Model For patients' surgery

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 $\widehat{\beta}$

```r
# 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

```r
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
```

2

```
##  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.