Psy 5011: Categorical Data Analysis **(Due: 28/04/20)**

Homework #5

Instructor: Yao, Kaiping Grace *Name:* Hao-Cheng Lo, *Id:* D08227104

Problem 5.4: The Students data file at the text website shows responses of a class of social science graduate students at the University of Florida to a questionnaire that asked about *gender* (1 = female, 0 = male), *age*, *hsgpa* = high school GPA (on a four-point scale), *cogpa* = college GPA, *dhome* = distance (in miles) of the campus from your home town, *dres* = distance (in miles) of the classroom from your current residence, *tv* = average number of hours perweek that youwatch TV, *sport* = average number of hours per week that you participate in sports or have other physical exercise, *news* = number of times a week you read a newspaper, *aids* = number of people you know who have died from AIDS or who are HIV+, *veg* = whether you are a vegetarian (1 = yes, 0 = no), *affil* = political affiliation (1 = Democrat, 2 = Republican, 3 = Independent), *ideol* = political ideology (1 = very liberal, 2 = liberal, 3 = slightly liberal, 4 = moderate, 5 = slightly conservative, 6 = conservative, 7 = very conservative), *relig* = how often you attend religious services (0 = never, 1 = occasionally, 2 = most weeks, 3 = every week), *abor* = opinion about whether abortion should be legal in the first three months of pregnancy (1 = yes, 0 = no), *affirm* = support affirmative action (1 = yes, 0 = no), and *life* = belief in life after death (1 = yes, 2 = no, 3 = undecided).

(a) Show all steps of a model-selection method such as purposeful selection for choosing a model for predicting *abor*, when the potential explanatory variables are *ideol*, *relig*, *news*, *hsgpa*, and *gender*.

Donate the *ideol* as I, *relig* as R, *news* as N, *hsgpa* as H, and *gender* as G. Note that the analysis performed below treats *ideol* and *relig* as quantitave variables because the **(b)** and **(c)** implies this information. *R code section is provided in supplementary information by the end of this homework.*

Table I Results of fitting several logistic regression models to predict *abor*.

Model	Explanatory Variables	Deviance	df	AIC	Models Compared	Deviance Difference	P-value
1	None	62.7	59	64.7			
2	I	45.5	58	49.5	(2)-(1)	17.3 (df = 1)	<.01
3	R	48.3	58	52.3	(3)-(1)	14.5 (df = 1)	<.01
4	N	55.4	58	59.4	(4)-(1)	7.3 (df = 1)	<.01
5	H	62.6	58	66.6	(5)-(1)	0.1 (df = 1)	.70
6	G	61.6	58	65.6	(6)-(1)	1.1 (df = 1)	.30
7	$I+R+N$	25.2	56	37.8	(7)-(2)	15.7 (df = 2)	<.01
					(7)-(3)	18.5 (df = 2)	<.01
					(7)-(4)	25.6 (df = 2)	<.01
8	$I+R+N+H+G$	25.2	54	37.2	(8)-(7)	4.6 (df = 2)	.10
9	$I+R+N+I\times R+I\times N+R\times N$	24.2	53	38.2	(9)-(7)	5.6 (df = 3)	.10

At step 1, I compare the null model (model 1 in Table I) to models that have *ideol*, *relig*, *news*, *hsgpa*, and *gender* as sole predictors (models 2, 3, 4, 5, and 6). The likelihood-ratio statistics equal the difference in deviances between the null model and each model. These show that *ideol*, *relig*, and *news* are statistically significant. Therefore, after step 1, the purposeful selection process includes *ideol*, *relig*, and *news* as initial explanatory variables, which is model 7 in Table I.

At step 2, backward elimination compares model 7 to models 2, 3, and 4 that remove *ideol* or *relig* or *news* alone. We can find large increases in deviance results from removing *ideol* or *relig* or *news*, so we leave them all.

At step 3, I compare model 8 to model 7 that adds spine *hsgpa* and *gender*, which were not the initially chosen variables. The decrease in deviance is not significant, so we keep *ideol*, *relig*, and *news* as the only predictors.

(a) *Conti.*

At step 4, model 9 adds the interaction between *ideol*, *relig*, and *news*. We implement this by adding three cross-product terms among *ideol*, *relig*, and *news*. The deviance decreases by 5.6 on $df = 3$, which is not significantly better.

The final model developed for diagnostic investigation has solely *ideol*, *relig*, and *news* explanatory variables as main effects. To check the goodness of fit of this ungroup data, we can find that the proposed model is better than others under LR model comparison test.

(b) Using an automated tool such as the stepAIC or bestglm function in R, construct a model to predict *abor*, selecting from the 14 binary and quantitative variables in the data file as explanatory variables.

Here I employ the stepAIC to select from the 14 binary and quantitative variables. *R code section is provided in supplementary information by the end of this homework.* This function regards AIC be the basis of stepwise model selection which is in a backward manner. We start with all 14 potential explanatory variables as main effects. At each step we remove the variable so that AIC decreases the most, until we get to the stage in which AIC increases if we remove any other variables. In the end, we conclude that we have *ideol*, *hsgpa*, and *news* explanatory variables as main effects to predict *abor*.

(c) With $y = veg$ and the 14 binary and quantitative variables in the data file as explanatory variables, show that the likelihood-ratio test of $H_0 : \beta_1 = \cdots = \beta_{14} = 0$ has P-value < 0.001, yet forward selection using Wald tests with 0.05 criterion selects the null model. Explain how this could happen.

