**Assignment 3**
**Course: STAT229**
**Name: Zihao Huang (Simon)**
**Date: 2/13/2017**

4

b)

```
Call:
glm(formula = DEATH ~ AGE + INH_INJ + AGE:INH_INJ, family = "binomial",
    data = BURN1000)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6605  -0.3861  -0.1314  -0.0583   3.5877

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.51007    0.52541 -12.390  < 2e-16 ***
AGE          0.08434    0.00828  10.186  < 2e-16 ***
INH_INJ      6.21215    0.65600   9.470  < 2e-16 ***
AGE:INH_INJ -0.06889    0.01164  -5.918 3.27e-09 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 845.42  on 999  degrees of freedom
Residual deviance: 498.34  on 996  degrees of freedom
AIC: 506.34

Number of Fisher Scoring iterations: 7
```

Table 4.1

```
             (Intercept)           AGE      INH_INJ   AGE:INH_INJ
(Intercept)   0.27605389 -4.188610e-03 -0.276053893  4.188610e-03
AGE          -0.00418861  6.855275e-05  0.004188610 -6.855275e-05
INH_INJ      -0.27605389  4.188610e-03  0.430335525 -7.015496e-03
AGE:INH_INJ   0.00418861 -6.855275e-05 -0.007015496  1.355111e-04
```

Table 4.2 covariance of variables

The four-step method is:

(INH_INJ=1,age)-(INH_INJ=0,age) = exp (beta1(1)+beta3(1)*age)

OR(age=20) = e^(6.212-20*0.069) = 125.462

OR(age=40) = e^(6.212-40*0.069) = 31.563

OR(age=60) = e^(6.212-60*0.069) = 7.941

OR(age=80) = e^(6.212-80*0.069) = 1.998

Se(beta1(1) + beta3(1) * age)= var(beta(1))+age^2*var(beta3)+2*age*cov(beta1,beta3)

Thus,95% confident interval is:

```
  Age Odds Ratio    95% confident interval
1  20     125.77 51.902             304.761
2  40      31.71 17.854              56.330
3  60       8.00  4.653              13.742
4  80       2.02  0.888               4.579
```

Table 4.3, 95% confident interval for estimations of age

c)

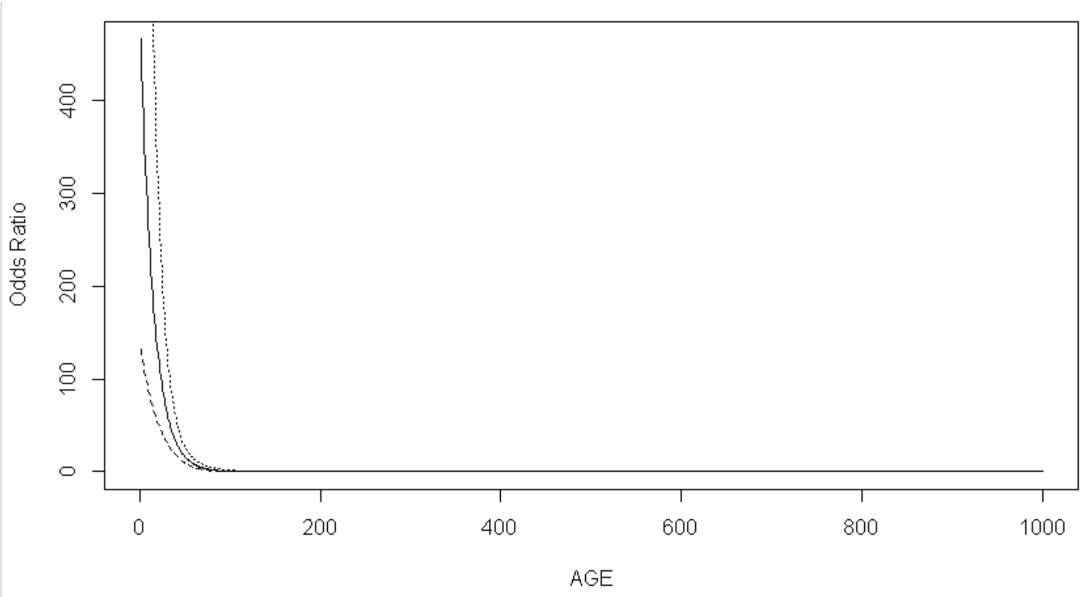



Table 4.4 inhalation injury as a function of age

5.

