# Homework 4

*Gustavo Esparza*

*05/09/2019*

## 1

Given only the Odds Ratio for a Bernoulli Random Variable, Y, we can not determine the Relative Risk. We will disprove the claim with a counterexample. Consider the following two tables:

| | x = 0 | x = 1 | Total |
|---|---|---|---|
| y=0 | 7 | 27 | 34 |
| y=1 | 41 | 25 | 66 |
| Total | 48 | 42 | 100 |

Table 1

| | x = 0 | x = 1 | Total |
|---|---|---|---|
| y=0 | 25 | 27 | 52 |
| y=1 | 41 | 7 | 48 |
| Total | 66 | 34 | 100 |

Table 2

For the first table, we have the following results:

$OR_1 = \frac{7\times25}{27\times41} \approx .16$ $RR_1 = \frac{7(27+25)}{27(7+41)} \approx .28$

For the second table, we have the following results:

$OR_2 = \frac{25\times7}{27\times41} \approx .16$ $RR_2 = \frac{25(27+7)}{27(25+41)} \approx .48$

Then, we have presented two different tables with a Bernoulli Random Variable, Y, that produce the same Odds Ratios but differing Relative Risks. Thus, the Relative Risk for a Bernoulli variable can not be determined if provided only the odds ratio.

## 2

Here is the table provided in the text:

| | Learning Impaired | Unimpaired | Total |
|---|---|---|---|
| Drink during Pregnancy | 24 (A) | 88 (B) | 112 |
| Did not drink during pregnancy | 16 (C) | 172 (D) | 188 |
| Total | 40 | 260 | 300 |

### A

Let the event of a **Mother Drinking During Pregnancy** be represented by **D** if they did drink during pregnancy and $\bar{\mathbf{D}}$ if they did not drink during pregnancy.

Let the outcome of a **Child Being Learning Impaired** be represented by **L** if they are learning impaired and $\bar{\mathbf{L}}$ if they are not learning impaired.

Then the odds ratio for learning impairement based on drinking during pregnancy is defined as

$$\widehat{OR} = \frac{(L \cap D)(\bar{L} \cap \bar{D})}{(\bar{L} \cap D)(L \cap \bar{D})} = \frac{24 * 172}{88 * 16} \approx 2.932$$

### B

We are creating a confidence interval for a model that has one nominal predictor (Drinking During Pregnancy), so an appropriate method for the interval would be Woolf's Method.

Letting $A = L \cap D$, $B = \bar{L} \cap D$, $C = L \cap \bar{D}$, and $D = \bar{L} \cap \bar{D}$, we can define Woolf's Confidence Interval as follows:

$$CI_{\widehat{OR}} = e^{ln(\widehat{OR}) \pm z^* \sqrt{\frac{1}{A}\frac{1}{B}\frac{1}{C}\frac{1}{D}}} = e^{ln(2.932) \pm 1.96 \sqrt{\frac{1}{24}\frac{1}{88}\frac{1}{16}\frac{1}{172}}} = [1.48, 5.80]$$

We can see that our Confidence Interval contains values greater than one, implying the odds ratio is reliable for the data presented.

### C

I am not convinced that that this is the best way to present information on the risks of drinking while pregnant. When considering the cause of learning Impairement in children, we should be considering far more factors than simply Drinking During Pregnancy. When considering social and economic factors, there are many variables that could also influence a mother to drink during pregnancy. Previous medical conditions, history of learning impairement, and age of the mother are just a few factors that should also be considered. Thus, we would have a confounding variable that has more to do with learning impairment of a child than drinking alone does.