Complete R code section and results are provided in supplementary information by the end of this homework.

```
fit11 <- glm(veg ~ gender+age+hsgpa+cogpa+dhome+dres+tv+sport+news+aids+abor+ideol+relig+affirm,
    family = binomial, data = stud)
summary(fit11)
```

As for the likelihood-ratio test, likelihood-ratio test that Y is jointly independent of the 14 explanatory variables simultaneously tests $H_0 : \beta_1 = \cdots = \beta_{14} = 0$. The test statistic is the difference between the null deviance and the residual deviance, which is $50.725 - 26.645$ with $df = 59 - 45$. This shows strong evidence that at least one explanatory variable has an effect.

```
for (x in colnames(stud)){
  print(x)
  print(summary(glm(veg ~ stud[,x], family = binomial, data = stud)))
}
```

However, the test results for individual variables are not that optimistic. Forward selection at its very first step is to check whether each of 14 variables can significantly predict the outcome variable. The results show that none of 14 variables reaches the significant level regarding their coefficients Wald tests. Hence, null model has been selected.

There are two related reasons. First, we can regard LRT for joint independence of the 14 explanatory variables as the test about the overall variance of outcome variable which can be explained by 14 explanatory variables. Similarly, we can also regard LRT for independence of one explanatory variable as the test about the overall variance of outcome variable which can be explained by that one explanatory variable. Hence, we can explain this phenomenon by that any single explanatory variable cannot significantly explained the variance of outcome variable, but the synergy of putting all variables together can significantly explain the variance of outcome variable. Second, we can find that LRT for individual explanatory variables show nothing significant at the 0.05 level, which means their marginal effect is not significant. The P-value for the overall test is small, yet the lack of significance for individual effects is a warning sign of *multicollinearity*.

Problem 5.6: Refer to the previous exercise. The data file also shows responses on whether a person smokes frequently. Software reports model -2 log-likelihood values of 1130.23 with only an intercept term, 1124.86 with also the main effect predictors, and 1119.87 with also all the two-factor interactions.

(a) Write the model for each case and show that the numbers of parameters are 1, 5, and 11.

For the Model 1:

$$\text{logit}[\pi(x)] = \beta_0$$

$|\{\beta_0\}| = 1$, hence, there is 1 parameter.

For the Model 2:

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 Ei + \beta_2 Sn + \beta_3 Tf + \beta_4 Jp$$

$|\{\beta_0, \ldots, \beta_4\}| = 5$, hence, there are 5 parameters.

For the Model 3:

$$\text{logit}[\pi(x)] = \beta_0 + \beta_1 Ei + \beta_2 Sn + \beta_3 Tf + \beta_4 Jp + \beta_5 EiSn + \beta_6 EiTf + \beta_7 EiJp + \beta_8 SnTf + \beta_9 SnJp + \beta_{10} TfJp$$

$|\{\beta_0, \ldots, \beta_{10}\}| = 11$, hence, there are 11 parameters.

(b) Find AIC values. Which of the three models is preferable?

The AIC values of Model 1 is $1130.23 + 2(1) = 1132.23$, of Model 2 is $1124.86 + 2(5) = 1134.86$, and of Model 3 is $1119.87 + 2(11) = 1141.87$. The AIC of model is the smallest is preferable, so Model 1 is preferable.

Problem 5.8: Refer to Table 2.9 on death penalty decisions. Fit a logistic model with the two race predictors. Conduct a residual analysis and interpret.

Table 2.9 Death penalty verdict by defendant's race and victims' race.

Victims' Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percentage Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

Source: M.L. Radelet and G.L. Pierce, *Florida Law Rev.* **43**: 1–34 (1991). Reprinted with permission of the *Florida Law Review*.

```
> vr <- c(1,1,0,0)
> dr <- c(1,0,1,0)
> yes <- c(53,11,0,4)
> no <- c(414,37,16,139)
> fit31 <- glm(yes/(no+yes)~vr+dr, weights = (no+yes), family = binomial)
> summary(fit31)
#Deviance Residuals:
#       1         2         3         4
# 0.02660  -0.06232  -0.60535   0.09379
#
#Coefficients:
#            Estimate Std. Error z value Pr(>|z|)
#(Intercept)  -3.5961     0.5069  -7.094 1.30e-12 ***
#vr            2.4044     0.6006   4.003 6.25e-05 ***
#dr           -0.8678     0.3671  -2.364   0.0181 *
#
#    Null deviance: 22.26591  on 3  degrees of freedom
#Residual deviance:  0.37984  on 1  degrees of freedom
#AIC: 19.3
> cbind(rstandard(fit31,type="pearson"),residuals(fit31,type="pearson"), rstandard(fit31,type="deviance"),residuals(fit31,type="deviance"))
#         [,1]        [,2]       [,3]        [,4]
#1  0.4447328  0.02661777  0.4445101  0.02660444
#2 -0.4447328 -0.06220533 -0.4455807 -0.06232392
#3 -0.4447329 -0.42927568 -0.6271475 -0.60535029
#4  0.4447329  0.09450780  0.4413600  0.09379105
```

After performing the residual analysis, since *df* of deviance is 1, only one absolute value occurs 0.44473 for the standardized residuals. Because the absolute value of standardized (pearson) residual is not greater than 2. The situation of lack of fit is not hazardous.

Problem 5.10: The Lungs data file at the text website summarizes eight studies in China about smoking and lung cancer. Analyze these data and prepare a short report that summarizes your analyses and interpretations.

