Assignment 1

Name: QINGTAN SHEN
Student number:1130945

1.(a) First step:

At first, we change the region, age and education information to factor

```
domviolence$reg.f <- factor(domviolence$reg)
domviolence$age.f <- factor(domviolence$age)
domviolence$educ.f <- factor(domviolence$educ)
domviolence$ms.f <- factor(domviolence$ms)
```

Then we create the whole model with all variables mentioned. This model is named as model0:

```
model0 <- glm(dv ~ age.f + ms.f + mmo + smok + alc + falc + educ.f + reg.f, family=binomial, data=domviolence)
summary(model0)
```

After that we use stepwise method to delete unsignificant variables, then we get model0.5:

```
#use stepwise in model0
model0.5 <- step(model0, trace=0)
summary(model0.5)
```

Then model0.5 = glm(dv ~ age.f + ms.f + smok + alc + falc + educ.f + reg.f), from the summary of model0.5, we find the p-value of ms.f6, alc, reg.f3 is much greater than 0.05, so anova test should be applied on these variables:

```
# measure the significant of reg.f (because the p-value in reg.f.south=0.898)
model0.6 <- glm(dv ~ age.f + ms.f + smok + alc + falc + educ.f, family=binomial, data=domviolence)
summary(model0.6)
anova(model0.6, model0.5, test='Chi')
# the p-value of Chi test is 1.43e-06, which means reg.f is significant and we cannot delete this variable

# next we should consider the variable ms
model0.65 <- glm(dv ~ age.f + smok + alc + falc + educ.f + reg.f, family=binomial, data=domviolence)
summary(model0.65)
anova(model0.65, model0.5, test='Chi')
# p-value=0.0008, means variable ms is significant

#next one is alc (the p-value for alc is the second largest, which is 0.090)
model0.7 <- glm(dv ~ age.f + ms.f + smok + falc + educ.f + reg.f, family=binomial, data=domviolence)
summary(model0.7)
anova(model0.7, model0.5, test='Chi')
# the p-value is 0.117, larger than 0.05, so alc can be ignored
```

We should delete alc variable (p_value > 0.05), Both ms and reg variables should be kept (p-value < 0.05), then we get model1 = glm(dv ~ age.f + ms.f + smok + falc + educ.f + reg.f) in the first step.

1.(b) Second step

Firstly, we need to create a model contains both age, education variables and age factor and education factor variables, then we get model_compare:

```
> model_compare <- glm(dv~ms.f + smok + falc + reg.f + age + educ + age.f + educ.f,family=binomial,data=domviolence)
> anova(model_compare, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: dv

Terms added sequentially (first to last)


       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                    1315     1561.6
ms.f    5   42.425      1310     1519.2 4.833e-08 ***
smok    1   23.608      1309     1495.6 1.181e-06 ***
falc    1   10.460      1308     1485.1  0.001220 **
reg.f   3   33.785      1305     1451.4 2.200e-07 ***
age     1    7.391      1304     1444.0  0.006556 **
educ    1   16.081      1303     1427.9 6.068e-05 ***
age.f   2    3.281      1301     1424.6  0.193911
educ.f  1    3.170      1300     1421.4  0.074995 .
```

Then we change the variable order to get model_compare1:

```
> model_compare1 <- glm(dv ~ ms.f + reg.f + smok + educ + falc + age + educ.f + age.f, family=binomial, data=domviolence)
> anova(model_compare1, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: dv

Terms added sequentially (first to last)


       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                    1315     1561.6
ms.f    5   42.425      1310     1519.2 4.833e-08 ***
reg.f   3   38.389      1307     1480.8 2.338e-08 ***
smok    1   20.407      1306     1460.4 6.260e-06 ***
educ    1   10.259      1305     1450.2  0.001360 **
falc    1    8.302      1304     1441.8  0.003960 **
age     1   13.967      1303     1427.9  0.000186 ***
educ.f  1    3.493      1302     1424.4  0.061629 .
age.f   2    2.958      1300     1421.4  0.227882
```

educ.f and age.f have the largest p-value (greater than 0.05), so these two variables can be replaced by educ and age. Then we get model2: glm(dv ~ ms.f + reg.f + smok + educ + falc + age)

