MMPH6117 Final Examination

Question 3

Name: Mingze Gao

UID: 3035426241

Read data sets.

library(readr)  
exam1 <- read\_csv("../Final/exam\_question1.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## study = col\_character(),  
## country = col\_character(),  
## sample\_size = col\_double(),  
## cfr = col\_double(),  
## cfr.lb = col\_double(),  
## cfr.ub = col\_double()  
## )

exam2 <- read\_csv("../Final/exam\_question2.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## id = col\_double(),  
## male = col\_double(),  
## age = col\_double(),  
## phy.index = col\_double(),  
## bmi = col\_double(),  
## low.income = col\_double(),  
## low.calc = col\_double(),  
## poor.bone = col\_double()  
## )

exam3 <- read\_csv("../Final/exam\_question3.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## id = col\_double(),  
## age = col\_double(),  
## comorbid = col\_double(),  
## death = col\_double()  
## )

1. A

exam3$agegp <- 1  
for(i in 1:nrow(exam3))   
{  
 if (exam3[i, "age"] >= 60)  
 {  
 exam3[i, "agegp"] = 2  
 }  
 else  
 {  
 if (exam3[i, "age"] <= 17)  
 {  
 exam3[i, "agegp"] = 0  
 }  
 }  
}  
round(with(exam3, prop.table(table(agegp, death),1)),3)

## death  
## agegp 0 1  
## 0 0.994 0.006  
## 1 0.990 0.010  
## 2 0.619 0.381

Agegp 0, 1, 2 corresponding to 0-17y, 18-59y and >= 60y respectively. The the case fatality ratio by age groups are 0.6%, 1% and 38.1% respectively. A high fatality rate can be observed in elderly.

1. B

require(MASS)

## Loading required package: MASS

reb <- glm(death~age+comorbid, data=exam3, family=binomial)  
summary(reb)

##   
## Call:  
## glm(formula = death ~ age + comorbid, family = binomial, data = exam3)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9968 -0.3430 -0.0783 -0.0176 4.0488   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -9.917227 0.775495 -12.79 <2e-16 \*\*\*  
## age 0.114736 0.009452 12.14 <2e-16 \*\*\*  
## comorbid 0.290657 0.234358 1.24 0.215   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 898.97 on 999 degrees of freedom  
## Residual deviance: 472.47 on 997 degrees of freedom  
## AIC: 478.47  
##   
## Number of Fisher Scoring iterations: 7

Based on the coefficients, increasing both age and comorbid could result in higher death rate, however, the p-value of comorbid for this model is not significant.

1. C

exam3$age2 <- exam3$age \* exam3$age  
rec <- glm(death~age2+age+comorbid, data=exam3, family=binomial)  
summary(rec)

##   
## Call:  
## glm(formula = death ~ age2 + age + comorbid, family = binomial,   
## data = exam3)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1980 -0.3213 -0.1237 -0.0789 3.3765   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.7945595 1.3581344 -4.267 1.99e-05 \*\*\*  
## age2 0.0008421 0.0002914 2.890 0.00385 \*\*   
## age -0.0061357 0.0398393 -0.154 0.87760   
## comorbid 0.3094653 0.2415124 1.281 0.20007   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 898.97 on 999 degrees of freedom  
## Residual deviance: 466.56 on 996 degrees of freedom  
## AIC: 474.56  
##   
## Number of Fisher Scoring iterations: 8

Based on the coefficients, increasing both age^2 and comorbid could result in higher death rate whilst age poses negative effect on the outcome. However, the p-value of comorbid and age for this model is not significant. Also, the coefficients for age and age2 are very small, might not be credible.

1. D

deviance(reb)/df.residual(reb)

## [1] 0.4738951

deviance(rec)/df.residual(rec)

## [1] 0.4684306

For reference, to evaluate satisfactory fitting of the model, the residual deviance / df should not be too much bigger than 1, better if it is below 1. The above model fits satisfactorily.

1. E

AIC(reb, rec)

## df AIC  
## reb 3 478.4734  
## rec 4 474.5569

Regarding AIC metric, a lower AIC indicates a better model. So based on the results of AIC and goodness of fit, model C is preferred.  
However, model C is quadratic effect on age which is not necessary I suppose and since the difference between AIC and residual is small, model B might be better.

1. F

require(car)

## Loading required package: car

## Loading required package: carData

require(MASS)  
library(survival)  
vif(rec)

## age2 age comorbid   
## 24.687148 24.723860 1.007409

exam3$agescale <- scale(exam3$age, scale = F)  
exam3$age2 <- exam3$agescale \* exam3$agescale  
ref <- glm(death~age2+age+comorbid, data=exam3, family=binomial)  
summary(ref)

##   
## Call:  
## glm(formula = death ~ age2 + age + comorbid, family = binomial,   
## data = exam3)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.1980 -0.3213 -0.1237 -0.0789 3.3765   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.0076815 0.7774313 -10.300 < 2e-16 \*\*\*  
## age2 0.0008421 0.0002914 2.890 0.00385 \*\*   
## age 0.0802031 0.0121731 6.589 4.44e-11 \*\*\*  
## comorbid 0.3094653 0.2415124 1.281 0.20007   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 898.97 on 999 degrees of freedom  
## Residual deviance: 466.56 on 996 degrees of freedom  
## AIC: 474.56  
##   
## Number of Fisher Scoring iterations: 8

vif(ref)

## age2 age comorbid   
## 2.298296 2.308318 1.007409

The multicollinearity among the variables of model C is high for age and age2 because vif result is far away from 1.  
We could see that after certering the data, the vif can be reduced as well as the collinearity.

1. G
2. The crude fatality rate for zoonotic disease is higher for elderly people, whilst the difference is not tremendous between other adults and teenagers.
3. Beside age, comorbid could also contribute to the death rate, however the effect might not be significant.
4. Certering is sometims useful for tackling the collinearity problem. After reducing, both age2 and age are significant regarding p value. The effect of age2 is smaller than age.