Question 1

A4 (a) For a GLM with canonical link function, explain how the likelihood equations imply that the residual vector $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{\mu}}$ is orthogonal to the column space of X. Here, $\boldsymbol{y}$ and $\hat{\boldsymbol{\mu}}$ denote the vectors of the observed responses and the fitted means, respectively.

$$\because \quad g = (b')^{-1} \quad \Rightarrow \quad \theta = X\beta = \eta$$

$$\text{Then} \quad \frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}$$

$$= \frac{y_i - b'(\theta_i)}{a(\phi)} \cdot \frac{1}{b''(\theta_i)} \cdot \frac{\partial \mu_i}{\partial \theta_i} \cdot x_{ij}$$

$$= \frac{y_i - b'(\theta_i)}{a(\phi)} \cdot x_{ij}$$

Then induce the score function from the likelihood equation:

$$\sum_{i=1}^{n} \frac{y_i - b'(\theta_i)}{a(\phi)} \cdot x_{ij} = 0 \qquad \text{for } j = 1, 2, \cdots p$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - b'(\theta_i))\, x_{ij} = 0$$

$$\sum_{i=1}^{n} (y_i - \hat{\mu}_i)\, x_{ij} = 0$$

$$\Rightarrow X^T(y - \hat{\mu}) = 0$$

$\therefore$ The column space of X is orthogonal to $y - \hat{\mu} = e$

(b) In a GLM that uses a noncanonical link function, explain why it need not be true that $\sum_i \hat{\mu}_i = \sum_i y_i$, that is, the residuals need not have mean zero.

$$\frac{\partial l_i}{\partial \beta_j} = \frac{y_i - b'(\theta_i)}{a(\phi)} \frac{1}{b''(\theta_i)} \frac{1}{g'(\mu_i)} x_{ij}$$

$$\text{The requirement is} \quad \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - b'(\theta_i)}{a(\phi)} \frac{1}{b''(\theta_i)} \frac{1}{g'(\mu_i)} x_{ij} = 0 \qquad \text{for } \forall j = 1, 2, \cdots p$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \hat{\mu}_i) \frac{1}{\mathrm{Var}(Y_i)} \cdot \frac{1}{g'(\mu_i)} x_{ij} = 0$$

This cannot imply that $\sum_{i=1}^{n} \hat{\mu}_i = \sum_{i=1}^{n} y_i$

In other words,

$$\sum_{i=1}^{n} e_i \frac{1}{Var(Y_i) g'(\mu_i)} x_{ij} = 0 \quad \text{cannot imply that} \sum_{i=1}^{n} e_i \to 0$$

$$\text{that is } \frac{1}{n}\sum_{i=1}^{n} e_i \xrightarrow{n\to\infty} E[e_i] = 0$$

Let $W = diag(w_1, \cdots w_n) = diag\left(\frac{1}{Var(Y_1)g'(\mu_1)}, \cdots, \frac{1}{Var(Y_n)g'(\mu_n)}\right)$

Then

$$X^T W e = 0$$

(c) Explain why a canonical link GLM needs an intercept term in order to ensure that $\sum_i \hat{\mu}_i = \sum_i y_i$.

For the canonical link GLM, $\frac{\partial l_i}{\partial \beta_j} = \frac{y_i - b'(\theta_i)}{a(\phi)} x_{ij}$

For the GLM model, we need

$$\sum_{i=1}^{n} \frac{y_i - b'(\theta_i)}{a(\phi)} x_{ij} = 0 \quad \text{for } \forall j = 1, 2, \cdots p$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \hat{\mu}_i) x_{ij} = 0$$

If there's an intercept term, then $X_{i1} = 1$ for $\forall i \in \{1, \cdots n\}$.

Then when $j=1$, we have

$$\sum_{i=1}^{n} (y_i - \hat{\mu}_i) X_{i1} = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \hat{\mu}_i) = 0$$

Then $\sum_{i=1}^{n} \hat{\mu}_i = \sum_{i=1}^{n} y_i$ can be ensured.