1.(c) Third step

To test their first order interaction terms, we put all the first order interaction into model3, and use stepwise method on model3 to get model3.1:

```
model3 <- glm(dv ~ (age + ms.f + smok + falc + educ + reg.f)^2, family=binomial, data=domviolence)
summary(model3)

model3.1 <- step(model3, trace=0)
summary(model3.1)
```

Then we get model3.1 = glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc + educ:reg.f), in the summary of model3.1, the interaction p-value of ms.f4:falc, smok:falc and educ:reg.f3 is much greater than 0.05, so we should test them using anova method separately:

```r
# consider the interaction between ms and falc
model3.2 <- glm(dv ~ age + ms.f + smok + falc + educ + reg.f + smok:falc + educ:reg.f, family=binomial, data=domviolence)
anova(model3.2, model3.1, test='Chi')
# Pr=0.037 < 0.05, the interaction cannot be ignored, also ms and falc cannot be ignored

# consider the interaction between educ and reg.f
model3.3 <- glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc, family=binomial, data=domviolence)
anova(model3.3, model3.1, test='Chi')
summary(model3.3)
# Pr=0.071 > 0.05, the interaction can be ignored

# consider the interaction between smok and falc
model3.4 <- glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc, family=binomial, data=domviolence)
anova(model3.4, model3.3, test='Chi')
# Pr=0.215 > 0.05, the interaction can be ignored
```

We only keep the interaction between ms and falc because its p-value is smaller than 0.05. So we get the final model: model4 = glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc), then we give model4 an anova test:

```
> anova(model4, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: dv

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                      1315     1561.6
age        1   23.321      1314     1538.3 1.371e-06 ***
ms.f       5   31.207      1309     1507.1 8.526e-06 ***
smok       1   19.669      1308     1487.4 9.210e-06 ***
falc       1   10.275      1307     1477.2  0.001348 **
educ       1   18.262      1306     1458.9 1.926e-05 ***
reg.f      3   31.016      1303     1427.9 8.435e-07 ***
ms.f:falc  5   12.791      1298     1415.1  0.025418 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We see that the p-value of all the variables and interactions are smaller than 0.05, So it can be our final model. Then we use summary method to get the parameters value:

$$
\begin{aligned}
\text{logit}(\theta) \;=\; & -0.038 + 0.797\times\texttt{ms.2} + 0.439\times\texttt{ms.3} + 1.312\times\texttt{ms.4} + 0.488\times\texttt{ms.5} + 0.143\times\texttt{ms.6} \\
& +0.533\times\texttt{smok} + 0.526\times\texttt{falc} - 0.908\times\texttt{reg.2} + 0.028\times\texttt{reg.3} - 0.424\times\texttt{reg.4} \\
& -0.347\times\texttt{age} - 0.490\times\texttt{educ} - 1.781\times\texttt{ms.2}:\texttt{falc} + 0.320\times\texttt{ms.3}:\texttt{falc} \\
& +0.249\times\texttt{ms.4}:\texttt{falc} + 0.595\times\texttt{ms.5}:\texttt{falc} + 0.118\times\texttt{ms.6}:\texttt{falc}
\end{aligned}
$$

2.(1) Marital status:

```
# 2.(1) ms
coe <- model4$coefficients
# for falc = 0
# ms2 vs ms1:
exp(coe[3]) #2.218
# ms3 vs ms1:
exp(coe[4]) #1.552
# ms4 vs ms1:
exp(coe[5]) #3.713
# ms5 vs ms1:
exp(coe[6]) #1.629
# ms6 vs ms1:
exp(coe[7]) #1.154
```

