**Statisical Modelling in Data Science: Assignment 1**

Chi Yin Wong

836872

**Question 1**

The first step was to rewrite all the categorical predictors using the factor function and then add them to the dataset. Then, a logistic regression model was fitted using all the predictors and captured by the variable model0.

Next, an ANOVA Chi-squared test is performed. A significance level of 5% is used to interpret the results of the hypothesis test. All the predictors have p-value < 0.05 except for alcohol which has a reported p-value of 0.115787. Thus, the alcohol predictor is removed from the model (model1) because it is not statistically significant. The test is repeated on model1 and there are no more predictors that need to be removed.

Next, the model is fitted (model2) with age and education as numerical variables instead of categorical ones and compare this with model1 using an ANOVA Chi – squared test. The reported p-value is 0.06863. Using a 5% significance level, the analysis of deviance table showed that there is no significant difference between the models in terms of adequacy of fit, so model2 is picked for simplicity.

Next, all first order interaction terms are included (model3). Then, a stepwise selection method with AIC is used to determine the final model (model4). Stepwise selection removed the married more than once predictor and its associated interaction terms, as well as many of the other interaction terms. All of them had large AIC values apart from the marital status : family alcohol cause for concern interaction term.

A final ANOVA Chi-squared test is performed on model4 to determine whether there are any more insignificant predictors in the model that should not be included. From the analysis of deviance table, all predictors and the interaction term appear to be significant, so it can be concluded that model4 is the final and “best” model of the form specified in the question sheet.

**Question 2**

Marital Status

|  |  |  |
| --- | --- | --- |
| Odds Ratios, at various levels of falc | falc = 0 | falc = 1 |
| For ms = 1 vs. ms = 1 | 0.9626 | 0.9626 |
| For ms = 2 vs. ms = 1 | 2.2178 | 0.3735 |
| For ms = 3 vs. ms = 1 | 1.5512 | 2.1374 |
| For ms = 4 vs. ms = 1 | 3.7132 | 4.7618 |
| For ms = 5 vs. ms = 1 | 1.6293 | 2.9536 |
| For ms = 6 vs. ms = 1 | 1.1540 | 1.2980 |

Considering that a woman who is currently married (ms = 1) is taken as a baseline, the predicted odds of a woman who is married is given by the intercept term. That is, the predicted odds ratio of domestic violence for a woman who is married, whether or not the woman has family members whose use alcohol has been a cause for concern when growing up, is = 0.9626.

Firstly, the women who grew up and didn’t have family members alcohol use leading to a cause for concern when growing up are considered (falc = 0). Keeping all other variables unchanged, the estimated odds ratio of domestic violence for a woman who is de facto, divorced, separated, widowed and never married against a woman who is married is, respectively, , , , and .

It can be concluded that for a woman who didn’t have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 121.78% for a woman who is in a de Facto relationship against a woman who is married.

For a woman who didn’t have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 55.12% for a woman who is divorced against a woman who is married.

For a woman who didn’t have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 271.32% for a woman who is separated against a woman who is married.

For a woman who didn’t have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 62.93% for a woman who is widowed against a woman who is married.

Finally, for a woman who didn’t have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 15.40% for a woman who has never been married against a woman who is married.

Next, the women who grew up and did have family members alcohol use leading to a cause for concern when growing up are considered (falc = 1). Keeping all other variables constant, the estimated odds ratio of domestic violence for a woman who is de facto, divorced, separated, widowed and never married against a woman who is married is, respectively, , , , and .

Interestingly, for a woman who did have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence actually decreased by 62.65% for a woman who is in a De Facto relationship against a woman who is married.

For a woman who did have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 113.74% for a woman who is divorced against a woman who is married.

For a woman who did have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 376.18% for a woman who is separated against a woman who is married.

For a woman who did have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 195.36% for a woman who is widowed against a woman who is married.

For a woman who did have family members who used alcohol leading to a cause for concern, the estimated odds of domestic violence increased by 29.80% for a woman who has never been married against a woman who is married.

Smoking

The odds ratio is

The estimated odds of domestic violence increased by 70.44% for a woman who smokes against a woman who doesn’t smoke.