## 3

### NEWTON METHOD

```r
library(readr)
```

```r
#Import Data
hw4= read.csv("/Users/gustavo/Desktop/MATH 535/13/hw4.csv", header=TRUE,
stringsAsFactors = FALSE)

#Cleaning Data
hw4=hw4[,-1]
hw4[hw4$x3 == "yes",4]=1
hw4[hw4$x3 == "no",4]=0
```

Here is our Newton-Raphson algorithm for estimated the coeffecients of our logistic regression model:

```
newton=function(beta0,X,y,maxit){

  for(i in 1:maxit){

    #Calculate pi and make it a vector
    p = exp(X%*%beta0) / (1+ exp(X%*%beta0))
    p=as.vector(p)

    W = diag(p*(1-p))  #Used for hessian

    gradient =  t(X) %*% (y-p)
    hessian = -t(X) %*% W %*% X

    beta1 = beta0 - solve(hessian) %*% gradient #Update

    converge = sqrt(sum((beta1-beta0)^2)) #Convergence

    cat("Iteration =",i,"\n")
    cat("Intercept =",beta1[1],"x1 = ",beta1[2],"x2 = ",beta1[3],"x3 = ",beta1[4],"\n")
    cat("\n")

    beta0=beta1
    if(converge<1e-6){break} #Convergence Check
  }}
```

Here is the initialization of our beta.

```
beta0=matrix(c(0,0,0,0),4,1)

x1=hw4$x1
x2=hw4$x2
x3=as.numeric(hw4$x3)

X=as.matrix(cbind(1,x1,x2,x3))
y=hw4[,1]

maxit=20
```

Finally, here is our result:

```
newton(beta0,X,y,maxit)
```

```
## Iteration = 1
## Intercept = 3.506191 x1 =  -0.3867816 x2 =  0.05410087 x3 =  -0.1681087
##
## Iteration = 2
## Intercept = 5.567462 x1 =  -0.5777306 x2 =  -0.03207953 x3 =  -0.2530813
##
## Iteration = 3
## Intercept = 6.452693 x1 =  -0.6567783 x2 =  -0.0804705 x3 =  -0.2847043
##
## Iteration = 4
## Intercept = 6.57154 x1 =  -0.6673461 x2 =  -0.08724733 x3 =  -0.2886978
##
```

```
## Iteration = 5
## Intercept = 6.573368 x1 =  -0.6675084 x2 =  -0.08735332 x3 =  -0.2887576
##
## Iteration = 6
## Intercept = 6.573368 x1 =  -0.6675085 x2 =  -0.08735335 x3 =  -0.2887576
```

Thus, we have the following estimates for the coeffecients of our linear logistic regression model: $\beta_0 = 6.57, \beta_1 = -.67, \beta_2 = -.09, \beta_3 = -.29$

## B-GLM CHECK

```r
y=hw4$y
x1=hw4$x1
x2=hw4$x2
x3=hw4$x3
```

```r
model=glm(y ~ x1+x2+x3, family="binomial",data=hw4)
summary(model)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3, family = "binomial", data = hw4)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0747  -0.6775  -0.2517   0.7180   2.7170
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.57337    3.01849   2.178   0.0294 *
## x1          -0.66751    0.09711  -6.873 6.27e-12 ***
## x2          -0.08735    0.96985  -0.090   0.9282
## x31         -0.28876    0.38474  -0.751   0.4529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 272.12  on 199  degrees of freedom
## Residual deviance: 176.78  on 196  degrees of freedom
## AIC: 184.78
##
## Number of Fisher Scoring iterations: 5
```

We can see that the coeffecients provided in the glm function are equivalent to the estimates found in our Newton method.

## C-ESTIMATED IMPACT

If $\Delta = 5$ is the change in $x_1$ and $\beta_1$ is the coeffecient found for $x_1$, then it's respective impact on the odds of $y = 1$ is equivalent to $e^{\Delta \beta_1}$. Then, we have the following result:

```r
delta = 5
x1=model$coefficients[2]
```

```r
impact = exp(delta*x1)

cat("The estimated impact of a 5 unit increase in x1 on the odds of Y=1 is ",impact,"")
```

```
## The estimated impact of a 5 unit increase in x1 on the odds of Y=1 is  0.03552416
```