```
# for falc = 1
# ms2 vs ms1:
exp(coe[3] + coe[14]) #0.374
# ms3 vs ms1:
exp(coe[4] + coe[15]) #2.137
# ms4 vs ms1:
exp(coe[5] + coe[16]) #4.762
# ms5 vs ms1:
exp(coe[6] + coe[17]) #2.954
# ms6 vs ms1:
exp(coe[7] + coe[18]) #1.298
```

Odds ratios, at various levels of falc	falc = 0	falc = 1
for ms=1 vs. ms=1	1	1
for ms=2 vs. ms=1	2.218	0.374
for ms=3 vs. ms=1	1.552	2.137
for ms=4 vs. ms=1	3.713	4.762
for ms=5 vs. ms=1	1.629	2.954
for ms=6 vs. ms=1	1.154	1.298

In the conditions when family member not use of alcohol for certain, the estimated domestic violence odds of de Facto status (ms=2) increase by 121.8% when comparing to married status (ms=1).

In the conditions when family member use of alcohol for certain, the estimated domestic violence odds of divorced (ms=3) increase by 113.7% when comparing to married status (ms=1).

(2) Smoking:

```
# smoking
exp(coe[8]) #1.704
```

The estimated domestic violence odds of non-smoking women increased by 70.4% when comparing to smoking women.

(3) Family alcohol:

```
# ms/falc = 0 vs falc = 0:
# 1

# ms/falc = 1 vs falc = 0:
#ms1-ms6:
exp(coe[9]) #1.693
exp(coe[9] + coe[14]) #0.285
exp(coe[9] + coe[15]) #2.331
exp(coe[9] + coe[16]) #2.171
exp(coe[9] + coe[17]) #3.068
exp(coe[9] + coe[18]) #1.904
```

Odds ratios, at various levels of ms	ms=1	ms=2	ms=3	ms=4	ms=5	ms=6
for falc=0 vs. falc=0	1	1	1	1	1	1
for falc=1 vs. falc=0	1.693	0.285	2.331	2.171	3.068	1.904

In the conditions when current marital status is married (ms=1), the estimated domestic violence odds of family use of alcohol for certain (falc=1) increase by 69.3% when comparing to family not use of alcohol for certain (falc=0).

In the conditions when current marital status is divorced (ms=3), the estimated domestic violence odds of family use of alcohol for certain (falc=1) increase by 133.1% when comparing to family not use of alcohol for certain (falc=0).

(4) Region:

```
# region vs north: north:north = 1
#other three regions
exp(coe[11]) #0.403
exp(coe[12]) #1.028
exp(coe[13]) #0.655
```

	north	east	south	west
Ratio of odds (OR) for each region vs. north	1	0.403	1.028	0.655

The estimated domestic violence odds for east region is decreased by 59.7% when comparing to north region.

The estimated domestic violence odds for south region is increased 2.8% when comparing to north region.

(5) Age:

```
#age
exp(coe[2]) #0.707
```

The estimated domestic violence odds for age in 18-29 is decreased 29.3% when comparing to age in 30-49, the odds ratio also decreased 29.3% from age 30-49 to 50-64, also decreased 29.3% from age 50-64 to 65+.

(6) Education:

```
#education
exp(coe[10]) #0.613
```

The estimated domestic violence odds for education period between 7 and 11 years decreased 38.7% when comparing to age education period smaller or equal to 6 years.

The estimated domestic violence odds for education period greater or equal to 11 years also decreased 38.7% when comparing to age education period between 7 and 11 years.

Appendix

```
> domviolence <- read.csv("/Users/apple/Desktop/MAST90139/domviolence.csv")
> dim(domviolence)
[1] 1316    9
> #1.(a)
> #change the region,age and education information to factor
> domviolence$reg.f <- factor(domviolence$reg)
> domviolence$age.f <- factor(domviolence$age)
> domviolence$educ.f <- factor(domviolence$educ)
> domviolence$ms.f <- factor(domviolence$ms)
> # the whole model
> model0 <- glm(dv ~ age.f + ms.f + mmo + smok + alc + falc + educ.f + reg.f, family=binomial, data=domviolence)
>
> #use stepwise in model0
> model0.5 <- step(model0, trace=0)
> summary(model0.5)

Call:
glm(formula = dv ~ age.f + ms.f + smok + alc + falc + educ.f +
    reg.f, family = binomial, data = domviolence)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7422  -0.8091  -0.5801   1.0542   2.3232