Family Alcohol

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Odds ratio at various levels of ms | ms=1 | ms=2 | ms=3 | ms=4 | ms=5 | ms=6 |
| For falc = 0 vs. falc = 0 | 0.9626 | 0.9626 | 0.9626 | 0.9626 | 0.9626 | 0.9626 |
| For falc = 1 vs. falc = 0 | 1.6926 | 0.2851 | 2.3313 | 2.1707 | 3.0684 | 1.9039 |

Considering that a woman who has no family members whose use of alcohol was a cause for concern growing up is used as a baseline, the predicted odds, no matter her current marital status, is given by the intercept term. That is, the predicted logs ratio is given by 0.9626 for everything in the second row. This can be interpreted as follows: the estimated odds ratio of domestic abuse for a woman who doesn’t have family members whose use of alcohol was a cause for concern is 0.9626, regardless of their current marital status.

Keeping all other variables unchanged, the estimated log odds ratio for a woman who had family members that used alcohol causing concern growing up and who is currently married, in a de facto relationship, divorced, separated, widowed and never married, is respectively and .

For a woman who is married, the estimated odds of domestic violence increased by 69.26% for a woman who did have family members whose use of alcohol was a cause for concern growing up as opposed to a woman who didn’t.

For a woman who is in a de facto relationship, the estimated odds of domestic violence decreased by 71.49% for a woman who did have family members whose use of alcohol was a cause for concern growing up as opposed to a woman who didn’t.

For a woman who is divorced, the estimated odds of domestic violence increased by 133.13% for a woman who did have family members whose use of alcohol was a cause for concern growing up as opposed to a woman who didn’t.

For a woman who is separated, the estimated odds of domestic violence increased by 117.07% for a woman who did have family members whose use of alcohol was a cause for concern growing up as opposed to a woman who didn’t.

For a woman who is widowed, the estimated odds of domestic violence increased by 206.84% for a woman who did have family members whose use of alcohol was a cause for concern growing up as opposed to a woman who didn’t.

For a woman who has never been married, the estimated odds of domestic violence increased by 90.39% for a woman who did have family members whose use of alcohol was a cause for concern growing up as opposed to a woman who didn’t.

Region

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | North | East | South | West |
| Ratio of odds for each region vs. north | 0.9626 | 0.4032 | 1.028 | 0.6547 |

Considering that the northern region is taken as a baseline, the predicted odds is just given by the intercept term. That is, the predicted odds ratio of domestic violence for a woman in the northern region is 0.9626. Keeping all the other variables unchanged except for the region, the estimated odds ratio of domestic violence for a woman living in the eastern, southern and western region against a woman living in the northern region is respectively , and .

It can be concluded that the estimated odds of domestic abuse for a woman living in the eastern region is decreased by 59.68% against a woman that lives in the northern region.

Similarly, the estimated odds of domestic abuse for a woman living in the southern region is increased by 2.8% against a woman that lives in the northern region.

Finally, the estimated odds of domestic abuse for a woman living in the western region is decreased by 34.53% against a woman that lives in the northern region.

Age

The odds ratio is

The estimated odds of domestic violence increases by (0.7068 – 1)\*100 = -29.32% for each increase in age by one factor unit. It can be concluded that the estimated odds of domestic violence for a woman in the 30 – 49 age range decreases by 29.32% against a woman in the 18-29 age range. Similarly, the estimated odds for a woman in the 50-64 age range decreases by 29.32% against a woman in the 30-49 age range, and the estimated odds for a woman in the 65+ range decreases by 29.32% against a woman in the 50-64 range.

Education

The odds ratio is

The odds of domestic violence increases by (0.6126 – 1)\*100 = -38.74% for each increase in education by one factor unit. It can be concluded that the estimated odds of domestic violence for a woman who has had between 7-11 years of education decreases by 38.74% against a woman who has had less than 6 years of education. Similarly, the estimated odds for a woman who has had more than 12 years of education decreases by 38.74% against a woman who has had between 7-11 years of education.

