Name ______________________

Student Number ______________________

STA 312f2012 Quiz 9

~~1.~~ (5 points) Mid-level executives at large corporations are classified by sex, and as either married or not, resulting in a 2×2 table of Sex by Marital Status.

(a) Write a linear regression model for $\log \mu_{ij}$, including the interaction of Sex ~~of Dog~~ by S~~ex of Owner~~. by marital status

$$\log \mu_{ij} = \beta_0 + \beta_1 s + \beta_2 m + \beta_3 sm$$

(b) Make a table with 4 rows showing how your dummy variables are defined. There will be a column for each dummy variable. Use effect coding (the scheme with the -1s). Make as many extra columns as necessary to show the product term(s) corresponding to the interaction(s). Add a last column showing $\log \mu_{ij}$ in terms of the β parameters.

Sex	Mar	s	m	sm	$\log \mu_{ij}$
M	Mar	1	1	1	$\beta_0 + \beta_1 + \beta_2 + \beta_3$
M	Sin	1	-1	-1	$\beta_0 + \beta_1 - \beta_2 - \beta_3$
F	Mar	-1	1	-1	$\beta_0 - \beta_1 + \beta_2 - \beta_3$
F	Sin	-1	-1	1	$\beta_0 - \beta_1 - \beta_2 + \beta_3$

(c) Make a 2×2 table showing μ_{ij} (not $\log \mu_{ij}$) in each cell. next page

	Married	Single
M	$\exp\{\beta_0 + \beta_1 + \beta_2 + \beta_3\}$	$\exp\{\beta_0 + \beta_1 - \beta_2 - \beta_3\}$
F	$\exp\{\beta_0 - \beta_1 + \beta_2 - \beta_3\}$	$\exp\{\beta_0 - \beta_1 - \beta_2 + \beta_3\}$

(d) Using your table, show that the odds ratio θ equals ~~zero if~~ one and only all interaction terms equal zero.

$$\theta = \frac{\operatorname{Exp}\{2\beta_0 + 2\beta_3\}}{\operatorname{Exp}\{2\beta_0 - 2\beta_3\}} = e^{4\beta_3}$$

So $\theta = 1$ iff $\beta_3 = 0$

2: (5 points) In your analysis of the Heart Attack Data, you fit a model with just age, cholesterol level, and family history of heart disease. Consider the test of Cholesterol level controlling for Age and Family history of heart disease.

(a) Give the value of G^2; the answer is a number from your printout.

$G^2 = 4.67$

(b) Give the Degrees of freedom. The answer is a number on your printout.

$df = 1$

(c) Give the p-value; the answer is a number from your printout.

$p = 0.0966$

(d) In plain, non-statistical language, what do you conclude from this test?

Allowing for age and family history of heart disease, there is no clear evidence of a link between cholesterol level and heart attack outcome

Attach your printout for Question 2 (Homework Question 1c). Make sure your name is written on the printout.

Quiz 9 Printout

```
> attack =
read.table("http://www.utstat.toronto.edu/~brunner/312f12/code_n_data/attac
k.data")
> head(attack)
  age diastol cholest ncigs height weight famhist   school   outcome
1  40      70     321     0   68.8    190     Yes  PostSec   Alive10
2  49      87     246    60   72.2    204      No       HS   Alive10
3  43      89     262     0   69.0    162      No       HS DiedFirst
4  50     105     275    15   62.5    152     Yes GradeSch   Alive10
5  50      88     261    30   68.0    142      No GradeSch    Dead10
6  47      79     372    30   67.0    193      No       HS   Alive10
> attach(attack)
>
> # Make an mlogit data frame
> longjohn = mlogit.data(attack,shape="wide",choice="outcome")
>
> model1 = mlogit(outcome ~ 0 | age+cholest+famhist, data=longjohn)
>
> summary(model1)

Call:
mlogit(formula = outcome ~ 0 | age + cholest + famhist, data = longjohn,
    method = "nr", print.level = 0)

Frequencies of alternatives:
  Alive10    Dead10 DiedFirst
  0.53846   0.12500   0.33654

nr method
5 iterations, 0h:0m:0s
g'(-H)^-1g = 5.32E-05
successive fonction values within tolerance limits

Coefficients :
                         Estimate Std. Error t-value Pr(>|t|)
Dead10:(intercept)     -9.2358517  4.5775569 -2.0176  0.04363 *
DiedFirst:(intercept) -3.1529671  3.0567050 -1.0315  0.30231
Dead10:age              0.1634238  0.0895093  1.8258  0.06788 .
DiedFirst:age           0.1077304  0.0595888  1.8079  0.07062 .
Dead10:cholest         -0.0004808  0.0051048 -0.0942  0.92496
DiedFirst:cholest      -0.0085263  0.0042825 -1.9910  0.04649 *
Dead10:famhistYes      -0.2413608  0.6552984 -0.3683  0.71263
```

```
DiedFirst:famhistYes  -0.9337608  0.5108499 -1.8279  0.06757 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -93.041
McFadden R^2:  0.067872
Likelihood ratio test : chisq = 13.549 (p.value = 0.035095)
> # Now test each expanatory variable controlling for the other 2. Full
model is always model1
> # Age controlling for cholest and famhist
> red1a = mlogit(outcome ~ 0 | cholest+famhist, data=longjohn)
> G2 = -2 * as.numeric(red1a$logLik - model1$logLik); G2
[1] 5.865214
> df = length(model1$coefficients)-length(red1a$coefficients); df
[1] 2
> pval = 1-pchisq(G2,df); pval
[1] 0.053258
> # Cholest controlling for age and famhist
> red1b = mlogit(outcome ~ 0 | age+famhist, data=longjohn)
> G2 = -2 * as.numeric(red1b$logLik - model1$logLik); G2
[1] 4.673621
> df = length(model1$coefficients)-length(red1b$coefficients); df
[1] 2
> pval = 1-pchisq(G2,df); pval
[1] 0.09663536
>
> # Famhist controlling for age and famhist
> red1c = mlogit(outcome ~ 0 | age+cholest, data=longjohn)
> G2 = -2 * as.numeric(red1c$logLik - model1$logLik); G2
[1] 3.58312
> df = length(model1$coefficients)-length(red1c$coefficients); df
[1] 2
> pval = 1-pchisq(G2,df); pval
[1] 0.1666999
>
>
```