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.5374727  0.3576465   1.503 0.132889
age.f1      -0.5794525  0.1768644  -3.276 0.001052 **
age.f2      -0.8955648  0.2394085  -3.741 0.000183 ***
age.f3      -0.9388325  0.2910096  -3.226 0.001255 **
ms.f2        0.2795192  0.2446046   1.143 0.253147
ms.f3        0.5611822  0.3097896   1.811 0.070064 .
ms.f4        1.3477283  0.3273620   4.117 3.84e-05 ***
ms.f5        0.4141109  0.4601644   0.900 0.368163
ms.f6        0.0003085  0.2123345   0.001 0.998841
smok         0.5117437  0.1478247   3.462 0.000537 ***
alc          0.3614774  0.2286011   1.581 0.113819
falc         0.4281975  0.1515403   2.826 0.004719 **
educ.f1     -0.9792360  0.2968630  -3.299 0.000972 ***
educ.f2     -1.3427248  0.3097845  -4.334 1.46e-05 ***
reg.f2      -0.9282729  0.2102969  -4.414 1.01e-05 ***
reg.f3       0.0037677  0.1766905   0.021 0.982987
reg.f4      -0.4348432  0.1863873  -2.333 0.019648 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1561.6  on 1315  degrees of freedom
Residual deviance: 1419.0  on 1299  degrees of freedom
AIC: 1453

Number of Fisher Scoring iterations: 4

> # model0.5 = glm(dv ~ age.f + ms.f + smok + alc + falc + educ.f + reg.f)
>
> # measure the significant of reg.f (because the p-value in reg.f.south=0.898)
> model0.6 <- glm(dv ~ age.f + ms.f + smok + alc + falc + educ.f, family=binomial, data=domviolence)
> anova(model0.6, model0.5, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ age.f + ms.f + smok + alc + falc + educ.f
Model 2: dv ~ age.f + ms.f + smok + alc + falc + educ.f + reg.f
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1302     1448.9
2      1299     1419.0  3   29.927 1.43e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # the p-value of Chi test is 1.43e-06, which means reg.f is significant and we cannot delete this variable
>
> # next we should consider the variable ms
> model0.65 <- glm(dv ~ age.f + smok + alc + falc + educ.f + reg.f, family=binomial, data=domviolence)
> anova(model0.65, model0.5, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ age.f + smok + alc + falc + educ.f + reg.f
Model 2: dv ~ age.f + ms.f + smok + alc + falc + educ.f + reg.f
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      1304     1440.1
2      1299     1419.0  5   21.112 0.0007716 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # p-value=0.0008, means variable ms is significant
>
> #next one is alc (the p-value for alc is the second largest, which is 0.090)
> model0.7 <- glm(dv ~ age.f + ms.f + smok + falc + educ.f + reg.f, family=binomial, data=domviolence)
> anova(model0.7, model0.5, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ age.f + ms.f + smok + falc + educ.f + reg.f
Model 2: dv ~ age.f + ms.f + smok + alc + falc + educ.f + reg.f
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1300     1421.4
2      1299     1419.0  1   2.4539   0.1172
> # the p-value is 0.117, larger than 0.05, so alc can be ignored
>
> # we can also consider age variable for accuracy
> model0.9 <- glm(dv ~ ms.f + smok + falc + educ.f + reg.f, family=binomial, data=domviolence)
> anova(model0.9, model0.7, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ ms.f + smok + falc + educ.f + reg.f
Model 2: dv ~ age.f + ms.f + smok + falc + educ.f + reg.f
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      1303     1439.1
2      1300     1421.4  3   17.645 0.0005206 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # Pr value for these two models are 0.0005, which is smaller than 0.05, so age cannot be ignored
>
> # so model1 is model0.8
> model1 <- model0.7
> summary(model1)

Call:
glm(formula = dv ~ age.f + ms.f + smok + falc + educ.f + reg.f,
    family = binomial, data = domviolence)


Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7848  -0.8195  -0.5870   1.0787   2.3143