(d) Considering a Gaussian GLM with the canonical link, recall that $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y}$ can be calculated explicitly. Show how the Fisher Scoring Algorithm proceeds in this case - does it find the explicit solution, and if so, in how many iterations? How does the Newton–Raphson Algorithm compare?

For Gaussian GLM : $\mu = \theta$, $b(\mu) = \frac{\mu^2}{2}$, $b'(\theta) = \theta = \mu$, canonical link : $g(x) = x$

let $D = \mathrm{diag}\left(\frac{1}{g'(\mu_i)},\ i=1, \cdots n \right) = I_n$

$W = \mathrm{diag}\left(\frac{1}{\mathrm{Var}(Y_i)} \left\{ \frac{1}{g'(\mu_i)} \right\}^2,\ i=1, \cdots n \right) = \mathrm{diag}\left(\frac{1}{\sigma_i^2},\ i=1, \cdots n \right)$

FS algorithm:

① Initialization : $\mu^{(0)} = (y_1, \cdots y_n)$, $\eta^{(0)} = g(\mu^{(0)}) = \mu^{(0)}$

$z^{(0)} = \eta^{(0)} + (y - \mu^{(0)}) g'(\mu^{(0)}) = \eta^{(0)} = \mu^{(0)}$

$W_i^{(0)} = \frac{1}{a(\phi)\, V(\mu_i^{(0)})} \left\{ \frac{1}{g'(\mu_i)} \right\}^2 = \frac{1}{\sigma_i^2}$

② Iteration :

$\beta^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W^{(t)} z^{(t)}$

$\eta^{(t+1)} = X \beta^{(t+1)}$

$\mu^{(t+1)} = g^{-1}(\eta^{(t+1)}) = \eta^{(t+1)}$

$W_i^{(t+1)} = \frac{1}{a(\phi)\, V(\mu_i^{(t+1)})} = \frac{1}{a(\phi)} = \frac{1}{\sigma_i^2}$

$z^{(t+1)} = \eta^{(t+1)} + g'(\mu^{(t+1)})(y - \mu^{(t+1)})$

$= \mu^{(t+1)} + (y - \mu^{(t+1)}) = y$

$D^{(t+1)} = I_n$

$t = t+1$

The Algorithm in this case can always find a solution, as $z^{(t+1)} = y$ always.

Then $\beta^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W^{(t)} z^{(t)}$

$= (X^T X)^{-1} X^T (W^{(t)})^{-1} W^{(t)} z^{(t)}$

$= (X^T X)^{-1} X^T Y$

The Algorithm can find the solution in one iteration.

- For the Newton-Raphson Algorithm: $\beta^{(t+1)} = \beta^{(t)} - H(\beta^{(t)})^{-1} \mu(\beta^t)$

Since we are using canonical link, then $H(\beta) = -J(\beta)$

$\Rightarrow \quad \beta^{(t+1)} = \beta^{(t)} + J(\beta^{(t)})^{-1} \cdot \mu(\beta^{(t)})$

$$= (X^T W(\beta^{(t)}) X)^{-1} X^T W(\beta^{(t)})\, z^{(t)} \quad \text{where} \quad z^{(t)} = X\beta^{(t)} + D^{-1}(\beta^{(t)})(y - \mu^{(t)})$$

① Initialization :

$$\mu^{(0)} = (y_1, \cdots y_n)$$

$$H^{(0)} = H(\beta^{(0)}) = -J(\beta^{(0)})$$

$$z^{(0)} = \eta^{(0)} + g'(\mu^{(0)})(y - \mu^{(0)}) = \mu^{(0)}$$

$$W_i^{(0)} = \frac{1}{a(\phi)\, v(\mu_i^{(0)})} \left\{ \frac{1}{g'(\mu_i^{(0)})} \right\}^2 = \frac{1}{a(\phi)}$$