**Appendix: R code**

Assignment1 Rcode

Chi Yin Wong

29/03/2021

setwd("~/Google Drive/Unimelb/Masters/Statistical Modelling for Data Science/Assignment 1")  
  
domviolence = read.csv("domviolence.csv")  
dim(domviolence)

## [1] 1316 9

summary(domviolence)

## age ms mmo smok   
## Min. :0.0 Min. :1.000 Min. :0.000 Min. :0.000   
## 1st Qu.:0.0 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:0.000   
## Median :1.0 Median :1.000 Median :1.000 Median :0.000   
## Mean :1.1 Mean :2.169 Mean :0.804 Mean :0.253   
## 3rd Qu.:2.0 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:1.000   
## Max. :3.0 Max. :6.000 Max. :1.000 Max. :1.000   
## alc falc educ reg   
## Min. :0.00000 Min. :0.0000 Min. :0.000 Min. :1.000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:2.000   
## Median :0.00000 Median :0.0000 Median :1.000 Median :3.000   
## Mean :0.08131 Mean :0.2158 Mean :1.432 Mean :2.606   
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:2.000 3rd Qu.:4.000   
## Max. :1.00000 Max. :1.0000 Max. :2.000 Max. :4.000   
## dv   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.2804   
## 3rd Qu.:1.0000   
## Max. :1.0000

# First step is to rewrite all the predictors as categorical variables i.e. use the factor function  
domviolence$age.f = factor(domviolence$age)  
domviolence$ms.f = factor(domviolence$ms)  
domviolence$mmo.f = factor(domviolence$mmo)  
domviolence$smok.f = factor(domviolence$smok)  
domviolence$alc.f = factor(domviolence$alc)  
domviolence$falc.f = factor(domviolence$falc)  
domviolence$educ.f = factor(domviolence$educ)  
domviolence$reg.f = factor(domviolence$reg)  
domviolence$dv.f = factor(domviolence$dv)  
  
model0 = glm(dv.f ~ age.f + ms.f + mmo.f + smok.f + alc.f + falc.f + educ.f + reg.f, family = binomial, data = domviolence)  
anova(model0, test = "Chi")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: dv.f  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 1315 1561.6   
## age.f 3 26.1373 1312 1535.5 8.926e-06 \*\*\*  
## ms.f 5 31.3925 1307 1504.1 7.835e-06 \*\*\*  
## mmo.f 1 4.0785 1306 1500.0 0.043431 \*   
## smok.f 1 17.9658 1305 1482.1 2.249e-05 \*\*\*  
## alc.f 1 2.4734 1304 1479.6 0.115787   
## falc.f 1 9.7522 1303 1469.8 0.001791 \*\*   
## educ.f 2 23.4457 1301 1446.4 8.106e-06 \*\*\*  
## reg.f 3 28.7213 1298 1417.7 2.563e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# alc insignificant, remove alc  
model1 = glm(dv.f ~ age.f + ms.f + mmo.f + smok.f + falc.f + educ.f + reg.f, family = binomial, data = domviolence)  
anova(model1, test = "Chi")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: dv.f  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 1315 1561.6   
## age.f 3 26.1373 1312 1535.5 8.926e-06 \*\*\*  
## ms.f 5 31.3925 1307 1504.1 7.835e-06 \*\*\*  
## mmo.f 1 4.0785 1306 1500.0 0.043431 \*   
## smok.f 1 17.9658 1305 1482.1 2.249e-05 \*\*\*  
## falc.f 1 10.5232 1304 1471.5 0.001179 \*\*   
## educ.f 2 22.6593 1302 1448.9 1.201e-05 \*\*\*  
## reg.f 3 28.7468 1299 1420.1 2.531e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# all predictors significant, nothing to remove  
# now we replace factor(age) and factor(educ) with age and educ (treating them as numerical)  
model2 = glm(dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f, family = binomial, data = domviolence)  
anova(model2, model1, test = "Chi")

## Analysis of Deviance Table  
##   
## Model 1: dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f  
## Model 2: dv.f ~ age.f + ms.f + mmo.f + smok.f + falc.f + educ.f + reg.f  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 1302 1427.2   
## 2 1299 1420.1 3 7.1049 0.06863 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# We see that the two models are not sigificantly different, thus we use model 2 as it is simpler in terms of model complexity  
# now we expand model 2 by including all the first order interaction terms   
model3 = glm(dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f + ms.f:falc.f + ms.f:mmo.f + ms.f:smok.f + ms.f:falc.f + ms.f:reg.f + mmo.f:smok.f + mmo.f:reg.f + smok.f:falc.f + smok.f:reg.f + falc.f:reg.f, family = binomial, data = domviolence)  
summary(model3)