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.52027    0.35749   1.455 0.145567
age.f1      -0.56976    0.17674  -3.224 0.001265 **
age.f2      -0.88831    0.23941  -3.710 0.000207 ***
age.f3      -0.92992    0.29064  -3.200 0.001376 **
ms.f2        0.30849    0.24311   1.269 0.204467
ms.f3        0.57716    0.30935   1.866 0.062078 .
ms.f4        1.38101    0.32680   4.226 2.38e-05 ***
ms.f5        0.42905    0.45987   0.933 0.350828
ms.f6        0.03606    0.21089   0.171 0.864219
smok         0.53873    0.14646   3.678 0.000235 ***
falc         0.44970    0.15069   2.984 0.002842 **
educ.f1     -0.97425    0.29679  -3.283 0.001029 **
educ.f2     -1.32296    0.30928  -4.278 1.89e-05 ***
reg.f2      -0.91585    0.20992  -4.363 1.28e-05 ***
reg.f3       0.01977    0.17628   0.112 0.910686
reg.f4      -0.43371    0.18632  -2.328 0.019929 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1561.6  on 1315  degrees of freedom
Residual deviance: 1421.4  on 1300  degrees of freedom
AIC: 1453.4

Number of Fisher Scoring iterations: 4

> # model1 = glm(dv ~ age.f + ms.f + smok + falc + educ.f + reg.f)
>
> #1.(b)
> # change both age and education variables from factor value to normal values
> model_compare <- glm(dv ~ ms.f + smok + falc + reg.f + age + educ + age.f + educ.f,family=binomial, data=domviolence)
```

```
> anova(model_compare, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: dv

Terms added sequentially (first to last)


       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                    1315     1561.6
ms.f    5   42.425      1310     1519.2 4.833e-08 ***
smok    1   23.608      1309     1495.6 1.181e-06 ***
falc    1   10.460      1308     1485.1  0.001220 **
reg.f   3   33.785      1305     1451.4 2.200e-07 ***
age     1    7.391      1304     1444.0  0.006556 **
educ    1   16.081      1303     1427.9 6.068e-05 ***
age.f   2    3.281      1301     1424.6  0.193911
educ.f  1    3.170      1300     1421.4  0.074995 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> model_compare1 <- glm(dv ~ ms.f + reg.f + smok + educ + falc + age + educ.f + age.f, family=binomial, data=domviolence)
> anova(model_compare1, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: dv

Terms added sequentially (first to last)


       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                    1315     1561.6
ms.f    5   42.425      1310     1519.2 4.833e-08 ***
reg.f   3   38.389      1307     1480.8 2.338e-08 ***
smok    1   20.407      1306     1460.4 6.260e-06 ***
educ    1   10.259      1305     1450.2  0.001360 **
falc    1    8.302      1304     1441.8  0.003960 **

age     1   13.967      1303     1427.9  0.000186 ***
educ.f  1    3.493      1302     1424.4  0.061629 .
age.f   2    2.958      1300     1421.4  0.227882
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # p-value for age.f is 0.194 > 0.05, p-value for educ.f is 0.075 > 0.05, so we can delete them
> model2 <- glm(dv ~ ms.f + reg.f + smok + educ + falc + age, family=binomial, data=domviolence)
> summary(model2)

Call:
glm(formula = dv ~ ms.f + reg.f + smok + educ + falc + age, family = binomial,
    data = domviolence)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8416  -0.8351  -0.6002   1.0940   2.3086

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.03066    0.26584  -0.115 0.908172
ms.f2        0.33731    0.24119   1.398 0.161963
ms.f3        0.56094    0.30786   1.822 0.068446 .
ms.f4        1.36151    0.32560   4.181 2.90e-05 ***
ms.f5        0.56923    0.44937   1.267 0.205250
ms.f6        0.18601    0.19852   0.937 0.348776
reg.f2      -0.91205    0.20919  -4.360 1.30e-05 ***
reg.f3       0.04867    0.17486   0.278 0.780745
reg.f4      -0.41621    0.18466  -2.254 0.024199 *
smok         0.51611    0.14554   3.546 0.000391 ***
educ        -0.48947    0.12230  -4.002 6.27e-05 ***
falc         0.42240    0.14963   2.823 0.004759 **
age         -0.33551    0.09131  -3.674 0.000239 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1561.6  on 1315  degrees of freedom
Residual deviance: 1427.9  on 1303  degrees of freedom
AIC: 1453.9
```