② Iteration :

$$\beta^{(t+1)} = \beta^{(t)} - \{H^{(t)}\}^{-1} \mu^{(t)}$$

$$= \beta^{(t)} + J(\beta^{(t)})^{-1} \mu^{(t)}$$

$$= (X^T W^{(t)} X)^{-1} X^T W(\beta^{(t)})\, z^{(t)}$$

② Iteration :

$$\beta^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W(\beta^{(t)})\, z^{(t)}$$

$$\eta^{(t+1)} = X\beta^{(t+1)}$$

$$\mu^{(t+1)} = g^{-1}(\eta^{(t+1)}) = \eta^{(t+1)}$$

$$W_i^{(t+1)} = \frac{1}{a(\phi)\, v(\mu_i^{(t+1)})} = \frac{1}{a(\phi)}$$

$$z^{(t+1)} = -\eta^{(t+1)} + g^{-1}(\mu^{(t+1)})(y - \mu^{(t+1)})$$

$$= -\mu^{(t+1)} + y - \mu^{(t+1)} = y - 2\mu^{(t+1)}$$

$$t = t + 1$$

$$\mu^{(t)} = \mu(\beta^{(t)}) = \left(\frac{\partial \ell}{\partial \beta_1^{(t+1)}} \cdots \frac{\partial \ell}{\partial \beta_p^{(t+1)}} \right)$$

$$= \left(\sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi)} x_{ij} \quad \text{for } j = 1, \cdots p \right)$$

$$H^{(t)} = H(\beta^{(t)}) = -J(\beta^{(t)})$$

$$W_i^{(t)} = \frac{1}{a(\phi)\, v(\mu_i^{(t+1)})} = \frac{1}{a(\phi)}$$

$$z^{(t)} = \eta^{(t+1)} + g^{-1}(\mu^{(t+1)})(y - \mu^{(t+1)}) = \mu^{(t+1)} + y - \mu^{(t+1)} = y$$

The Newton Raphson stops when $\| \beta^{(t)} - \beta^{(t+1)} \| < \varepsilon$, then $\beta^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W(\beta^{(t)})\, Y$, so it can finally converge to $(X^T X)^{-1} X^T Y$ in 1 iteration, where $\beta^{(t)} = \beta^{(t+1)}$

Question 2

A5 As in **A2** on the previous assignment, consider the Inverse Gaussian GLM with the log-link and one continuous predictor, viz.

$$\log(\mu_i) = \alpha + \beta x_i, \quad i = 1, \ldots, n.$$

(a) Compute the Fisher information matrix for this model.

Inverse Gaussian : $f(y; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left[-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right]$

Then $Var(Y) = \frac{\mu^3}{\lambda}$ and $E[Y] = \mu$

Since $\eta = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{12} \\ & \vdots & \\ x_{n1} & \cdots & x_{n2} \end{bmatrix}$

Let $J(\beta)$ be the matrix of Fisher information, then

$$J(\beta)_{jk} = \sum_{i=1}^{n} x_{ij} W_i x_{ik} \quad \text{where } W_i = \frac{1}{Var(Y_i)}\left(\frac{1}{g'(\mu_i)}\right)^2$$

$$= \sum_{i=1}^{n} x_{ij} \frac{\lambda}{\mu_i} x_{ik} \qquad = \frac{\lambda}{\mu_i^3} \mu_i^2 = \frac{\lambda}{\mu_i}$$

Then $J(\beta) = \begin{bmatrix} \sum_{i=1}^{n} x_{i1}\frac{\lambda}{\mu_i}x_{i1} & \sum_{i=1}^{n} x_{i1}\frac{\lambda}{\mu_i}x_{i2} \\ \sum_{i=1}^{n} x_{i2}\frac{\lambda}{\mu_i}x_{i1} & \sum_{i=1}^{n} x_{i2}\frac{\lambda}{\mu_i}x_{i2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{\lambda}{\mu_i} & \sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i \\ \sum_{i=1}^{n} x_i \frac{\lambda}{\mu_i} & \sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i^2 \end{bmatrix}$

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

$$[H(\beta)]_{jk} = \frac{\partial^2 \ell}{\partial\beta_j \partial\beta_k}(\beta) \qquad \text{where } \beta_1 = \alpha,\ \beta_2 = \beta$$