##   
## Call:  
## glm(formula = dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ +   
## reg.f + ms.f:falc.f + ms.f:mmo.f + ms.f:smok.f + ms.f:falc.f +   
## ms.f:reg.f + mmo.f:smok.f + mmo.f:reg.f + smok.f:falc.f +   
## smok.f:reg.f + falc.f:reg.f, family = binomial, data = domviolence)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8154 -0.8153 -0.5822 0.9883 2.4737   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.63379 0.52147 -1.215 0.224220   
## age -0.36206 0.09581 -3.779 0.000158 \*\*\*  
## ms.f2 1.48329 0.67403 2.201 0.027762 \*   
## ms.f3 0.98281 0.87396 1.125 0.260781   
## ms.f4 1.26864 0.94362 1.344 0.178809   
## ms.f5 -11.86397 495.87740 -0.024 0.980912   
## ms.f6 -0.29941 0.65973 -0.454 0.649948   
## mmo.f1 0.39704 0.44899 0.884 0.376540   
## smok.f1 1.54514 0.46819 3.300 0.000966 \*\*\*  
## falc.f1 1.16140 0.33231 3.495 0.000474 \*\*\*  
## educ -0.49992 0.12652 -3.951 7.77e-05 \*\*\*  
## reg.f2 -1.01596 0.74036 -1.372 0.169989   
## reg.f3 0.81085 0.51505 1.574 0.115414   
## reg.f4 0.15032 0.55054 0.273 0.784814   
## ms.f2:falc.f1 -1.82729 0.62313 -2.932 0.003363 \*\*   
## ms.f3:falc.f1 0.47447 0.73613 0.645 0.519225   
## ms.f4:falc.f1 0.24147 0.85981 0.281 0.778832   
## ms.f5:falc.f1 0.23891 1.23615 0.193 0.846746   
## ms.f6:falc.f1 0.30226 0.42219 0.716 0.474029   
## ms.f2:mmo.f1 -0.03421 0.52897 -0.065 0.948433   
## ms.f3:mmo.f1 -0.81970 0.72870 -1.125 0.260640   
## ms.f4:mmo.f1 0.18087 0.75340 0.240 0.810271   
## ms.f5:mmo.f1 13.31689 495.87667 0.027 0.978575   
## ms.f6:mmo.f1 0.07549 0.51022 0.148 0.882373   
## ms.f2:smok.f1 -0.47924 0.52032 -0.921 0.357023   
## ms.f3:smok.f1 -0.19889 0.67948 -0.293 0.769749   
## ms.f4:smok.f1 -0.28050 0.74732 -0.375 0.707409   
## ms.f5:smok.f1 0.06441 1.46049 0.044 0.964821   
## ms.f6:smok.f1 -0.30208 0.37066 -0.815 0.415088   
## ms.f2:reg.f2 -0.05115 0.86744 -0.059 0.952981   
## ms.f3:reg.f2 0.43531 1.29138 0.337 0.736048   
## ms.f4:reg.f2 0.44942 1.12745 0.399 0.690178   
## ms.f5:reg.f2 -0.95111 1.50489 -0.632 0.527378   
## ms.f6:reg.f2 0.55770 0.59140 0.943 0.345670   
## ms.f2:reg.f3 -0.74290 0.68942 -1.078 0.281230   
## ms.f3:reg.f3 -0.41931 0.92143 -0.455 0.649066   
## ms.f4:reg.f3 0.07435 1.06260 0.070 0.944221   
## ms.f5:reg.f3 -0.89408 1.16327 -0.769 0.442132   
## ms.f6:reg.f3 0.32137 0.50559 0.636 0.525019   
## ms.f2:reg.f4 -0.91425 0.77234 -1.184 0.236518   
## ms.f3:reg.f4 -0.26049 1.01988 -0.255 0.798405   
## ms.f4:reg.f4 -0.37407 0.93482 -0.400 0.689047   
## ms.f5:reg.f4 -1.34375 1.52436 -0.882 0.378039   
## ms.f6:reg.f4 0.74045 0.49914 1.483 0.137957   
## mmo.f1:smok.f1 -0.59568 0.37431 -1.591 0.111516   
## mmo.f1:reg.f2 0.22956 0.69786 0.329 0.742193   
## mmo.f1:reg.f3 -0.47958 0.48995 -0.979 0.327662   
## mmo.f1:reg.f4 -0.35402 0.51802 -0.683 0.494340   
## smok.f1:falc.f1 -0.54640 0.34713 -1.574 0.115474   
## smok.f1:reg.f2 -0.25900 0.50097 -0.517 0.605157   
## smok.f1:reg.f3 -0.48869 0.40722 -1.200 0.230107   
## smok.f1:reg.f4 -0.24674 0.41016 -0.602 0.547467   
## falc.f1:reg.f2 -0.32242 0.49193 -0.655 0.512192   
## falc.f1:reg.f3 -0.81031 0.40439 -2.004 0.045096 \*   
## falc.f1:reg.f4 -0.78984 0.44677 -1.768 0.077075 .   
## ---  
## 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: 1391.6 on 1261 degrees of freedom  
## AIC: 1501.6  
##   
## Number of Fisher Scoring iterations: 13

