Name Jerry

Student Number ______

STA 312f2012 Quiz 8

1. (5 points) In a large corporation, a sample of office workers are given a typing test that counts the number of errors they make typing a standard document. It is natural to think of number of errors as a Poisson random variable. Then they are randomly assigned to one of three conditions: Training programme A, Training programme B, or a control condition designated as C, in which they receive the afternoon off with pay instead of being trained. We will treat score on the pre-test as a covariate. In your model, call it x.

*

(a) Write a Poisson regression model for this problem. You need not say how your dummy variables are defined; you will do that in the next part. I will start the answer for you. Letting $E(Y) = \lambda$,

$$\log \lambda = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2$$

(b) Make a table showing how you would set up the dummy variables for this problem. Use indicator dummy variable coding, with the control condition C as the reference category. There will be one row for each experimental condition. In addition to the columns for the dummy variables, make another column showing λ (not $\log \lambda$).

Prog	d_1	d_2	
A	1	0	$e^{\beta_0+\beta_1 x+\beta_2}$
B	0	1	$e^{\beta_0+\beta_1 x+\beta_3}$
C	0	0	$e^{\beta_0+\beta_1 x}$

(c) For a given score on the pretest, the expected number of errors is e^{β_2} times as great for training programme A as for the control condition.

(d) Write a full model for $\log \lambda$ that would let you test the proportional means assumption. Again, it starts with

$$\log \lambda = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 x d_1 + \beta_5 x d_2$$

(e) In the symbols of your model above, what null hypothesis would you test to check the proportional means assumption?

$H_0: \beta_4 = \beta_5 = 0$

2. (3 points) This question is based on your printout from the Awards data. ***Write your answers in the blanks below, and also circle them on your printout.*** **On the printout, label the answers 2a, 2b etc.**

   (a) 3.025 Give just the value of the Z or G^2 statistic used to answer this question: Controlling for score on the test of general knowledge, do students in the Academic program get more awards on average than students in the General program? The answer is a single number from your printout.

   (b) 0.348 Give just the value of the Z or G^2 statistic used to answer this question: Is the proportional means assumption reasonable for these data? The answer is a single number from your printout.

   (c) 6.619 Give just the value of the Z or G^2 statistic used to answer this question: Controlling for academic program, is score on the test of general knowledge related to the expected number of awards? The answer is a single number from your printout.

3. (2 points) In plain, non-statistical language, what do you conclude from the test of Question 2c? You have more room than you need.

Even allowing for what academic programme they are in, students with higher scores on the test of general knowledge tend to get more awards.

No marks unless they say MORE.

At least one point off for any mention of Ho, p-value, etc.

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

312f22 Quiz 10 Printout

```
> # Awards Work
> # Took data from http://www.ats.ucla.edu/stat/r/dae/poissonreg.htm
> # Changed the example a little: score = math+20
> rm(list=ls())
> prizes =
read.table("http://www.utstat.toronto.edu/~brunner/312f12/code_n_data/award
s.data")
> attach(prizes); head(prizes)
  awards prog score
1      0    3    61
2      0    1    61
3      0    3    64
4      0    3    62
5      0    3    60
6      0    1    62
> program = factor(prog, labels = c("General", "Academic", "Vocational"))
> model1 = glm(awards ~ score + program, family=poisson); summary(model1)

Call:
glm(formula = awards ~ score + program, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2043  -0.8436  -0.5106   0.2558   2.6796

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -6.65017    0.84808  -7.841 4.46e-15 ***
score              0.07015    0.01060   6.619 3.63e-11 ***
programAcademic    1.08386    0.35825   3.025  0.00248 **
programVocational  0.36981    0.44107   0.838  0.40179
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 287.67  on 199  degrees of freedom
Residual deviance: 189.45  on 196  degrees of freedom
AIC: 373.5

Number of Fisher Scoring iterations: 6
```

2c

2a

```
> anova(model1,test='Chisq')
Analysis of Deviance Table

Model: poisson, link: log

Response: awards

Terms added sequentially (first to last)


        Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                      199     287.67
score    1   83.651       198     204.02 < 2.2e-16 ***
program  2   14.572       196     189.45 0.0006852 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # Switch dummy var coding to test Academic vs Vocational
> kon2 = contr.treatment(3,base=2) # Academic is ref
> plabs = c("General", "Academic", "Vocational")
> colnames(kon2) <- plabs[c(1,3)] # Giving it col names works!
> program2 = factor(prog); contrasts(program2) = kon2
> model1b = glm(awards ~ score + program2, family=poisson);
summary(model1b)

Call:
glm(formula = awards ~ score + program2, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2043  -0.8436  -0.5106   0.2558   2.6796

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -5.56631    0.87311  -6.375 1.83e-10 ***
score               0.07015    0.01060   6.619 3.63e-11 ***
program2General    -1.08386    0.35825  -3.025  0.00248 **
program2Vocational -0.71405    0.32001  -2.231  0.02566 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 287.67  on 199  degrees of freedom
Residual deviance: 189.45  on 196  degrees of freedom
```

```
AIC: 373.5

Number of Fisher Scoring iterations: 6

> model2 = glm(awards ~ score + program + score:program, family=poisson);
summary(model2)

Call:
glm(formula = awards ~ score + program + score:program, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2295  -0.7958  -0.5298   0.2528   2.6826

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)             -4.74173    3.43126  -1.382    0.167
score                    0.04400    0.04721   0.932    0.351
programAcademic         -1.00926    3.57006  -0.283    0.777
programVocational       -1.30271    3.94471  -0.330    0.741
score:programAcademic    0.02841    0.04870   0.583    0.560
score:programVocational  0.02290    0.05421   0.422    0.673

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 287.67  on 199  degrees of freedom
Residual deviance: 189.10  on 194  degrees of freedom
AIC: 377.16

Number of Fisher Scoring iterations: 6

> anova(model1,model2,test='Chisq')
Analysis of Deviance Table

Model 1: awards ~ score + program
Model 2: awards ~ score + program + score:program
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       196     189.45
2       194     189.10  2    0.348   0.8403
```

2b