$j, k \in \{1, 2\}$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial\beta_k}\left\{\frac{y_i - b'(\theta_i)}{a(\phi)} \cdot \frac{1}{b''(\theta_i)} \cdot \mu_i \cdot x_{ij}\right\}$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial\beta_k}\left\{\frac{y_i - \mu_i}{\mu_i^3/\lambda} * \mu_i \cdot x_{ij}\right\}$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial\beta_k}\left\{\frac{\lambda(y_i - \mu_i)}{\mu_i^2} x_{ij}\right\}$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial\mu_i}\left\{\frac{\lambda(y_i - \mu_i)}{\mu_i^2} x_{ij}\right\} \cdot \frac{\partial\mu_i}{\partial\eta_i} \frac{\partial\eta_i}{\partial\beta_k}$$

$$\Rightarrow [H(\beta)]_{jk} = \sum_{i=1}^{n} \left\{ -\frac{\lambda y_i x_{ij}}{\mu_i^3} + \frac{\lambda x_{ij}}{\mu_i^2} \right\} \mu_i \; x_{ik}$$

$$= \sum_{i=1}^{n} x_{ij} \left\{ \frac{\lambda}{\mu_i} - \frac{\lambda y_i}{\mu_i^2} \right\} x_{ik}$$

$$\Rightarrow -H(\beta) = \begin{bmatrix} -\sum_{i=1}^{n} \left(\frac{\lambda}{\mu_i} - \frac{\lambda y_i}{\mu_i^2} \right) & -\sum_{i=1}^{n} \left(\frac{\lambda}{\mu_i} - \frac{\lambda y_i}{\mu_i^2} \right) x_i \\ -\sum_{i=1}^{n} \left(\frac{\lambda}{\mu_i} - \frac{\lambda y_i}{\mu_i^2} \right) x_i & -\sum_{i=1}^{n} \left(\frac{\lambda}{\mu_i} - \frac{\lambda y_i}{\mu_i^2} \right) x_i^2 \end{bmatrix}$$

This is not the same as the Fisher-information matrix. Since we use log link here, so $\frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \neq 1$, where $\frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} = 1$ for canonical link and $J(\beta) = -H(\beta)$ for the canonical link. Then the the weight of $H(\beta)$: $\left\{ \frac{\lambda}{\mu_i} - \frac{\lambda y_i}{\mu_i^2} \right\}$ is different from $W_i = \frac{\lambda}{\mu_i}$.

(c) Give an expression for the asymptotic variance of the MLE $\hat{\alpha}$ of the intercept.

MLE $\hat{\beta} \xrightarrow{P} \beta$ $\quad \{J(\beta)\}^{\frac{1}{2}} (\hat{\beta} - \beta) \xrightarrow{n \to \infty} N(0, I_p)$

As n get large : $\hat{\beta} \approx N(\beta, \{J(\beta)\}^{-1})$

$$\because J(\beta) = \begin{bmatrix} \sum_{i=1}^{n} \frac{\lambda}{\mu_i} & \sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i \\ \sum_{i=1}^{n} x_i \frac{\lambda}{\mu_i} & \sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i^2 \end{bmatrix}$$

$$\therefore J(\beta)^{-1} = \frac{1}{\left(\sum_{i=1}^{n} \frac{\lambda}{\mu_i} \sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i^2 \right) - \left(\sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i \sum_{i=1}^{n} \frac{\lambda}{\mu_i} x_i \right)} \begin{bmatrix} \sum_{i=1}^{n} \frac{1}{\mu_i} x_i^2 & -\sum_{i=1}^{n} \frac{1}{\mu_i} x_i \\ -\sum_{i=1}^{n} x_i \frac{1}{\mu_i} & \sum_{i=1}^{n} \frac{1}{\mu_i} \end{bmatrix}$$

$\Rightarrow$ the asymptotic variance will be
$$\left(\frac{1}{\lambda} \right) \frac{\sum_{i=1}^{n} \frac{1}{\mu_i} x_i^2}{\left(\sum_{i=1}^{n} \frac{1}{\mu_i} \right) \left(\sum_{i=1}^{n} \frac{x_i^2}{\mu_i} \right) - \left(\sum_{i=1}^{n} \frac{x_i}{\mu_i} \right) \left(\sum_{i=1}^{n} \frac{x_i}{\mu_i} \right)}$$

(d) Derive the deviance for this model.

