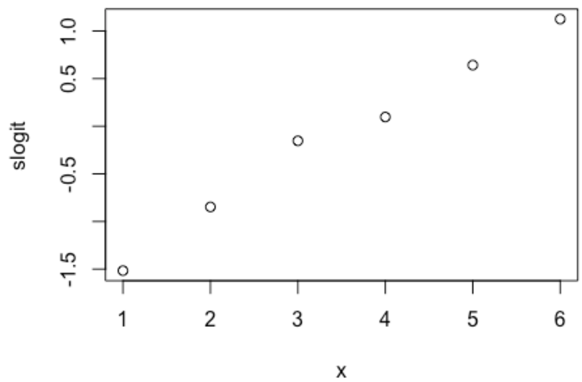
HW#8, Nan Deng

(1)

(a)

```
ymat <- matrix(c(13,61,55,129,489,570,475,431,293,154,38,12),ncol=2,nrow=6,byrow=T)
slogit <- log((ymat[,1]+0.5)/(ymat[,2]+0.5))
x <- c(1, 2, 3, 4, 5, 6)
plot(x,slogit)
```



(b)

```
cancer_fit <- glm(ymat~x,family=binomial)
summary(cancer_fit)

##
## Call:
## glm(formula = ymat ~ x, family = binomial)
##
## Deviance Residuals:
##       1        2        3        4        5        6
## -1.3369  -0.9074   1.7517  -1.3168   0.1211   0.2286
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.60032    0.15466  -10.35   <2e-16 ***
## x            0.44630    0.04159   10.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 130.9010  on 5  degrees of freedom
## Residual deviance:   7.4797  on 4  degrees of freedom
## AIC: 46.383
##
## Number of Fisher Scoring iterations: 3

paste('p-value of Chi-square test = ',pchisq(7.4797,4,lower.tail=F))

## [1] "p-value of Chi-square test =  0.112607775626486"
```

```
#qchisq(0.95,4)
ei <- residuals(cancer_fit,type="pearson")
chi_square <- sum(ei^2)
paste('X2 = ',chi_square)

## [1] "X2 =  7.36205612902065"

paste('p-value of Pearson = ',pchisq(chi_square,4,lower.tail=F))

## [1] "p-value of Pearson =  0.117948132433085"
```

The Residual Deviance of this model is 7.4797 with 4 degree of freedom. For the Chi-square test, the p-value is 0.1126078, which is greater than 0.05, so this model fits the data well. On the other hand, the Pearson chi-squared statistic X2 is 7.362056, while the corresponding p-value is 0.1179481. Similarly, the hypothesis that the model fits well fails to be rejected.

(c)

```
cancer_base_fit <- glm(ymat~1,family=binomial)
anova(cancer_base_fit,cancer_fit)

## Analysis of Deviance Table
##
## Model 1: ymat ~ 1
## Model 2: ymat ~ x
##   Resid. Df Resid. Dev Df Deviance
## 1         5     130.90
## 2         4       7.48  1   123.42

paste('p-value of Likelihood Ratio Test = ',pchisq(123.42,1,lower.tail=F))

## [1] "p-value of Likelihood Ratio Test =  1.12845581968394e-28"

#qchisq(0.95,1)
varb <- vcov(cancer_fit)[-1,-1]
varbi <- solve(varb)
beta <- coef(cancer_fit)[-1]
w <- t(beta) %*% varbi %*% beta
paste('w =',w)

## [1] "w = 115.152964000646"

paste('p-value of Wald Test = ',pchisq(w,4,lower.tail=F))

## [1] "p-value of Wald Test =  5.78861765735426e-24"
```

Based on the analysis, the deviance of Likelihood Ratio Test is 123.42 with 1 degree of freedom. According to the p-value 1.12845581968394e-28, which if far less than 0.05, the null hypothesis should be rejected. For the Wald Test, the statistic is 115.152964000646 with 1 degree of freedom. The corresponding p-value is 5.78861765735426e-24 and should also be rejected due to less than 0.05.

(d)

```
exp(beta)

##        x
## 1.562519
```

```
paste('Confidence interval for the odds ratio = ',paste(exp(beta-1.96*sqrt(vcov(cancer_fit)[2,2])),exp(beta+1.96*sqrt(vcov(cancer_fit)[2,2]))))

## [1] "Confidence interval for the odds ratio =  1.44020150546514 1.69522550535127"
```

For a unit increase in x (a one-level increase in daily average number of cigarettes), the estimated odds ratio of lung cancer is 1.562519, while the 95% confidence interval for the odds ratio is (1.44020150546514, 1.69522550535127).

