Stat 203 HW 4 Solution

1. Some students may treat any number > 0 as 1, but this will cause a lost of information. So they can get at most 8 points.

A sample of R code is attached in the end.

(a) The fitted result is:

```
Call:
glm(formula = Y ~ T, family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.95227  -0.78299  -0.54117  -0.04379   2.65152

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.08498    3.05247   1.666   0.0957 .
T           -0.11560    0.04702  -2.458   0.0140 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 24.230  on 22  degrees of freedom
Residual deviance: 18.086  on 21  degrees of freedom
AIC: 35.647

Number of Fisher Scoring iterations: 5
```

We can see that the temperature has a coefficient of -0.1156, meaning what when temperature decreases by 1, the log odds ratio of O-ring failure will increase by 0.1156 (or the odds ratio of O-ring failure will increase to exp(0.1156) times).

(b) The fitted result is:

```
Call:
glm(formula = Y[-18, ] ~ T[-18], family = binomial)

Deviance Residuals:
```

```
   Min      1Q   Median      3Q     Max  
-0.7608  -0.5742  -0.3320  -0.1861   1.5204  

Coefficients:
          Estimate Std. Error z value Pr(>|z|)   
(Intercept)  8.66157    3.63441   2.383  0.01716 * 
T[-18]     -0.17680    0.05869  -3.013  0.00259 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 20.0667  on 21  degrees of freedom
Residual deviance:  9.4096  on 20  degrees of freedom
AIC: 24.748

Number of Fisher Scoring iterations: 6
```

We can see that the temperature has a coefficient of -0.1768, meaning what when temperature decreases by 1, the log odds ratio of O-ring failure will increase by 0.1768 (or the odds ratio of O-ring failure will increase to exp(0.1768) times).

(c) The probability is 0.96.

(d) The probability that at least one out of six O-rings fails is $1-(1-0.96)^6 \approx 100\%$. The probability that all six O-rings fail is $0.96^6 \approx 78\%$. Our judgment should be based on one of the above two percentage, which are both too high. So I would not advise that; launching at so low temperature was obviously unwise. (Students will get full credit if only they make right judgment. Any reason is OK.)

A sample of source code:

```
orings <- read.table("Orings.txt", header = T)
T <- orings$Temp
S <- orings$Damaged
F <- 6 - S
Y <- cbind(S, F)

######## Part (a)
orings.glm <- glm(Y ~ T, family = binomial)
```

```
print(summary(orings.glm))

######## Part (b)
orings.glm.d <- glm(Y[-18,] ~ T[-18], family = binomial)
print(summary(orings.glm.d))

######## Part (c)
z <- exp(31 * coef(orings.glm.d)[2] + coef(orings.glm.d)[1])
t31 <- z / (1 + z)
print(t31)
```

2. A sample of R code is attached in the end.

(a) The fitted result for NFL is:

```
Call:
glm(formula = cbind(Success, Attempts - Success) ~ Distance +
    I(Distance^2), family = binomial(link = "logit"), data = nfl[Z ==
    0, ])

Deviance Residuals:
         1           2           3           4           5
 0.1162755  -0.0004784  -0.4017326   0.6420910  -0.9146466

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.490203   1.018620   2.445   0.0145 *
Distance     -0.013167   0.065990  -0.200   0.8419
I(Distance^2) -0.001513   0.001008  -1.500   0.1335
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 147.7816  on 4  degrees of freedom
Residual deviance:   1.4238  on 2  degrees of freedom
AIC: 28.890

Number of Fisher Scoring iterations: 4
```

The fitted result for AFL is:

```
Call:
glm(formula = cbind(Success, Attempts - Success) ~ Distance +
    I(Distance^2), family = binomial(link = "logit"), data = nfl[Z ==
    1, ])

Deviance Residuals:
      6        7        8        9       10
 0.3187  -0.6829   0.7721  -0.5231   0.2853

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.892466   1.189274   4.114 3.89e-05 ***
Distance     -0.197046   0.074348  -2.650  0.00804 **
```

```
I(Distance^2)  0.001604   0.001098   1.461  0.14395
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 78.7794  on 4  degrees of freedom
Residual deviance:  1.5192  on 2  degrees of freedom
AIC: 28.443

Number of Fisher Scoring iterations: 3
```

(b) The fitted model is:

```
Call:
glm(formula = cbind(Success, Attempts - Success) ~ Distance +
    I(Distance^2) + Z, family = binomial(link = "logit"), data = nfl)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.86350  -0.20086   0.03301   0.55505   1.60112

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.5241844  0.7747832   4.549  5.4e-06 ***
Distance     -0.0958710  0.0490210  -1.956   0.0505 .
I(Distance^2) -0.0001086  0.0007365  -0.147   0.8828
Z             0.1037533  0.1698311   0.611   0.5413
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 228.5180  on 9  degrees of freedom
Residual deviance:   8.9776  on 6  degrees of freedom
AIC: 59.367

Number of Fisher Scoring iterations: 4
```

(c) From the above output, we can see that the quadratic term is not significant, with p-value=0.88.