First the HOURS is created by summing the SPORTHR, READHR, COMPHR, STUDYHR and TVHR

```
Call:
glm(formula = MYOPIC ~ GENDER + HOURS, family = "binomial", data = myopia)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-0.6316  -0.5700  -0.5072  -0.4674   2.2501  

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept) -1.812237   0.326131  -5.557 2.75e-08 ***
GENDER       0.340057   0.241915   1.406    0.160    
HOURS       -0.009642   0.009916  -0.972    0.331    
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 480.08  on 617  degrees of freedom
Residual deviance: 476.77  on 615  degrees of freedom
AIC: 482.77

Number of Fisher Scoring iterations: 4
```

Table 5.1

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6607  -0.5729  -0.5419  -0.4266   2.5577

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.29349    0.42952  -3.012   0.0026 **
GENDER       -0.59719    0.57685  -1.035   0.3005
HOURS        -0.02937    0.01549  -1.897   0.0579 .
GENDER:HOURS  0.03603    0.02037   1.769   0.0768 .
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 480.08  on 617  degrees of freedom
Residual deviance: 473.56  on 614  degrees of freedom
AIC: 481.56

Number of Fisher Scoring iterations: 5
```

Table 5.2

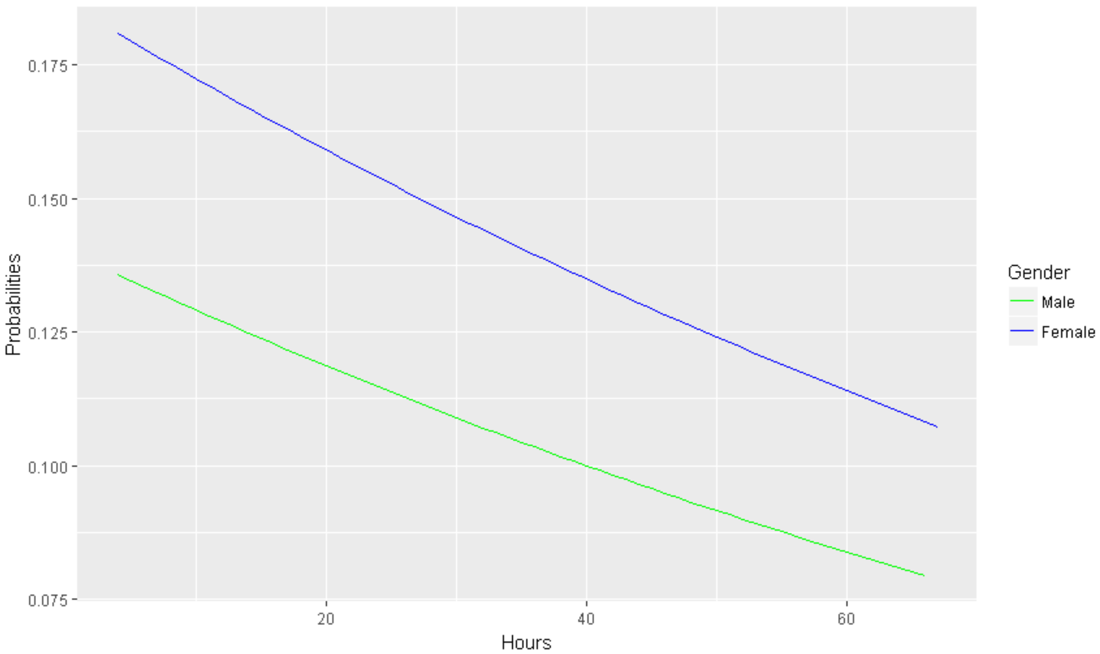



Table 5.3 Probabilities for Male and Female becoming myopic

Here shows that Male is not more likely to have myopic than Female.

Assignment3

1. The power 2 model is chosen on the Burn data.