Since the model contains an intercept and

$$\frac{\partial l_i}{\partial \alpha} = \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha}$$

$$= \frac{y_i - b'(\theta_i)}{a(\phi)} \cdot \frac{1}{b''(\theta_i)} \cdot \frac{1}{g'(\mu_i)} \cdot x_{i1}$$

$$= \frac{y_i - \mu_i}{\mu_i^3/\lambda} \cdot \mu_i$$

Then $\sum_{i=1}^{n} \frac{\partial l_i}{\partial \alpha} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mu_i^3/\lambda} \mu_i = 0$

$$\sum_{i=1}^{n} (y_i - \mu_i) \frac{\lambda}{\mu_i^2} = 0$$

$$\sum_{i=1}^{n} \frac{y_i}{\mu_i^2} = \sum_{i=1}^{n} \frac{1}{\mu_i}$$

$\hat{\theta}_i = (b')^{-1}(\hat{\mu}_i) = -\frac{1}{2\hat{\mu}_i^2}$ and $\tilde{\theta}_i = (b')^{-1}(y_i) = -\frac{1}{2y_i^2}$

the Deviance $D(y, \hat{\mu}) = 2 \sum_{i=1}^{n} w_i [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)]$

$$= 2 \sum_{i=1}^{n} [y_i \cdot (-\frac{1}{2y_i^2} + \frac{1}{2\hat{\mu}_i^2}) - \sqrt{-2\hat{\theta}_i} + \sqrt{-2\tilde{\theta}_i}]$$

$$= 2 \sum_{i=1}^{n} [\frac{1}{2y_i} - \frac{y_i}{2\hat{\mu}_i^2} - \sqrt{-2 \cdot (-\frac{1}{2\hat{\mu}_i^2})} + \sqrt{-2 \cdot (-\frac{1}{2y_i^2})}]$$

$$= 2 \sum_{i=1}^{n} [\frac{1}{2y_i} - \frac{y_i}{2\hat{\mu}_i^2} - \frac{1}{\hat{\mu}_i} + \frac{1}{y_i}]$$

$$= \sum_{i=1}^{n} \frac{y_i^2 - 2\hat{\mu}_i y_i + \hat{\mu}_i^2}{\hat{\mu}_i^2 y_i}$$

$$= \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}$$

A2_Q6

Question 3

A6 R excercise.Load the data set SoreThroat.dat available on MyCourses under Assigmnents:

```
SoreThroat<-read.table("SoreThroat.dat",header=TRUE)
attach(SoreThroat)
```

```
## The following object is masked from package:base:
##
##     T
```

In this data set, the response variable Y describes whether a patient having surgery experienced a sore throat on waking (1 = "yes", 0 = "no"). The explanatory variables are D, a continuous variable giving the duration of the surgery in minutes, and T, a factor variable with levels 1 = "Tracheal tube", 0 = "Laryngeal mask airway" indicating the type of device used to secure the airway.

(a) Fit a GLM to the data, using the intercept, D, T, and the interaction between D and T. Write down the model for the probability of experiencing a sore throat on waking and interpret it (qualitative statements are fully sufficient).

```
# probability of experiencing a sore throat
# use the Gaussian GLM
Y<-SoreThroat$Y
D<-SoreThroat$D
T0<- factor(SoreThroat$T)
logitmod<- glm(cbind(Y,1-Y)~D+T0+T0*D,family=binomial,x=TRUE)
summary(logitmod)
```

```
##
## Call:
## glm(formula = cbind(Y, 1 - Y) ~ D + T0 + T0 * D, family = binomial,
##     x = TRUE)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9707  -0.3779   0.3448   0.7292   1.9961
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.04979    1.46940   0.034   0.9730
## D            0.02848    0.03429   0.831   0.4062
## T01         -4.47224    2.46707  -1.813   0.0699 .
## D:T01        0.07460    0.05777   1.291   0.1966
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 46.180  on 34  degrees of freedom
## Residual deviance: 28.321  on 31  degrees of freedom
## AIC: 36.321
##
## Number of Fisher Scoring iterations: 6
```

We are using the binomial GLM with canonical link here, so $g(\mu_i) = log(\frac{p_i}{1-p_i}) = \eta_i$. The probability of experiencing a sore throat P will be $log(\frac{p}{1-p}) = 0.366973 + 0.09062D - 3.19094T$ and $P = \frac{p}{1-p}$. The P is approximatly the probability of Y=1. If the probability is higher than 0.5, then it is more likely to have a sore throat and otherwise it is more likely to have no sore throat.