First, we propose the only *city* effect and no *smoking* effect model: $\text{logit}(\pi) = \alpha + \beta_{city}$.

```
> fit40=glm(Yes/(Yes+No)~factor(City),family=binomial(link=logit),data=lungs,weights=(Yes+No))
> summary(fit40)
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -7.5110  -3.1926  -0.4589   2.0808   6.7543
#
# Coefficients:
#                          Estimate Std. Error z value Pr(>|z|)
# (Intercept)           -2.930e-16  1.115e-01   0.000 1.000000
# factor(City)Harbin     1.854e-16  1.275e-01   0.000 1.000000
# factor(City)Nanchang   3.723e-16  1.686e-01   0.000 1.000000
# factor(City)Nanjing    9.577e-17  1.387e-01   0.000 1.000000
# factor(City)Shanghai  -6.209e-02  1.175e-01  -0.528 0.597167
# factor(City)Shenyang  -7.405e-02  1.182e-01  -0.627 0.530926
# factor(City)Taiyuan   -6.931e-01  1.832e-01  -3.784 0.000154 ***
# factor(City)Zhengzhou  1.898e-16  1.425e-01   0.000 1.000000
#
#     Null deviance: 310.90  on 15  degrees of freedom
# Residual deviance: 288.27  on  8  degrees of freedom
# AIC: 402.12
#
> 1-pchisq(fit40$deviance,fit40$df.residual)
# [1] 0
> cbind(rstandard(fit40,type="pearson"), residuals(fit40,type="pearson"), rstandard(fit40,type="
  deviance"),residuals(fit40,type="deviance"))
#          [,1]      [,2]       [,3]      [,4]
# 1    3.167462  1.729494   3.170972  1.731411
# 2   -3.167462 -2.653614  -3.187353 -2.670277
# 3    6.224363  3.527757   6.233512  3.532943
# 4   -6.224363 -5.128120  -6.266146 -5.162545
# 5   10.066113  6.749966  10.072520  6.754262
# 6  -10.066113 -7.467569 -10.124704 -7.511035
# 7    2.444682  1.414214   2.445890  1.414912
# 8   -2.444682 -1.994109  -2.449488 -1.998029
# 9    9.309164  5.585977   9.308002  5.585280
# 10  -9.309164 -7.446972  -9.385363 -7.507929
# 11   2.338830  1.177622   2.315367  1.165808
# 12  -2.338830 -2.020726  -2.440511 -2.108577
# 13   5.650224  3.122794   5.661603  3.129084
# 14  -5.650224 -4.708841  -5.711317 -4.759756
# 15   2.261232  1.079724   2.262373  1.080268
# 16  -2.261232 -1.986799  -2.274618 -1.998559
```

The model is inadequate because *smoking* effect exists in some cities. The model's residual deviance is 288.27 with df = 8, which indicates that it fits rather poorly (P-value = 0). Also, by checking lack of fit of each city, we can find that for almost every city, the observed values are far from the model predicted.

Next, we consider the *city* effect and *smoking* effect model: $\text{logit}(\pi) = \alpha + \beta_{city} + \beta_{smoking}$.

```
> fit41=glm(Yes/(Yes+No)~factor(City)+Smoking,family=binomial(link=logit),data=lungs,weights=(Yes+No
  ))
> summary(fit41)
# Deviance Residuals:
#      Min        1Q    Median        3Q       Max
# -1.21781  -0.14842  -0.00012   0.16817   1.35470
#
# Coefficients:
#                        Estimate Std. Error z value Pr(>|z|)
# (Intercept)           -0.548682   0.118022  -4.649 3.34e-06 ***
# factor(City)Harbin     0.018187   0.129473   0.140    0.888
# factor(City)Nanchang  -0.054906   0.170996  -0.321    0.748
# factor(City)Nanjing    0.005764   0.140911   0.041    0.967
# factor(City)Shanghai   0.055618   0.119570   0.465    0.642
# factor(City)Shenyang  -0.027739   0.120071  -0.231    0.817
# factor(City)Taiyuan   -0.745683   0.185519  -4.019 5.83e-05 ***
# factor(City)Zhengzhou  0.028782   0.144755   0.199    0.842
```

```
# SmokingYes             0.777062   0.046775  16.613  < 2e-16 ***
#
#     Null deviance: 310.8951  on 15  degrees of freedom
# Residual deviance:   5.1958  on  7  degrees of freedom
# AIC: 121.05
#
> 1-pchisq(fit41$deviance,fit41$df.residual)
# [1] 0.6360822
> cbind(rstandard(fit41,type="pearson"), residuals(fit41,type="pearson"), rstandard(fit41,type="
    deviance"),residuals(fit41,type="deviance"))
#            [,1]          [,2]         [,3]          [,4]
# 1   0.038863100  0.0203993045  0.038865118  0.0204003636
# 2  -0.038863101 -0.0322738531 -0.038874999 -0.0322837339
# 3   0.500221992  0.2612895164  0.500428301  0.2613972816
# 4  -0.500221996 -0.3902934956 -0.501198475 -0.3910553844
# 5  -0.247142035 -0.1294368396 -0.247104237 -0.1294170435
# 6   0.247142036  0.1460958592  0.247059441  0.1460470334
# 7  -1.711678487 -0.9445258372 -1.708291496 -0.9426568531
# 8   1.711678519  1.3657735102  1.697802613  1.3547017208
# 9   0.001263846  0.0006159253  0.001263846  0.0006159256
# 10 -0.001263846 -0.0008511773 -0.001263849 -0.0008511796
# 11  0.229556858  0.1010236316  0.229398021  0.1009537302
# 12 -0.229556864 -0.2040742691 -0.231069689 -0.2054191585
# 13  1.483800386  0.7769830112  1.486229265  0.7782548786
# 14 -1.483800419 -1.2067620236 -1.497383609 -1.2178091144
# 15 -0.268376554 -0.1227104248 -0.268309716 -0.1226798644
# 16  0.268376557  0.2352749248  0.267550417  0.2345506807
```