```
Call:
glm(formula = DEATH ~ AGEFP1 + TBSAFP1 + RACEC + INH_INJ, family = "binomial",
    data = burn)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.97653  -0.20373  -0.08907  -0.04407   3.05707

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.957085   0.596694 -13.335  < 2e-16 ***
AGEFP1       0.086708   0.008245  10.516  < 2e-16 ***
TBSAFP1      0.936365   0.087402  10.713  < 2e-16 ***
RACEC       -0.608872   0.309589  -1.967   0.0492 *
INH_INJ      1.433392   0.342153   4.189  2.8e-05 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 845.42  on 999  degrees of freedom
Residual deviance: 321.94  on 995  degrees of freedom
AIC: 331.94

Number of Fisher Scoring iterations: 7
```

Table 3.1.1 selection of power 2 model parameters on the Burn data

Here,

AGEFP1 = (AGE / 10)^2

TBSAFP1 = sqrt(TBSA)

```
glm(formula = DEATH ~ AGEFP1 + TBSAFP1 + RACEC + INH_INJ + AGEFP1
    INH_INJ, family = "binomial", data = burn)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.73588  -0.19395  -0.07796  -0.03850   2.96643

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -8.215445   0.631340 -13.013  < 2e-16 ***
AGEFP1          0.096068   0.009583  10.025  < 2e-16 ***
TBSAFP1         0.912097   0.087762  10.393  < 2e-16 ***
RACEC          -0.622957   0.310032  -2.009   0.0445 *
INH_INJ         2.420008   0.545173   4.439 9.04e-06 ***
AGEFP1:INH_INJ -0.034088   0.014515  -2.349   0.0188 *
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 845.42  on 999  degrees of freedom
Residual deviance: 316.55  on 994  degrees of freedom
AIC: 328.55

Number of Fisher Scoring iterations: 7
```

3.

First, we run univariable analysis for each variables to find out whether they are significant at 25% significant level.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.09139    0.76590  -2.731  0.00632 **
AGE          0.14376    0.05329   2.698  0.00698 **
---
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.75769    0.31389  -5.600 2.15e-08 ***
BEHAV        0.29018    0.05012   5.789 7.07e-09 ***
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.98426    0.15262  -6.449 1.13e-10 ***
LOS          0.06564    0.01065   6.161 7.22e-10 ***
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -26.57   22128.47  -0.001    0.999
PLACE31        53.13   38278.41   0.001    0.999
PLACE32        53.13   39438.81   0.001    0.999
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2271     0.1327  -1.711   0.0871 .
RACE          0.3423     0.1790   1.912   0.0559 .
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.1904     0.1214  -1.568   0.1168
GENDER        0.3273     0.1786   1.832   0.0669 .
---
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.1144     0.1071  -1.068    0.285
NEURO1        0.2877     0.2474   1.163    0.245
NEURO2        0.1834     0.3867   0.474    0.635
NEURO3        0.1978     0.3081   0.642    0.521
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.1701     0.1018  -1.671  0.09469 .
EMOT          0.5616     0.2128   2.639  0.00831 **
---
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2528     0.3273  -3.827 0.000130 ***
DANGER1       1.4315     0.3835   3.733 0.000189 ***
DANGER2       1.4951     0.3687   4.055 5.01e-05 ***
DANGER3       1.1961     0.3550   3.369 0.000753 ***
---
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2288     0.1132  -2.021  0.04330 *
ELOPE         0.5013     0.1845   2.717  0.00659 **
---

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.0365     0.1278  -8.113 4.94e-16 ***
CUSTD         2.9296     0.2497  11.733  < 2e-16 ***
---
```

