A4 As in **A2** on the previous assignment, consider the Negative Binomial GLM with known θ_z. Suppose that the log-link is used and that the model contains the intercept and one factor predictor with L levels.

(a) Write down the GLM model equation $g(\mu_i) = \eta_i$, i.e. specify g and the linear predictor η_i. How many parameters does the model have?

(b) In the special case $L = 2$, suppose that the levels of the factor are A and B, say, and that $x_i = A$ for $i = 1, \ldots, n_A$ (group A) and $x_i = B$ for $i = n_A+1, \ldots, n_A+n_B$ (group B); here, $n = n_A + n_B$. Write down the log-likelihood and derive the score equations.

(c) For the model in part (b), calculate the parameter estimates explicitly. Determine the fitted means? Would the fitted means and/or the estimates of β differ if the canonical link was used? Explain.

(d) For the model in part (b), compute the Fisher information matrix explicitly (i.e. do NOT leave it in the form $X^T WX$). How would it differ if the canonical link were used (again, calculate the Fisher information matrix explicitly in this case)?

(e) Compute the observed information for the model in part (b), i.e. minus the Hessian matrix of the log-likelihood function. Is this matrix the same as the Fisher information matrix calculated in part (c)? Why or why not?

(f) Give an expression for the asymptotic variance of the MLEs for the model in part (b).

(g) Derive the deviance for the model in part (b).

(h) **Bonus question for extra marks**. For the model in part (b), suppose that $\log(\mu_i) = \beta_0$ for $i = 1, \ldots, n_A$, and set $\bar{Y}_A = (1/n_A)(Y_1 + \cdots + Y_{n_A})$ to be the group mean of group A. Calculate the variance of $\bar{Y}_A$: (i) directly; (ii) using the GLM formula for the variance of $\hat{\beta}_0$ when the log link is used; (iii) using the GLM formula for the variance of $\hat{\beta}_0$ when the canonical link is used. Observe that the answer is the same in all three cases.

A5 For n independent observations from a Poisson distribution with parameter μ, show that Fisher scoring gives $\mu(t+1) = \bar{y} = (1/n)(y_1 + \cdots + y_n)$ for all $t > 0$. By contrast, what happens with the Newton–Raphson method?

A6 `R` **exercise.** Load the data set on the number of awards received by high school students in a given year, viz.

```
awards <- read.csv("awards.csv")
attach(awards)
```

The data set is available on MyCourses under Course Content - Assignments. In this data set, the response variable `numawards` is a discrete variable indicating the number of awards received. The explanatory variables are `math`, a continuous variable giving the students' score on their final mathematics exam, and `prog`, a factor variable with levels `1 = "General", 2 = "Academic", 3 = Vocational"` indicating the type of program in which the students were enrolled.

(a) Fit a Poisson GLM with the log-link for `numawards` with `math` as a predictor, viz.

```
glm(numawards~1+math,family=poisson(link=log))
```

Report the estimated parameters. Does `math` seem to be a significant predictor? Construct a 95% confidence interval for β_1 (i.e. the parameter pertaining to `math`).

(b) Fit a Poisson GLM with the log link and `prog` as predictor (careful, `prog` is a factor predictor, so use `as.factor(prog)` in `R`). How many parameters does the model have? Print the estimated parameters and comment on their significance. Use a Wald test as well as a likelihood ratio test to determine whether `prog` is a significant predictor.

(c) Fit the following Poisson GLMs

```
numawards~1+math+as.factor(prog)
numawards~1+math*as.factor(prog)
```

How many parameters does each of these models have? Interpret each model in terms of the effect of `math` and `prog` on the expected number of awards of a student.

(d) Construct a plot to assess how well the models from parts (a), (b), and (c) fit. Because `math` is a continuous predictor, you may consider binning it into a reasonable number of categories, say into nine intervals $(30, 35], \ldots, (70, 75]$, and computing the mean number of awards for each category. *Hint: explore the* `R` *functions* `cut` *and* `split`.

(e) Consider again the models fitted in parts (a), (b), and (c). Use deviance comparisons to select the model that is best suited for these data and interpret it.