This time, we can find that the model is adequate. The model's residual deviance is 5.19 with df = 7, which indicates that it fits well (P-value = 0.63). Also, by checking lack of fit of each city, we can find that there is no lack of fit issue.

```
> odds.ratio <-  function(x){
+   return ((x[1,1]*x[2,2])/(x[1,2]*x[2,1]))
+ }
>
> for (i in seq(0,4)){
+   print(odds.ratio(matrix(strtoi(l[(1+2*i):(2+2*i),3:4]),nrow=2,ncol=2)))
+ }
#[1] 2.196
#[1] 2.319148
#[1] 2.142962
#[1] 1.587963
#[1] 2.175265
> x = matrix(c(0,0,0,0),nrow=2,ncol=2)
> for (i in seq(0,4)){
+   a<-matrix(strtoi(l[(1+2*i):(2+2*i),3:4]),nrow=2,ncol=2)
+   for (j in seq(2)){
+     for (k in seq(2)){
+       x[j,k] = x[j,k]+a[j,k]
+     }
+   }
+ }
> print(x)
#     [,1] [,2]
#[1,] 2531 1999
#[2,] 1061 1779
> print(odds.ratio(x))
#[1] 2.122951
```

We can find that this model has a ML estimate of $e^{0.77} = 2.17$ for the smoking-cancer conditional odd ratio, which indicates the estimated odds of smoking were 117% higher for people who smoke than people who don't. Also, we can check marginal table of smoking-cancer observed odd ratios which is 2.12. There is no Simpson's paradox because the conditional association has the same direction as the marginal association.

Problem 5.14: Table 5.8 is from a study of nonmetastatic osteosarcoma described in the *LogXact* 7 manual (Cytel Software, 2005, p. 171). The response is whether the subject achieved a three-year disease-free interval.

Table 5.8 Data for Exercise 5.14.

Lymphocytic Infiltration	Sex	Osteoblastic Pathology	Disease-Free Yes	Disease-Free No
High	Female	No	3	0
High	Female	Yes	2	0
High	Male	No	4	0
High	Male	Yes	1	0
Low	Female	No	5	0
Low	Female	Yes	3	2
Low	Male	No	5	4
Low	Male	Yes	6	11

(a) Show that each explanatory variable has a significant effect when it is used as the sole predictor in logistic regression. Try to fit a main-effects model containing all three predictors. Explain why the ML estimate for the effect of lymphocytic infiltration is actually infinite.

```
> LX<-data.frame(LI=c('H','H','H','H','L','L','L','L'),sex=c('f','f','m','m','f','f','m','m'),OP=c
    (0,1,0,1,0,1,0,1),yes=c(3,2,4,1,5,3,5,6),no=c(0,0,0,0,0,2,4,11))
> fit51 <- glm(yes/(yes+no)~factor(LI),family=binomial(link=logit),data=LX,weights=(yes+no))
> summary(fit51)
#
# Deviance Residuals:
#      Min        1Q    Median        3Q       Max
# -1.44956   0.00008   0.00012   0.20659   2.52800
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)    20.06    4357.04   0.005    0.996
# factor(LI)L   -19.95    4357.04  -0.005    0.996
#
#     Null deviance: 19.4327  on 7  degrees of freedom
# Residual deviance:  8.6256  on 6  degrees of freedom
# AIC: 20.671
#
# Number of Fisher Scoring iterations: 18 # very slow convergence
#
# > Anova(fit51)
# Response: yes/(yes + no)
#            LR Chisq Df Pr(>Chisq)
# factor(LI)   10.807  1   0.001011 **
> confintModel(fit51, objective="ordinaryDeviance", method="zoom", endpoint.tolerance = 1e-08)
#                 Lower     Upper
# (Intercept) 1.552307        Inf
# factor(LI)L     -Inf -1.346829
> fit52 <- glm(yes/(yes+no)~factor(sex),family=binomial(link=logit),data=LX,weights=(yes+no))
> summary(fit52)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -1.4792  -0.1607   0.8416   1.1617   2.3003
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept)    1.8718     0.7595   2.464   0.0137 *
# factor(sex)m  -1.8073     0.8403  -2.151   0.0315 *
#
#     Null deviance: 19.433  on 7  degrees of freedom
# Residual deviance: 13.553  on 6  degrees of freedom
# AIC: 25.598
#
# Number of Fisher Scoring iterations: 4
```

```
#
> fit53 <- glm(yes/(yes+no)~OP,family=binomial(link=logit),data=LX,weights=(yes+no))
> summary(fit53)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -1.7360   0.1389   1.1688   1.3385   1.7134
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)   1.4469     0.5557   2.604  0.00922 **
# OP           -1.5270     0.6849  -2.230  0.02578 *
#
#     Null deviance: 19.433  on 7  degrees of freedom
# Residual deviance: 13.898  on 6  degrees of freedom
# AIC: 25.943
#
# Number of Fisher Scoring iterations: 4
#
> fit54 <- glm(yes/(yes+no)~factor(LI)+factor(sex)+OP,family=binomial(link=logit),data=LX,weights=(
    yes+no))
> summary(fit54)
# Deviance Residuals:
#        1         2         3         4         5         6         7         8
#  0.00002   0.00003   0.00005   0.00005   1.07088  -0.51727  -0.36813   0.27912
#
# Coefficients:
#                Estimate Std. Error z value Pr(>|z|)
# (Intercept)     23.4920 11084.3781   0.002   0.9983
# factor(LI)L    -21.3842 11084.3781  -0.002   0.9985
# factor(sex)m    -1.6362     0.9123  -1.794   0.0729 .
# OP              -1.2204     0.7712  -1.582   0.1135
#
#     Null deviance: 19.4327  on 7  degrees of freedom
# Residual deviance:  1.6278  on 4  degrees of freedom
# AIC: 17.673
#
# Number of Fisher Scoring iterations: 20
#
> Anova(fit54)
#             LR Chisq Df Pr(>Chisq)
# factor(LI)    6.9149  1   0.008548 **
# factor(sex)   3.7210  1   0.053731 .
# OP            2.6362  1   0.104451
> confintModel(fit54, objective="ordinaryDeviance", method="zoom", endpoint.tolerance = 1e-08)
#                  Lower       Upper
# (Intercept)   2.679308         Inf
# factor(LI)L       -Inf -0.80598077
# factor(sex)m -3.699682  0.02505359
# OP           -2.827147  0.24901300
```

