Analysis

Panfeng Liang

June 22, 2020

```
clean.data=read.csv("data.csv")
attach(clean.data)
```

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

```
##
## 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
```

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

fit_infected=glm(formula = n ~ factor(gender) + factor(occupation)+factor(province), dat
a = 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
## -24.2551   -0.3660    0.0000    0.6734   16.8233
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.42376    0.35405   4.021 5.79e-05 ***
## factor(gender)M              -0.87591    0.03655 -23.964  < 2e-16 ***
## factor(occupation)OTHER      -1.04924    0.04641 -22.609  < 2e-16 ***
## factor(occupation)PHYSICIAN -0.42457    0.03745 -11.338  < 2e-16 ***
## factor(province)四川省      -0.12328    1.06130  -0.116    0.908
## factor(province)天津        -0.12328    1.06130  -0.116    0.908
## factor(province)安徽省      -0.99919    1.06085  -0.942    0.346
## factor(province)山东省      -0.16375    0.57021  -0.287    0.774
## factor(province)广西省      -1.42376    1.06082  -1.342    0.180
## factor(province)江苏省       0.55962    0.44320   1.263    0.207
## factor(province)江西省      -0.65422    0.79082  -0.827    0.408
## factor(province)河南省      -0.56154    0.61256  -0.917    0.359
## factor(province)海南省      -0.25677    0.79105  -0.325    0.745
## factor(province)湖北         2.23327    0.36632   6.096 1.08e-09 ***
## factor(province)湖北省       5.68441    0.35410  16.053  < 2e-16 ***
## factor(province)湖南省      -1.42376    1.06082  -1.342    0.180
## factor(province)甘肃省      -0.12328    1.06130  -0.116    0.908
## factor(province)福建省      -1.42376    1.06082  -1.342    0.180
## factor(province)重庆        -1.42376    1.06082  -1.342    0.180
## factor(province)陕西省      -0.99919    1.06085  -0.942    0.346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 14124.1  on 36  degrees of freedom
## Residual deviance:  1291.4  on 17  degrees of freedom
## AIC: 1462.3
##
## Number of Fisher Scoring iterations: 5
```