# now we check to see if the model can be simplified  
model4 = step(model3)

## Start: AIC=1501.62  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + ms.f:mmo.f + ms.f:smok.f + ms.f:falc.f + ms.f:reg.f +   
## mmo.f:smok.f + mmo.f:reg.f + smok.f:falc.f + smok.f:reg.f +   
## falc.f:reg.f  
##   
## Df Deviance AIC  
## - ms.f:reg.f 15 1399.3 1479.3  
## - ms.f:smok.f 5 1392.9 1492.9  
## - ms.f:mmo.f 5 1394.3 1494.3  
## - smok.f:reg.f 3 1393.1 1497.1  
## - mmo.f:reg.f 3 1393.5 1497.5  
## - falc.f:reg.f 3 1396.7 1500.7  
## <none> 1391.6 1501.6  
## - smok.f:falc.f 1 1394.1 1502.1  
## - mmo.f:smok.f 1 1394.2 1502.2  
## - ms.f:falc.f 5 1404.1 1504.1  
## - age 1 1406.4 1514.4  
## - educ 1 1407.3 1515.3  
##   
## Step: AIC=1479.33  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + ms.f:mmo.f + ms.f:smok.f + mmo.f:smok.f + mmo.f:reg.f +   
## smok.f:falc.f + smok.f:reg.f + falc.f:reg.f  
##   
## Df Deviance AIC  
## - ms.f:smok.f 5 1400.8 1470.8  
## - ms.f:mmo.f 5 1402.2 1472.2  
## - mmo.f:reg.f 3 1400.3 1474.3  
## - smok.f:reg.f 3 1400.7 1474.7  
## - falc.f:reg.f 3 1404.2 1478.2  
## <none> 1399.3 1479.3  
## - mmo.f:smok.f 1 1401.5 1479.5  
## - smok.f:falc.f 1 1401.5 1479.5  
## - ms.f:falc.f 5 1410.7 1480.7  
## - age 1 1413.9 1491.9  
## - educ 1 1416.1 1494.1  
##   
## Step: AIC=1470.79  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + ms.f:mmo.f + mmo.f:smok.f + mmo.f:reg.f + smok.f:falc.f +   
## smok.f:reg.f + falc.f:reg.f  
##   
## Df Deviance AIC  
## - ms.f:mmo.f 5 1404.1 1464.1  
## - mmo.f:reg.f 3 1401.8 1465.8  
## - smok.f:reg.f 3 1402.3 1466.3  
## - falc.f:reg.f 3 1405.5 1469.5  
## - mmo.f:smok.f 1 1402.6 1470.6  
## - smok.f:falc.f 1 1402.8 1470.8  
## <none> 1400.8 1470.8  
## - ms.f:falc.f 5 1412.7 1472.7  
## - age 1 1416.0 1484.0  
## - educ 1 1417.3 1485.3  
##   
## Step: AIC=1464.14  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + mmo.f:smok.f + mmo.f:reg.f + smok.f:falc.f +   
## smok.f:reg.f + falc.f:reg.f  
##   
## Df Deviance AIC  
## - smok.f:reg.f 3 1405.5 1459.5  
## - mmo.f:reg.f 3 1405.8 1459.8  
## - falc.f:reg.f 3 1408.9 1462.9  
## - mmo.f:smok.f 1 1405.9 1463.9  
## - smok.f:falc.f 1 1406.0 1464.0  
## <none> 1404.1 1464.1  
## - ms.f:falc.f 5 1416.5 1466.5  
## - age 1 1419.8 1477.8  
## - educ 1 1420.4 1478.4  
##   
## Step: AIC=1459.46  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + mmo.f:smok.f + mmo.f:reg.f + smok.f:falc.f +   
## falc.f:reg.f  
##   
## Df Deviance AIC  
## - mmo.f:reg.f 3 1406.9 1454.9  
## - falc.f:reg.f 3 1410.2 1458.2  
## - mmo.f:smok.f 1 1407.0 1459.0  
## - smok.f:falc.f 1 1407.2 1459.2  
## <none> 1405.5 1459.5  
## - ms.f:falc.f 5 1417.7 1461.7  
## - age 1 1421.2 