From observing the results of analysis, the model *fit51*, *fit52*, and *fit53* represents the sole LI, Sex, and OP effect on disease-free or not respectively. The LRT for these three model indicate that each explanatory variable has a significant effect when it is used as the sole predictor in logistic regression. Next, we use the model *fit54* to fit main effects model containing all three predictors.

We can find that in *fit51* and *fit54*, LI has infinite effect. There are many facts indicating these phenomenon. Here I take *fit51* to illustrate this. Since the huge SE value, the Wald statistic is worthless. In this example, $z = 0$ and the P-value is near 1. By contrast, even with a truly infinite ML estimate, the likelihood-ratio test is valid. The difference between the null deviance and the residual deviance is 10.81 with df = 1. This test has P-value = 0.001 and yields very strong evidence of an effect. A 95% profile likelihood confidence interval for β is $(-\infty, -1.34)$, corresponding to a 1-unit multiplicative effect on the odds of at least $e^{-1.34} = 0.26$. The infinite lower endpoint reflects that the likelihood function keeps decreasing all the way out to $\hat{\beta} = -\infty$.

The reason why the results show infinite effect is when the x values having $y = 0$ are completely below or completely above those having $y = 1$. We can almost find that when LI is high, the prob. of yes all are 1, indicating the (quasi)-complete separation.

Supplementary Information

Problem 5.4 (a)

```
> stud<-read.table("http://users.stat.ufl.edu/~aa/cat/data/Students.dat",header=T)
> fit1 <- glm(abor ~ 1, family = binomial, data = stud)
> summary(fit1)
#Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept)   1.2852     0.3134   4.101 4.11e-05 ***
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 62.719  on 59  degrees of freedom
#AIC: 64.719
> fit2 <- glm(abor ~ ideol, family = binomial, data = stud)
> summary(fit2)
#Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept)   4.4205     1.0649   4.151 3.31e-05 ***
#ideol        -0.8789     0.2524  -3.482 0.000498 ***
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 45.464  on 58  degrees of freedom
#AIC: 49.464
> anova(fit2,fit1,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
#1        58     45.464
#2        59     62.719 -1  -17.255 3.269e-05 ***
> fit3 <- glm(abor ~ relig, family = binomial, data = stud)
> summary(fit3)
#Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept)   3.1762     0.7363   4.314 1.61e-05 ***
#relig        -1.2974     0.3837  -3.381 0.000722 ***
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 48.262  on 58  degrees of freedom
#AIC: 52.262
> anova(fit3,fit1,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
#1        58     48.262
#2        59     62.719 -1  -14.457 0.0001434 ***
> fit4 <- glm(abor ~ news, family = binomial, data = stud)
> summary(fit4)
#Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept)  -0.0385     0.5849  -0.066   0.9475
#news          0.4032     0.1769   2.280   0.0226 *
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 55.389  on 58  degrees of freedom
#AIC: 59.389
> anova(fit4,fit1,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
#1        58     55.389
#2        59     62.719 -1  -7.3299 0.006782 **
> fit5 <- glm(abor ~ hsgpa, family = binomial, data = stud)
> summary(fit5)
#Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept)   1.9124     2.3597   0.810    0.418
#hsgpa        -0.1889     0.7022  -0.269    0.788
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 62.645  on 58  degrees of freedom
#AIC: 66.645
> anova(fit5,fit1,test="Chisq")
#  Resid. Df Resid. Dev Df  Deviance Pr(>Chi)
#1        58     62.645
#2        59     62.719 -1 -0.073479   0.7863
> fit6 <- glm(abor ~ factor(gender), family = binomial, data = stud)
> summary(fit6)
#Coefficients:
#                 Estimate Std. Error z value Pr(>|z|)
```

```
#(Intercept)       0.9651     0.4155   2.323   0.0202 *
#factor(gender)1   0.6836     0.6412   1.066   0.2863
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 61.554  on 58  degrees of freedom
#AIC: 65.554
> anova(fit6,fit1,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
#1        58     61.554
#2        59     62.719 -1  -1.1648   0.2805
> fit7 <- glm(abor ~ ideol+relig+news, family = binomial, data = stud)
> summary(fit7)
#Coefficients:
#            Estimate Std. Error z value Pr(>|z|)
#(Intercept)   3.5205     1.2513   2.814  0.00490 **
#ideol        -1.2515     0.4671  -2.679  0.00738 **
#relig        -0.7198     0.4982  -1.445  0.14854
#news          1.1292     0.4574   2.469  0.01356 *
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 29.791  on 56  degrees of freedom
#AIC: 37.791
> anova(fit7,fit2,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
#1        56     29.791
#2        58     45.464 -2  -15.673 0.0003951 ***
> anova(fit7,fit3,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
#1        56     29.791
#2        58     48.262 -2  -18.471 9.752e-05 ***
> anova(fit7,fit4,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
#1        56     29.791
#2        58     55.389 -2  -25.598 2.764e-06 ***
> fit8 <- glm(abor ~ ideol+relig+news+gender+hsgpa, family = binomial, data = stud)
> summary(fit8)
#Coefficients:
#            Estimate Std. Error z value Pr(>|z|)
#(Intercept)  11.9148     5.0030   2.382  0.01724 *
#ideol        -1.3736     0.5184  -2.650  0.00805 **
#relig        -0.8606     0.5587  -1.540  0.12344
#news          1.4333     0.5566   2.575  0.01002 *
#gender        0.8837     1.1668   0.757  0.44883
#hsgpa        -2.6030     1.3359  -1.949  0.05135 .
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 25.188  on 54  degrees of freedom
#AIC: 37.188
> anova(fit8,fit7,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
#1        54     25.188
#2        56     29.791 -2  -4.6034   0.1001
> fit9 <- glm(abor ~ ideol+relig+news+ideol:relig+ideol:news+relig:news, family = binomial, data =
    stud)
> summary(fit9)
#Coefficients:
#            Estimate Std. Error z value Pr(>|z|)
#(Intercept)   0.4344     2.3249   0.187    0.852
#ideol        -1.3909     1.0038  -1.386    0.166
#relig         3.6192     2.3932   1.512    0.130
#news          1.1512     1.2536   0.918    0.358
#ideol:relig  -0.5236     0.5484  -0.955    0.340
#ideol:news    0.4923     0.3445   1.429    0.153
#relig:news   -1.1119     0.9542  -1.165    0.244
#
#    Null deviance: 62.719  on 59  degrees of freedom
#Residual deviance: 24.228  on 53  degrees of freedom
#AIC: 38.228
> anova(fit9,fit7,test="Chisq")
#  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
#1        53     24.228
#2        56     29.791 -3  -5.5638   0.1349
```