Table 3.3.1 univariable analysis for each variable

From table 3.3.1 we found NEURO and PLACE3 are not significant, thus we have a multivariable analysis for other significant variables.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1901  -0.5571  -0.2023   0.4748   2.4748

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.54089    1.45440  -5.185 2.16e-07 ***
AGE          0.17649    0.08378   2.107  0.03515 *
BEHAV        0.25585    0.09810   2.608  0.00911 **
LOS          0.07482    0.01352   5.533 3.14e-08 ***
RACE         0.69803    0.27608   2.528  0.01146 *
GENDER       0.51187    0.30546   1.676  0.09378 .
NEURO.1      0.16742    0.35729   0.469  0.63938
EMOT         0.44776    0.32869   1.362  0.17311
DANGER1      1.02227    0.62475   1.636  0.10178
DANGER2      0.46298    0.69615   0.665  0.50601
DANGER3      0.34618    0.68149   0.508  0.61147
ELOPE       -0.02616    0.29826  -0.088  0.93012
CUSTD        3.43088    0.30961  11.081  < 2e-16 ***
VIOL        -0.11684    0.45465  -0.257  0.79719
---
```

Table 3.3.2 multivariable analysis for significant variables

From table 3.3.2 we found AGE, BEHAV, LOS, RACE, GENDER, CUSTD are significant,

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.06221    1.37833  -5.124 3.00e-07 ***
AGE          0.19593    0.08218   2.384  0.01712 *
BEHAV        0.22694    0.07207   3.149  0.00164 **
LOS          0.07507    0.01330   5.643 1.67e-08 ***
RACE         0.64793    0.27115   2.390  0.01687 *
CUSTD        3.48410    0.30057  11.592  < 2e-16 ***
GENDER       0.42622    0.28505   1.495  0.13484
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
```