(e)

```
exp(beta*2)

##        x
## 2.441466

paste('Confidence interval for the odds ratio = ',paste(exp(beta*2-1.96*2*sqrt(vcov(cancer_fit)[2,2])),exp(beta*2+1.96*2*sqrt(vcov(cancer_fit)[2,2]))))

## [1] "Confidence interval for the odds ratio =  2.07418037634406 2.87378951399348"
```

For a 2-unit increase in x (a one-level increase in daily average number of cigarettes), the estimated odds ratio of lung cancer is 2.441466, while the 95% confidence interval for the odds ratio is (2.07418037634406, 2.87378951399348).

(2)

(a)

```
acc <- c(188,107,63,23,241,92,61,19,200,118,22,19,21,13,11,1,31,11,5,0,26,17,2,6)
trav <- c(204433874,177250749,41949294,29883757,128647023,59822202,17642351,6838521,31765363,59730974,1289058,3775431,23163210,21162524,9473358,3474259,15040022,5529527,2400560,459525,3207263,4951688,224036,340844)
truck <- gl(2,12)
road <- gl(3,4,length=24)
time <- gl(2,2,length=24)
area <- gl(2,1,length=24)
logtrav <- log(trav)
dataf <- data.frame(truck=truck,road=road,time=time,area=area,acc=acc,trav=trav,logtrav=logtrav)
m1 <- glm(acc ~ truck + road + time + area + road*time + road*area,family=poisson, offset=logtrav, data=dataf)
summary(m1)

##
## Call:
## glm(formula = acc ~ truck + road + time + area + road * time +
##     road * area, family = poisson, data = dataf, offset = logtrav)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6388  -0.5644  -0.1062   0.4557   2.2073
##
## Coefficients:
##              Estimate Std. Error  z value Pr(>|z|)
## (Intercept) -13.89100    0.06673 -208.180  < 2e-16 ***
## truck2        0.09039    0.08861    1.020  0.30772
## road2         0.70600    0.08857    7.971 1.57e-15 ***
## road3         1.92451    0.09271   20.758  < 2e-16 ***
```

```
## time2          0.36547    0.11540    3.167  0.00154 **
## area2         -0.47201    0.10255   -4.603 4.17e-06 ***
## road2:time2    0.17539    0.16662    1.053  0.29252
## road3:time2    0.61055    0.19121    3.193  0.00141 **
## road2:area2    0.28262    0.14726    1.919  0.05496 .
## road3:area2   -0.63468    0.14424   -4.400 1.08e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 711.086  on 23  degrees of freedom
## Residual deviance:  18.982  on 14  degrees of freedom
## AIC: 156.06
##
## Number of Fisher Scoring iterations: 4
```

(b)

```
acc[acc == 0] <- 0.5
obsrate <- acc/trav
mu <- obsrate
eta <- log(mu)
z <- eta
w <- acc
w[w == 0] <- 0.5
for (i in 1:5) {
  fit_model <- lm(z~truck+road+time+area+road*time+road*area,weights=w,data=dataf)
  eta <- fitted.values(fit_model)
  mu <- exp(eta)
  z <- eta+(obsrate-mu)/mu
  w <- trav*mu
}
coef(fit_model)
```

```
##  (Intercept)       truck2        road2        road3        time2
## -13.89141453   0.09393788   0.70483636   1.92456069   0.36525866
##        area2  road2:time2  road3:time2  road2:area2  road3:area2
##  -0.47195958   0.18162753   0.61063814   0.28696277  -0.63465071
```

```
coef(m1)
```

```
##  (Intercept)       truck2        road2        road3        time2
## -13.89099812   0.09038607   0.70600305   1.92450739   0.36547438
##        area2  road2:time2  road3:time2  road2:area2  road3:area2
##  -0.47200876   0.17538832   0.61054921   0.28262247  -0.63468355
```

```
varbeta <- summary(fit_model)$cov.unscaled
sqrt(diag(varbeta))
```

```
## (Intercept)      truck2       road2       road3       time2       area2
##  0.06672939  0.08847666  0.08858474  0.09271375  0.11540434  0.10254564
## road2:time2 road3:time2 road2:area2 road3:area2
##  0.16641772  0.19120946  0.14715039  0.14423589
```

```
sqrt(diag(vcov(m1)))
```

```
## (Intercept)      truck2       road2       road3       time2       area2
##  0.06672590  0.08861235  0.08856862  0.09271284  0.11540368  0.10254526
## road2:time2 road3:time2 road2:area2 road3:area2
##  0.16662364  0.19120930  0.14726350  0.14423575
```

The result shown in part b is quite similar to those from part a, which uses Fisher scoring algorithm.