Problem 5.4 (b)

```
> fit10 <- glm(abor ~ gender+age+hsgpa+cogpa+dhome+dres+tv+sport+news+aids+veg+ideol+relig+affirm,
    family = binomial, data = stud)
> stepAIC(fit10)
# Start:  AIC=51.37
# abor ~ gender + age + hsgpa + cogpa + dhome + dres + tv + sport +
#     news + aids + veg + ideol + relig + affirm
#
#          Df Deviance    AIC
# - sport   1   21.380 49.380
# - gender  1   21.665 49.665
# - age     1   21.752 49.752
# - cogpa   1   22.028 50.028
# - aids    1   22.197 50.197
# - relig   1   22.355 50.355
# - dhome   1   22.466 50.466
# - affirm  1   22.664 50.664
# - dres    1   22.927 50.927
# - tv      1   23.147 51.147
# <none>        21.368 51.368
# - veg     1   23.389 51.389
# - hsgpa   1   24.924 52.924
# - ideol   1   32.261 60.261
# - news    1   34.371 62.371
#
# Step:  AIC=49.38
# abor ~ gender + age + hsgpa + cogpa + dhome + dres + tv + news +
#     aids + veg + ideol + relig + affirm
#
#          Df Deviance    AIC
# - gender  1   21.686 47.686
# - age     1   21.754 47.754
# - aids    1   22.199 48.199
# - cogpa   1   22.261 48.261
# - relig   1   22.397 48.397
# - dhome   1   22.497 48.497
# - affirm  1   22.689 48.689
# - dres    1   22.927 48.927
# - tv      1   23.172 49.172
# <none>        21.380 49.380
# - veg     1   23.778 49.778
# - hsgpa   1   24.990 50.990
# - ideol   1   32.418 58.418
# - news    1   35.239 61.239
#
# Step:  AIC=47.69
# abor ~ age + hsgpa + cogpa + dhome + dres + tv + news + aids +
#     veg + ideol + relig + affirm
#
# glm.fit: fitted probabilities numerically 0 or 1 occurred         Df Deviance    AIC
# - age     1   22.094 46.094
# - relig   1   22.418 46.418
# - aids    1   22.680 46.680
# - dhome   1   22.713 46.713
# - affirm  1   22.787 46.787
# - dres    1   23.051 47.051
# - cogpa   1   23.200 47.200
# <none>        21.686 47.686
# - veg     1   24.103 48.103
# - tv      1   24.238 48.238
# - hsgpa   1   25.008 49.008
# - ideol   1   33.813 57.813
# - news    1   35.965 59.965
#
# Step:  AIC=46.09
# abor ~ hsgpa + cogpa + dhome + dres + tv + news + aids + veg +
#     ideol + relig + affirm
#
# Step:  AIC=44.69
# abor ~ hsgpa + cogpa + dhome + dres + tv + news + aids + veg +
#     ideol + affirm
#
#          Df Deviance    AIC
```

```
# - affirm  1   23.286 43.286
# - aids    1   23.371 43.371
# - dhome   1   23.773 43.773
# - veg     1   24.626 44.626
# - cogpa   1   24.653 44.653
# <none>        22.691 44.691
# - dres    1   24.784 44.784
# - tv      1   25.364 45.364
# - hsgpa   1   26.035 46.035
# - news    1   36.921 56.921
# - ideol   1   40.943 60.943
#
# Step:  AIC=43.29
# abor ~ hsgpa + cogpa + dhome + dres + tv + news + aids + veg +
#     ideol
#
#         Df Deviance    AIC
# - aids   1   23.754 41.754
# - dhome  1   23.901 41.901
# - veg    1   24.658 42.658
# - dres   1   24.785 42.785
# - cogpa  1   25.135 43.135
# <none>       23.286 43.286
# - tv     1   25.430 43.430
# - hsgpa  1   26.426 44.426
# - news   1   37.250 55.250
# - ideol  1   41.782 59.782
#
# Step:  AIC=41.75
# abor ~ hsgpa + cogpa + dhome + dres + tv + news + veg + ideol
#
#         Df Deviance    AIC
# - dhome  1   24.266 40.266
# - veg    1   24.712 40.712
# - dres   1   24.790 40.790
# - cogpa  1   25.197 41.197
# - tv     1   25.450 41.450
# <none>       23.754 41.754
# - hsgpa  1   26.694 42.694
# - news   1   37.343 53.343
# - ideol  1   43.856 59.856
#
# Step:  AIC=40.27
# abor ~ hsgpa + cogpa + dres + tv + news + veg + ideol
#
#         Df Deviance    AIC
# - veg    1   25.004 39.004
# - dres   1   25.279 39.279
# - tv     1   25.716 39.716
# - cogpa  1   25.790 39.790
# <none>       24.266 40.266
# - hsgpa  1   27.251 41.251
# - news   1   39.648 53.648
# - ideol  1   46.859 60.859
#
# Step:  AIC=39
# abor ~ hsgpa + cogpa + dres + tv + news + ideol
#
#         Df Deviance    AIC
# - cogpa  1   25.912 37.912
# - tv     1   26.009 38.009
# - dres   1   26.151 38.151
# <none>       25.004 39.004
# - hsgpa  1   27.460 39.460
# - news   1   39.657 51.657
# - ideol  1   50.040 62.040
#
# Step:  AIC=37.91
# abor ~ hsgpa + dres + tv + news + ideol
#
#         Df Deviance    AIC
# - dres   1   27.146 37.146
# - tv     1   27.160 37.160
```

```
# - hsgpa  1   27.839 37.839
# <none>       25.912 37.912
# - news   1   40.892 50.892
# - ideol  1   50.933 60.933
#
# Step:  AIC=37.15
# abor ~ hsgpa + tv + news + ideol
#
#         Df Deviance    AIC
# - tv     1   27.944 35.944
# <none>       27.146 37.146
# - hsgpa  1   30.248 38.248
# - news   1   42.143 50.143
# - ideol  1   52.334 60.334
#
# Step:  AIC=35.94
# abor ~ hsgpa + news + ideol
#
#         Df Deviance    AIC
# <none>       27.944 35.944
# - hsgpa  1   32.014 38.014
# - news   1   44.667 50.667
# - ideol  1   54.654 60.654
#
# Call:  glm(formula = abor ~ hsgpa + news + ideol, family = binomial,
#     data = stud)
#
# Coefficients:
# (Intercept)        hsgpa         news        ideol
#      11.287       -2.338        1.291       -1.594
#
# Degrees of Freedom: 59 Total (i.e. Null);  56 Residual
# Null Deviance:      62.72
# Residual Deviance: 27.94  AIC: 35.94
```