Table 3.3.3 reduced model

From table 3.3.3, GENDER is not significant.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -6.77033    1.36217  -4.970 6.69e-07 ***
AGE          0.18003    0.08163   2.205  0.02743 *  
BEHAV        0.25539    0.06954   3.673  0.00024 ***
LOS          0.07441    0.01316   5.656 1.55e-08 ***
RACE         0.62555    0.26959   2.320  0.02032 *  
CUSTD        3.43495    0.29513  11.639  < 2e-16 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
```

Table 3.3.4 reduced model

Then we check the confounder of the variables

AGE: 2.0057794%

BEHAV: -0.1797928%

LOS: -0.5479818%

RACE: -10.3835079%

CUSTD: 0.1186285%

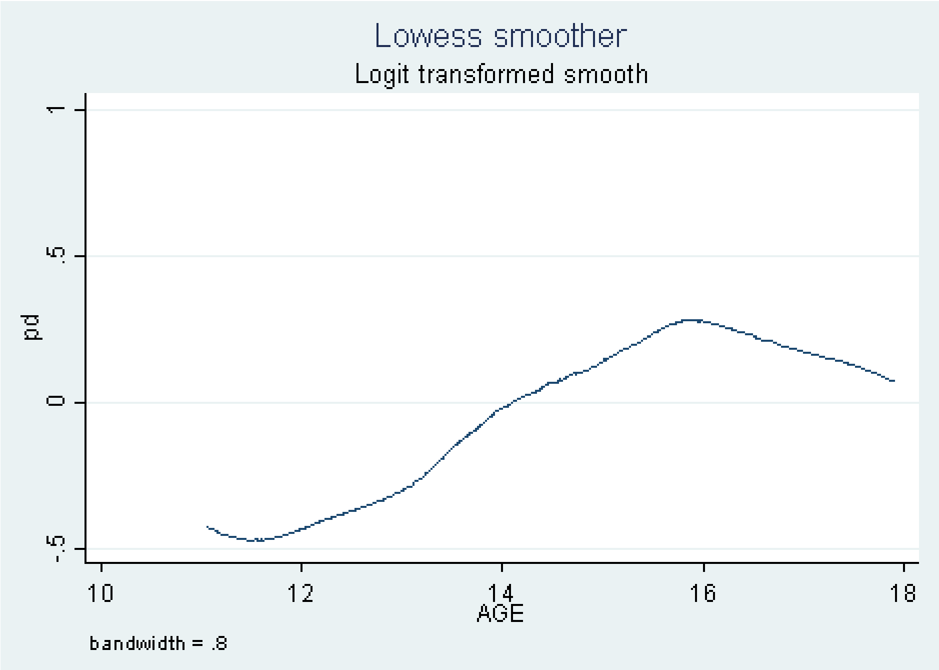
It shows none of the parameters are bigger than 20%, thus, there is no confounder. Then we add the variables which are not significant in the 1st step into this model.

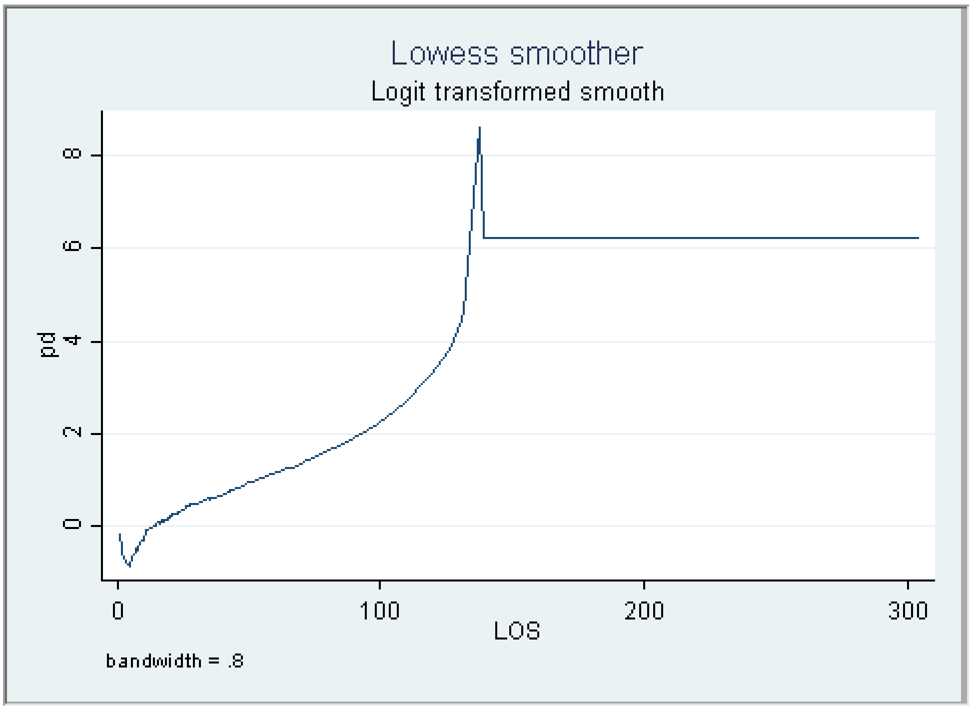
```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -6.95902    1.36949  -5.081 3.75e-07 ***
AGE          0.17165    0.08155   2.105 0.035309 *  
BEHAV        0.26063    0.07036   3.704 0.000212 ***
LOS          0.07290    0.01302   5.598 2.17e-08 ***
RACE         0.62176    0.27431   2.267 0.023413 *  
CUSTD        3.59239    0.30931  11.614  < 2e-16 ***
NEURO1       0.38894    0.36539   1.064 0.287122    
NEURO2       0.68738    0.52172   1.318 0.187664    
NEURO3       1.11993    0.42998   2.605 0.009197 ** 