1473.2  
## - educ 1 1422.1 1474.1  
##   
## Step: AIC=1454.89  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + mmo.f:smok.f + smok.f:falc.f + falc.f:reg.f  
##   
## Df Deviance AIC  
## - falc.f:reg.f 3 1411.4 1453.4  
## - mmo.f:smok.f 1 1408.3 1454.3  
## - smok.f:falc.f 1 1408.7 1454.7  
## <none> 1406.9 1454.9  
## - ms.f:falc.f 5 1419.1 1457.1  
## - age 1 1422.3 1468.3  
## - educ 1 1423.2 1469.2  
##   
## Step: AIC=1453.44  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + mmo.f:smok.f + smok.f:falc.f  
##   
## Df Deviance AIC  
## - mmo.f:smok.f 1 1412.9 1452.9  
## - smok.f:falc.f 1 1413.2 1453.2  
## <none> 1411.4 1453.4  
## - ms.f:falc.f 5 1423.4 1455.4  
## - age 1 1426.6 1466.6  
## - educ 1 1427.5 1467.5  
## - reg.f 3 1439.3 1475.3  
##   
## Step: AIC=1452.93  
## dv.f ~ age + ms.f + mmo.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f + smok.f:falc.f  
##   
## Df Deviance AIC  
## - mmo.f 1 1413.5 1451.5  
## - smok.f:falc.f 1 1414.5 1452.5  
## <none> 1412.9 1452.9  
## - ms.f:falc.f 5 1424.5 1454.5  
## - age 1 1427.7 1465.7  
## - educ 1 1429.0 1467.0  
## - reg.f 3 1440.7 1474.7  
##   
## Step: AIC=1451.55  
## dv.f ~ age + ms.f + smok.f + falc.f + educ + reg.f + ms.f:falc.f +   
## smok.f:falc.f  
##   
## Df Deviance AIC  
## - smok.f:falc.f 1 1415.1 1451.1  
## <none> 1413.5 1451.5  
## - ms.f:falc.f 5 1425.3 1453.3  
## - age 1 1428.2 1464.2  
## - educ 1 1429.5 1465.5  
## - reg.f 3 1442.1 1474.1  
##   
## Step: AIC=1451.09  
## dv.f ~ age + ms.f + smok.f + falc.f + educ + reg.f + ms.f:falc.f  
##   
## Df Deviance AIC  
## <none> 1415.1 1451.1  
## - ms.f:falc.f 5 1427.9 1453.9  
## - smok.f 1 1428.1 1462.1  
## - age 1 1429.9 1463.9  
## - educ 1 1430.9 1464.9  
## - reg.f 3 1444.4 1474.4

summary(model4)

##   
## Call:  
## glm(formula = dv.f ~ age + ms.f + smok.f + falc.f + educ + reg.f +   
## ms.f:falc.f, 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.f1 0.53324 0.14649 3.640 0.000273 \*\*\*  
## falc.f1 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.f1 -1.78134 0.57027 -3.124 0.001786 \*\*   
## ms.f3:falc.f1 0.32013 0.65570 0.488 0.625388   
## ms.f4:falc.f1 0.24874 0.83284 0.299 0.765197   
## ms.f5:falc.f1 0.59486 1.13590 0.524 0.600494   
## ms.f6:falc.f1 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

# check to see if model still has any insignificant variables  
anova(model4, test = "Chi")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: dv.f  
##   
## 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.f 1 19.669 1308 1487.4 9.210e-06 \*\*\*  
## falc.f 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.f 5 12.791 1298 1415.1 0.025418 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1