Problem 5.4 (c)

```
fit11 <- glm(veg ~ gender+age+hsgpa+cogpa+dhome+dres+tv+sport+news+aids+abor+ideol+relig+affirm,
    family = binomial, data = stud)
summary(fit11)
anova(fit11,glm(veg ~ 1, family = binomial, data = stud),test="Chisq")
# Call:
# glm(formula = veg ~ gender + age + hsgpa + cogpa + dhome + dres +
#     tv + sport + news + aids + abor + ideol + relig + affirm,
#     family = binomial, data = stud)
#
# Deviance Residuals:
#      Min        1Q    Median        3Q       Max
# -1.40618  -0.23680  -0.00534   0.00000   1.90975
#
# Coefficients:
#               Estimate Std. Error z value Pr(>|z|)
# (Intercept)  5.541e+00  2.902e+03   0.002   0.9985
# gender      -1.844e+00  1.641e+00  -1.124   0.2612
# age          4.772e-02  7.677e-02   0.622   0.5342
# hsgpa       -3.839e+00  2.885e+00  -1.331   0.1833
# cogpa       -2.937e+00  2.269e+00  -1.294   0.1956
# dhome       -1.123e-03  7.130e-04  -1.575   0.1153
# dres         4.552e-01  2.562e-01   1.776   0.0757 .
# tv          -2.897e-02  1.289e-01  -0.225   0.8221
# sport       -5.885e-01  3.945e-01  -1.492   0.1358
# news         1.344e-01  2.928e-01   0.459   0.6462
# aids         2.069e-01  2.541e-01   0.814   0.4155
# abor        -3.573e+00  3.248e+00  -1.100   0.2713
# ideol       -2.458e+00  1.469e+00  -1.673   0.0942 .
# relig        1.636e+00  1.071e+00   1.528   0.1266
# affirm       2.355e+01  2.902e+03   0.008   0.9935
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 26.645  on 45  degrees of freedom
# AIC: 56.645
#
# Number of Fisher Scoring iterations: 19
#
# Analysis of Deviance Table
#
# Model 1: veg ~ gender + age + hsgpa + cogpa + dhome + dres + tv + sport +
#     news + aids + abor + ideol + relig + affirm
# Model 2: veg ~ 1
#   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
# 1        45     26.645
# 2        59     50.725 -14   -24.08  0.04482 *
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
for (x in colnames(stud)){
  print(x)
  print(summary(glm(veg ~ stud[,x], family = binomial, data = stud)))
}
# [1] "gender"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.6559  -0.6559  -0.4673  -0.4673   2.1301
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)  -2.1595     0.6097  -3.542 0.000397 ***
# stud[, x]     0.7324     0.7605   0.963 0.335568
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
```

```
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 49.753  on 58  degrees of freedom
# AIC: 53.753
#
# Number of Fisher Scoring iterations: 4
#
# [1] "age"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.6615  -0.5735  -0.5632  -0.5571   1.9696
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.964853   1.252769  -1.568    0.117
# stud[, x]    0.007841   0.040571   0.193    0.847
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.689  on 58  degrees of freedom
# AIC: 54.689
#
# Number of Fisher Scoring iterations: 4
#
# [1] "hsgpa"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.7437  -0.6044  -0.5431  -0.5090   2.0928
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)  -0.2165     2.4818  -0.087    0.930
# stud[, x]    -0.4636     0.7577  -0.612    0.541
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.360  on 58  degrees of freedom
# AIC: 54.36
#
# Number of Fisher Scoring iterations: 4
#
# [1] "cogpa"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.7178  -0.6079  -0.5426  -0.4977   2.0473
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)   0.3777     3.5170   0.107    0.914
# stud[, x]    -0.6164     1.0286  -0.599    0.549
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.367  on 58  degrees of freedom
# AIC: 54.367
#
# Number of Fisher Scoring iterations: 4
#
```

```
# [1] "dhome"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.6459  -0.6262  -0.5760  -0.3960   2.0642
#
# Coefficients:
#               Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.4613079  0.4514569  -3.237  0.00121 **
# stud[, x]   -0.0002714  0.0003241  -0.837  0.40243
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 49.782  on 58  degrees of freedom
