve by ggplot2.

Here Death Ratio = confirmed deaths in a fixed age/ confirmed cases in that fixed age

Then I using the Linear Regression method to Fit a linear model.

To analysis this model,I draw a residuals graph and a box plot,which show that Most of the points are randomly distributed around 0,this is the evidence to the normal distribution.And they also show that There is no outlier, and the most of the points are within 0.04 and -0.06 .So the fit is good enough.

~~5.CONCLUSIONS & INFORMATION~~

直接说（In conclusion)

Firstly，Males are more likely to die than females.

Secondly，The relationship between age and mortality is more obvious, and the death rate increases with age.(which is opposite with many other infectious diseases)

That means for many infectious diseases young children are most at risk. For instance, in the case of malaria, the [57% of deaths](https://ourworldindata.org/grapher/malaria-deaths-by-age) are in children under five.

That is the most interesting thing that I found.

So if I have time,I would like to do research for why the old are at the greatest risk of dying in Covid-19.If that is due to the fact that older persons are more unhealthy and more likely to get some disease(such as cardiovascular diseases and diabetes).

However, through collecting some information on the Internet, I found that the relationship between death rate and gender is also quite significant. The death rate of males is usually higher than that of females。

So we can find that the conclusion drawn from data2 is more consistent with expectations.

Why there are some differences.

To explain that we can go to the next part

(6. THE ERROR ANALYSIS)(不读）

By finding some information, I find 3 possible reasons.

First,The Mosaic plot in data1 shows that the death ratio in children is not in the range of acceptance,so it should be questioned.

Second,the mortality rate changes with time[5], but these 2 datas end at different time.

Third,Data1 misses many datas, and it is diffificult to find whether those patients died or not, more than 12thousand rows were deleted, so It may cause errors.

**In the end,the main conclusion and the application in this project is that** We can better protect the health of all people and allocate medical resources more rationally.That means we already know that coVID-19 mortality increases with age, so we can focus on protecting older people and allocating more medical resources to them to improve their cure rates.