(b)If $\widehat{\beta}$ denotes that parameter pertaining to the interaction, construct a 95% confidence interval for $\widehat{\beta}$

```
# construct the 95% CI
I<-t(logitmod$x)%*%diag(logitmod$weights)%*%(logitmod$x)
I.inv<-solve(I)

sd<-sqrt(diag(I.inv))
# here we want 95% CI
z<-qnorm(0.975)
beta<-logitmod$coefficients

c.upper<-beta+z*sd
c.lower<-beta-z*sd
CI<- cbind(c.lower,c.upper)
colnames(CI)<-c("2.5%","97.5%")
CI
```

```
##                    2.5%      97.5%
## (Intercept) -2.83018681 2.92976028
## D           -0.03872086 0.09567689
## T01         -9.30760657 0.36312369
## D:T01       -0.03862224 0.18782478
```

Then the 95% confidence interval for each paraeters is shown above, the 95% confidence interval for $\widehat{\beta}$ is [-0.0386,0.1878].

(c)Using (i) a Wald test and (ii) a likelihood ratio test, test whether the interaction between D and T is significant. Test at the 5% level.

For (i) wald test, we build a new model by adding a connection between D and T. Then if

```
library(Matrix)
#test whether each beta=0 using Wald tests
beta/sd
```

```
## (Intercept)           D         T01       D:T01
##  0.03388233  0.83060747 -1.81277565  1.29139077
```

```r
pchisq((beta/sd)^2,df=1,lower.tail=FALSE)
```

```
## (Intercept)           D         T01        D:T01
##  0.97297098  0.40619541  0.06986643  0.19656821
```

Then we can see that the p-value is 0.1966>0.05, so the null hypothsis is sitting in the 95% convidence interval.Then the null hypothesis should not be rejected, so the interaction between D and T is not significant.

```r
# use the likelihood ratio test
logit_new1<-glm(cbind(Y,1-Y)~D+T0,family=binomial,x=TRUE)
L1<--2*(logLik(logit_new1)-logLik(logitmod))
L1
```

```
## 'log Lik.' 1.816886 (df=3)
```

```r
p.val<-pchisq(L1,df=1,lower.tail=FALSE)
p.val
```

```
## 'log Lik.' 0.1776844 (df=3)
```

The p value is equal to 0.1777>0.05, we should accpet the hull hypothsis that the interation is not significant.

(d)Using (i) a Wald test and (ii) a likelihood ratio test, test whether the interaction between D and T and T are significant predictors (that is, whether we can remove T from the model altogether). Test at the 5% level.

```r
#Wald test
var<-solve(I.inv[3:4,3:4])
beta<-logitmod$coefficients
beta2<-c(beta[3:4])
W<-t(beta2)%*%var%*%(beta2)
pchisq(W,df=2,lower.tail=FALSE)
```

```
##            [,1]
## [1,] 0.1287759
```

```r
#likelihood ratio test
logit_new2<-glm(cbind(Y,1-Y)~D,family=binomial,x=TRUE)
L2<--2*(logLik(logit_new2)-logLik(logitmod))
L2
```

```
## 'log Lik.' 5.330286 (df=2)
```

```r
p.val<-pchisq(L2,df=2,lower.tail=FALSE)
p.val
```

```
## 'log Lik.' 0.06958941 (df=2)
```

According to the Wald test, we have 0.129>0.05 and the likelihood ratio test shows that 0.0696>0.05. We accept the null hypothesis test that the interaction of D and T and T are not significant, so T can beremoved from the model.