```
Number of Fisher Scoring iterations: 4

>
> #1.(c)
> model3 <- glm(dv ~ (age + ms.f + smok + falc + educ + reg.f)^2, family=binomial, data=domviolence)
Warning message:
glm.fit:拟合概率算出来是数值零或一
>
> model3.1 <- step(model3, trace=0)
There were 12 warnings (use warnings() to see them)
> summary(model3.1)

Call:
glm(formula = dv ~ age + ms.f + smok + falc + educ + reg.f +
    ms.f:falc + smok:falc + educ:reg.f, family = binomial, data = domviolence)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9001  -0.8120  -0.6126   1.0027   2.3227

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.13536    0.38978   0.347 0.728387
age         -0.37147    0.09332  -3.980 6.88e-05 ***
ms.f2        0.76595    0.27660   2.769 0.005620 **
ms.f3        0.46111    0.38588   1.195 0.232102
ms.f4        1.31717    0.36560   3.603 0.000315 ***
ms.f5        0.53650    0.50969   1.053 0.292523
ms.f6        0.07792    0.22467   0.347 0.728714
smok         0.67308    0.17019   3.955 7.66e-05 ***
falc         0.63780    0.20800   3.066 0.002167 **
educ        -0.62665    0.25004  -2.506 0.012204 *
reg.f2      -1.14883    0.55007  -2.089 0.036753 *
reg.f3       0.25858    0.44167   0.585 0.558242
reg.f4      -1.29780    0.50056  -2.593 0.009522 **
ms.f2:falc  -1.64084    0.57575  -2.850 0.004373 **
ms.f3:falc   0.50303    0.66004   0.762 0.445985
ms.f4:falc   0.23587    0.82867   0.285 0.775926
ms.f5:falc   0.48348    1.15166   0.420 0.674621

ms.f6:falc   0.24844    0.40581   0.612 0.540400
smok:falc   -0.47591    0.33210  -1.433 0.151841
educ:reg.f2  0.19103    0.37798   0.505 0.613283
educ:reg.f3 -0.15840    0.30731  -0.515 0.606259
educ:reg.f4  0.60461    0.33414   1.809 0.070379 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1561.6  on 1315  degrees of freedom
Residual deviance: 1406.5  on 1294  degrees of freedom
AIC: 1450.5

Number of Fisher Scoring iterations: 4

> # model3.1 = glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc + educ:reg.f)
>
> # consider the interaction between ms and falc
> model3.2 <- glm(dv ~ age + ms.f + smok + falc + educ + reg.f + smok:falc + educ:reg.f, family=binomial, data=domviolence)
> anova(model3.2, model3.1, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ age + ms.f + smok + falc + educ + reg.f + smok:falc + educ:reg.f
Model 2: dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc +
    educ:reg.f
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1299     1418.4
2      1294     1406.5  5   11.843    0.037 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # Pr=0.037 < 0.05, the interaction cannot be ignored, also ms and falc cannot be ignored
>
> # consider the interaction between educ and reg.f
> model3.3 <- glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc, family=binomial, data=domviolence)
> anova(model3.3, model3.1, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc
Model 2: dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc +
```

```
    educ:reg.f
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1297     1413.5
2      1294     1406.5  3   7.0232  0.07116 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # Pr=0.071 > 0.05, the interaction can be ignored
>
> # consider the interaction between smok and falc
> model3.4 <- glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc, family=binomial, data=domviolence)
> anova(model3.4, model3.3, test='Chi')
Analysis of Deviance Table

Model 1: dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc
Model 2: dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc + smok:falc
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1298     1415.1
2      1297     1413.5  1   1.5394   0.2147
> # Pr=0.215 > 0.05, the interaction can be ignored
>
> # get model4 glm(dv ~ age + ms.f + smok + falc + educ + reg.f + ms.f:falc)
> model4 <- model3.4
> anova(model4, test='Chi')
Analysis of Deviance Table

Model: binomial, link: logit

Response: dv

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                      1315     1561.6
age        1   23.321      1314     1538.3 1.371e-06 ***
ms.f       5   31.207      1309     1507.1 8.526e-06 ***
smok       1   19.669      1308     1487.4 9.210e-06 ***
falc       1   10.275      1307     1477.2  0.001348 **
educ       1   18.262      1306     1458.9 1.926e-05 ***


reg.f      3   31.016      1303     1427.9 8.435e-07 ***
ms.f:falc  5   12.791      1298     1415.1  0.025418 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(model4)