# AIC: 53.782
#
# Number of Fisher Scoring iterations: 5
#
# [1] "dres"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.7223  -0.5719  -0.5440  -0.5290   2.0272
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.93218    0.50273  -3.843 0.000121 ***
# stud[, x]    0.04810    0.07912   0.608 0.543171
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.379  on 58  degrees of freedom
# AIC: 54.379
#
# Number of Fisher Scoring iterations: 4
#
# [1] "tv"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.9549  -0.5765  -0.5396  -0.5084   2.0869
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept) -2.05731    0.53830  -3.822 0.000132 ***
# stud[, x]    0.04077    0.04637   0.879 0.379349
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.013  on 58  degrees of freedom
# AIC: 54.013
#
# Number of Fisher Scoring iterations: 4
#
# [1] "sport"
```

```
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.7836  -0.6438  -0.5621  -0.3826   2.0889
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)  -1.0234     0.6244  -1.639    0.101
# stud[, x]    -0.1484     0.1205  -1.231    0.218
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 48.908  on 58  degrees of freedom
# AIC: 52.908
#
# Number of Fisher Scoring iterations: 5
#
# [1] "news"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.5797  -0.5710  -0.5691  -0.5672   1.9544
#
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.749599   0.613576  -2.851  0.00435 **
# stud[, x]    0.003663   0.120730   0.030  0.97580
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.724  on 58  degrees of freedom
# AIC: 54.724
#
# Number of Fisher Scoring iterations: 4
#
# [1] "aids"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.7930  -0.5630  -0.5379  -0.5379   2.0025
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept) -1.86031    0.42897  -4.337 1.45e-05 ***
# stud[, x]    0.07861    0.12961   0.607    0.544
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.387  on 58  degrees of freedom
# AIC: 54.387
#
# Number of Fisher Scoring iterations: 4
#
# [1] "ideol"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
```

```
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.8075  -0.6555  -0.4210  -0.3173   2.0212
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)  -0.4782     0.8004  -0.597    0.550
# stud[, x]    -0.4752     0.3065  -1.550    0.121
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 47.624  on 58  degrees of freedom
# AIC: 51.624
#
# Number of Fisher Scoring iterations: 5
#
# [1] "relig"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.6702  -0.5579  -0.5579  -0.5080   2.0551
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)  -1.9827     0.5929  -3.344 0.000826 ***
# stud[, x]     0.2011     0.3608   0.558 0.577141
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 50.420  on 58  degrees of freedom
# AIC: 54.42
#
# Number of Fisher Scoring iterations: 4
#
# [1] "abor"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -0.6109  -0.6109  -0.6109  -0.4001   2.2649
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)  -2.4849     1.0408  -2.387    0.017 *
# stud[, x]     0.9008     1.1108   0.811    0.417
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 49.936  on 58  degrees of freedom
# AIC: 53.936
#
# Number of Fisher Scoring iterations: 5
#
# [1] "affirm"
#
# Call:
# glm(formula = veg ~ stud[, x], family = binomial, data = stud)
#
# Deviance Residuals:
#      Min        1Q    Median        3Q       Max
```

```
# -0.68533  -0.68533  -0.68533  -0.00008   1.76860
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept)   -19.57    2608.23  -0.008    0.994
# stud[, x]      18.24    2608.23   0.007    0.994
#
# (Dispersion parameter for binomial family taken to be 1)
#
#     Null deviance: 50.725  on 59  degrees of freedom
# Residual deviance: 44.121  on 58  degrees of freedom
# AIC: 48.121
#
# Number of Fisher Scoring iterations: 18
```