

Example of Analysis

Panfeng Liang

June 22, 2020

Analysis on the Number of Infected Cases through Log Linear Model

```
count.data=clean.data %>% count(gender, occupation, province)

fit_infected=glm(formula = n ~ factor(gender) + factor(occupation)+factor(province),
                  data = count.data, family = poisson(link=log))
summary(fit_infected)

##
## Call:
## glm(formula = n ~ factor(gender) + factor(occupation) + factor(province),
##      family = poisson(link = log), data = count.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -24.391   -2.964   -1.176    4.246   16.960
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      7.14046    0.02591  275.64  <2e-16 ***
## factor(gender)M    -0.87891    0.03649  -24.09  <2e-16 ***
## factor(occupation)OTHER -1.05903    0.04635  -22.85  <2e-16 ***
## factor(occupation)PHYSICIAN -0.41995    0.03735  -11.24  <2e-16 ***
## factor(province)Other  -4.39865    0.15168  -29.00  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 7077.7  on 11  degrees of freedom
## Residual deviance: 1295.2  on 7  degrees of freedom
## AIC: 1371.2
##
## Number of Fisher Scoring iterations: 5
```

According to the data, female medical works are more likely to be infected by Covid19.

- In each category of occupation and location, the number of infected male medical workers is $\exp(-0.88) = 0.42$ that of females.

Nurses are more likely to be infected by Covid19 than physicians and other medical staffs.

- Among both genders, physicians infected by Covid19 is just $\exp(-0.42) = 0.66$ times that of nurses.
- Among both genders, other medical staffs infected by Covid19 is just $\exp(-1.06) = 0.35$ times that of nurses.

In Hubei province, medical workers are more likely to be infected.

- Medical workers infected in other provinces are just $\exp(-4.4) = 0.01$ of that in Hubei provinces.

Analysis on the Death Rate through Logistic Model :

```
fit_death=glm(formula = survival ~ factor(gender)+factor(occupation),
              data = clean.data, family = binomial(link = logit))
summary(fit_death)

##
## Call:
## glm(formula = survival ~ factor(gender) + factor(occupation),
##      family = binomial(link = logit), data = clean.data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7048   0.0457   0.0697   0.1543   0.2295
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)         6.8618    0.7087   9.682 < 2e-16 ***
## factor(gender)M       -0.8425    0.4295  -1.961  0.04982 *
## factor(occupation)OTHER -1.5945    0.8687  -1.835  0.06645 .
## factor(occupation)PHYSICIAN -2.3956    0.7926  -3.022  0.00251 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 356.92  on 3622  degrees of freedom
## Residual deviance: 320.03  on 3619  degrees of freedom
## AIC: 328.03
##
## Number of Fisher Scoring iterations: 9
```

Among all medical occupations, females are more likely to survive from the infection of Covid19 according to the data.

- Among all medical occupations, the estimated conditional odds ratio between the survival of Covid19 and gender is $\exp(-0.84) = 0.43$.
- Among male medical works, the estimated odds of survival of Covid19 infection is just 43% of that in female ones.

Among both genders of infected medical workers, nurses are more likely to survive than physicians and other medical staffs.

- Among a specific gender, the estimated odds that physicians survive from Covid19 infection is $\exp(-2.4) = 0.09$ times the estimated odds for nurses.

- Among a specific gender, the estimated odds that other medical staffs survive from Covid19 infection is $\exp(-1.59) = 0.2$ times the estimated odds for nurses.