Call:
glm(formula = dv ~ age + ms.f + smok + falc + educ + reg.f +
    ms.f:falc, family = binomial, data = domviolence)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9645  -0.8312  -0.5834   1.0333   2.3199

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.03815    0.27058  -0.141 0.887864
age         -0.34707    0.09181  -3.780 0.000157 ***
ms.f2        0.79651    0.27412   2.906 0.003665 **
ms.f3        0.43946    0.38210   1.150 0.250094
ms.f4        1.31189    0.36282   3.616 0.000299 ***
ms.f5        0.48817    0.50342   0.970 0.332192
ms.f6        0.14320    0.22283   0.643 0.520462
smok         0.53324    0.14649   3.640 0.000273 ***
falc         0.52629    0.19063   2.761 0.005766 **
educ        -0.49007    0.12337  -3.972 7.12e-05 ***
reg.f2      -0.90821    0.21067  -4.311 1.63e-05 ***
reg.f3       0.02792    0.17609   0.159 0.874038
reg.f4      -0.42353    0.18623  -2.274 0.022953 *
ms.f2:falc  -1.78134    0.57027  -3.124 0.001786 **
ms.f3:falc   0.32013    0.65570   0.488 0.625388
ms.f4:falc   0.24874    0.83284   0.299 0.765197
ms.f5:falc   0.59486    1.13590   0.524 0.600494
ms.f6:falc   0.11761    0.40123   0.293 0.769425
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1561.6  on 1315  degrees of freedom
```

```
Residual deviance: 1415.1  on 1298  degrees of freedom
AIC: 1451.1

Number of Fisher Scoring iterations: 4

>
> # 2.(1) ms
> coe <- model4$coefficients
> # for falc = 0
> # ms2 vs ms1:
> exp(coe[3]) #2.218
   ms.f2
2.217785
> # ms3 vs ms1:
> exp(coe[4]) #1.552
   ms.f3
1.551876
> # ms4 vs ms1:
> exp(coe[5]) #3.713
   ms.f4
3.713199
> # ms5 vs ms1:
> exp(coe[6]) #1.629
   ms.f5
1.629334
> # ms6 vs ms1:
> exp(coe[7]) #1.154
   ms.f6
1.153958
>
> # for falc = 1
> # ms2 vs ms1:
> exp(coe[3] + coe[14]) #0.374
    ms.f2
0.3735025
> # ms3 vs ms1:
> exp(coe[4] + coe[15]) #2.137
   ms.f3
2.137418


> # ms4 vs ms1:
> exp(coe[5] + coe[16]) #4.762
   ms.f4
4.761826
> # ms5 vs ms1:
> exp(coe[6] + coe[17]) #2.954
   ms.f5
2.953621
> # ms6 vs ms1:
> exp(coe[7] + coe[18]) #1.298
   ms.f6
1.297979
>
> # smoking
> exp(coe[8]) #1.704
    smok
1.704437
>
> # ms/falc = 0 vs falc = 0:
> # 1
>
> # ms/falc = 1 vs falc = 0:
> #ms1-ms6:
> exp(coe[9]) #1.693
    falc
1.692634
> exp(coe[9] + coe[14]) #0.285
     falc
0.2850606
> exp(coe[9] + coe[15]) #2.331
    falc
2.331285
> exp(coe[9] + coe[16]) #2.171
    falc
2.170643
> exp(coe[9] + coe[17]) #3.068
    falc
3.068371
```

```
> exp(coe[9] + coe[18]) #1.904
    falc
1.903886
>
> # region vs north: north:north = 1
> #other three regions
> exp(coe[11]) #0.403
   reg.f2
0.4032442
> exp(coe[12]) #1.028
  reg.f3
1.028309
> exp(coe[13]) #0.655
   reg.f4
0.6547292
>
> # age and educ are different kinds of variables than others
> #age
> exp(coe[2]) #0.707
      age
0.7067593
> #education
> exp(coe[10]) #0.613
     educ
0.6125824
```