(Students may use other kind of tests, like likelihood ratio test and F-test, and they are all right if they get similar p-value.)

(d) As (c) suggested, we first delete the quadratic term and then test whether Z is significant. The p-value = 0.54 from the summary table, so we accept that the probabilities are the same for each league. (Students may use other kind of tests, like F-test, and they are all right if they get similar p-value.)

A sample of source code:

```
#2a
nfl=read.table("http://www-stat.stanford.edu/~nzhang/203_web/Data/NFL.txt",
header=T)
attach(nfl)
nfl.glm=glm(cbind(Success,Attempts-Success)~Distance+I(Distance^2),
family=binomial(link='logit'),data=nfl[Z==0,])
summary(nfl.glm)
afl.glm=glm(cbind(Success,Attempts-Success)~Distance+I(Distance^2),
family=binomial(link='logit'),data=nfl[Z==1,])
summary(afl.glm)
#2b
all.glm=glm(cbind(Success,Attempts-Success)~Distance+I(Distance^2)+Z,
family=binomial(link='logit'),data=nfl)
summary(all.glm)
#2d
glm.nq <- glm(cbind(Success,Attempts-Success)~Distance+Z, family = binomial)
print(summary(glm.nq))
```

3. A sample of source code is attached in the end.

(a) The result of the regression is:

```
Call:
glm(formula = Falls ~ Intervention + Gender + Balance + Strength,
    family = poisson())

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1854  -0.7819  -0.2564   0.5449   2.3625

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)   0.489467   0.336869   1.453  0.14623
Intervention -1.069403   0.133154  -8.031 9.64e-16 ***
Gender      -0.046606   0.119970  -0.388  0.69766
Balance      0.009470   0.002953   3.207  0.00134 **
Strength     0.008566   0.004312   1.986  0.04698 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 199.19  on 99  degrees of freedom
Residual deviance: 108.79  on 95  degrees of freedom
AIC: 377.29

Number of Fisher Scoring iterations: 5
```

So the coefficients and their estimated standard deviations are in the above. The so the estimated response function is:

$$\hat{Y} = \exp(0.489 - 1.069X_1 - 0.0466X_2 + 0.00947X_3 + 0.00857X_4)$$

(b) The null hypothesis is: all coefficients of X_1, X_2, X_3, X_4 are zero. The alternative hypothesis is: not all coefficients of X_1, X_2, X_3, X_4 are zero. Decision rule: likelihood ratio test. We get that the likelihood ratio statistic ($\sim \chi^2_{(4)}$) = 90.4, with p-value almost zero ($< 1 \times 10^{-5}$), so we reject the null hypothesis. (Students may use other kind of tests, like F-test, ANOVA table, and they are all right if they get similar p-value.)

(c) The null hypothesis is: the coefficient of X_2 is zero. The alternative hypothesis is: the coefficient of X_2 is not zero. Decision rule: likelihood ratio test. We get that the likelihood ratio statistic ($\sim \chi^2_{(1)}$) = 0.151, with p-value = 0.70, so we accept the null hypothesis. (Students may use other kind of tests, like F-test, or ANOVA, and they are all right if they get similar p-value.)

Source code:

```
ger <- read.table('Geriatrics.txt', header = T)
print(ger)
attach(ger)

########### Part (a)
ger.glm <- glm(Falls ~ Intervention + Gender + Balance + Strength, family = poisson())
print(summary(ger.glm))
```

```
########### Part (b)
lr.n <- ger.glm$null.deviance - ger.glm$deviance
p.val.n <- 1 - pchisq(lr.n, ger.glm$df.null - ger.glm$df.residual)
cat("The likelihood ratio statistics is:", lr.n, ", p-value is: ", p.val.n, "\n")

########### Part (c)
ger.glm.ng <- glm(Falls ~ Intervention + Balance + Strength, family = poisson())
print(summary(ger.glm.ng))
lr.ng <- ger.glm.ng$deviance - ger.glm$deviance
p.val.ng <- 1 - pchisq(lr.ng, ger.glm.ng$df.residual - ger.glm$df.residual)
cat("The likelihood ratio statistics is:", lr.ng, ", p-value is: ", p.val.ng, "\n")
```