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 704.04  on 507  degrees of freedom
Residual deviance: 364.84  on 499  degrees of freedom
AIC: 382.84
```

Table 3.3.5

We found NEURO3 is significant.

Then check the continuous variables by using Lowess Smooth.


Lowess smoother
Logit transformed smooth
pd
1
.5
0
-.5
10
12
14
16
18
AGE
bandwidth = .8


Lowess smoother
Logit transformed smooth
pd
8
6
4
2
0
0
100
200
300
LOS
bandwidth = .8

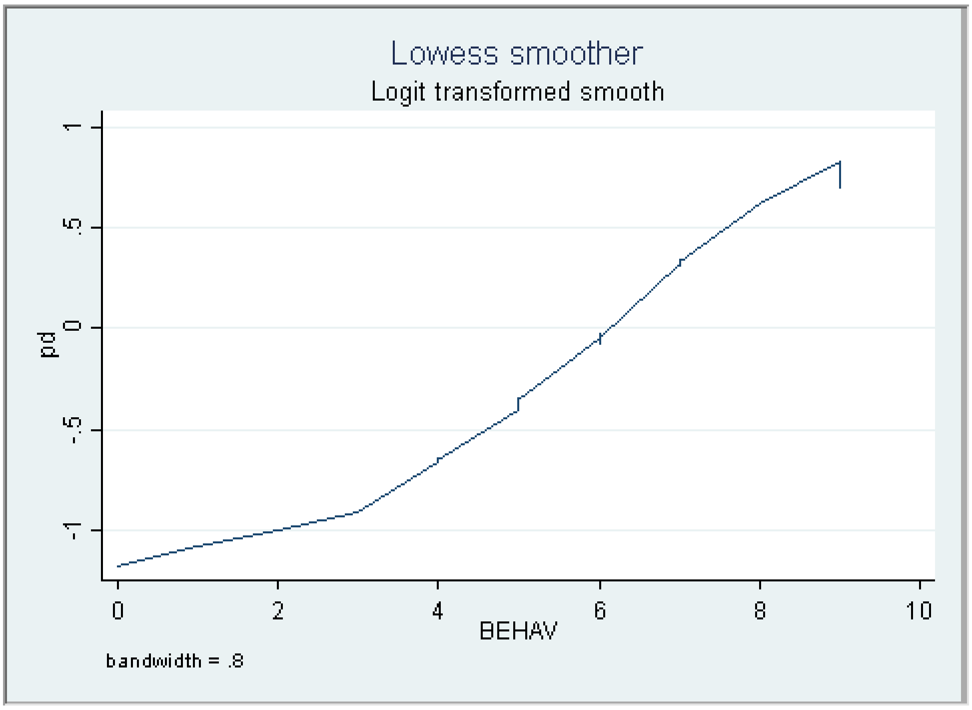



Table 3.3.6 Checking the linearity of three variables

```
Coefficients:
                 Estimate Std. Error z value Pr(>|z|)    
(Intercept)       -7.5799     1.4659  -5.171 2.33e-07 ***
AGE.s              2.0802     0.8550   2.433  0.01498 *  
BEHAV.s            2.3369     0.7672   3.046  0.00232 ** 
LOS.s              0.9238     0.1671   5.530 3.20e-08 ***
RACE               0.4691     0.2930   1.601  0.10935    
CUSTD              4.2339     0.4365   9.699  < 2e-16 ***
NEURO.n           -3.9787     2.0702  -1.922  0.05462 .  
LOS.s:CUSTD       -0.6442     0.2607  -2.470  0.01349 *  
RACE:NEURO.n       2.6525     1.0803   2.455  0.01408 *  
BEHAV.s:NEURO.n    5.2577     2.5346   2.074  0.03805 *  
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 704.04  on 507  degrees of freedom
Residual deviance: 354.53  on 498  degrees of freedom
AIC: 374.53

Number of Fisher Scoring iterations: 7
```

Table 3.3.7

Then we filter the interactions and at last here is the fitted model.