Stat 207 HW7

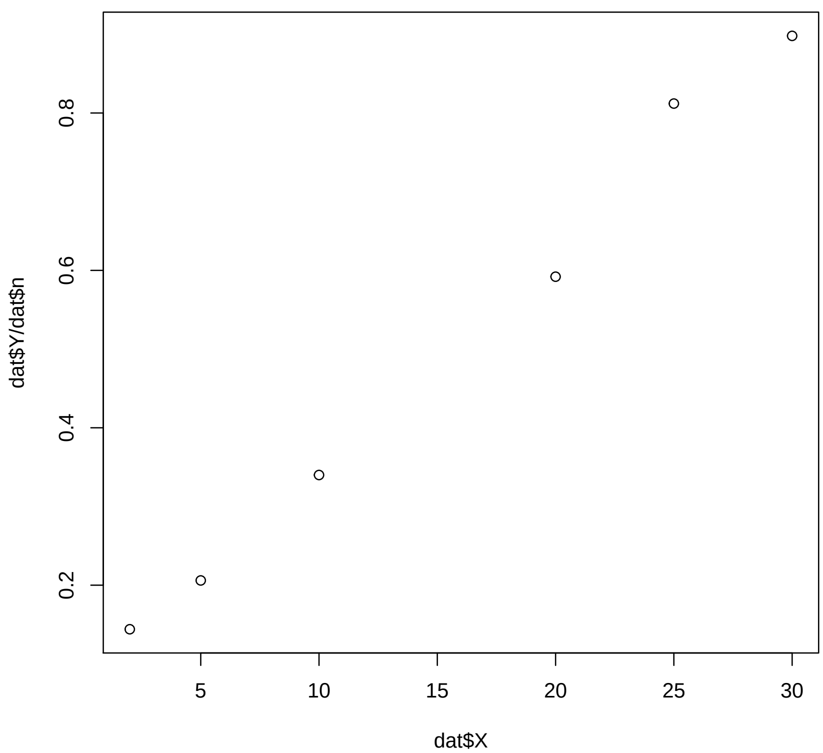
Cheng Luo 912466499
Fan Wu 912538518

March 3, 2015

1 14.11

(a)

```
dat = read.table("CH14PR11.txt")
names(dat) = c("X", "n", "Y")
plot(dat$X, dat$Y/dat$n)
```



The plot support the analyst's belief that the logistic response functiion is appropriate.

(b)

```
logit = glm(Y/n ~ X, data = dat, family = "binomial")
## Warning:  non-integer #successes in a binomial glm!
summary(logit)
##
## Call:
```

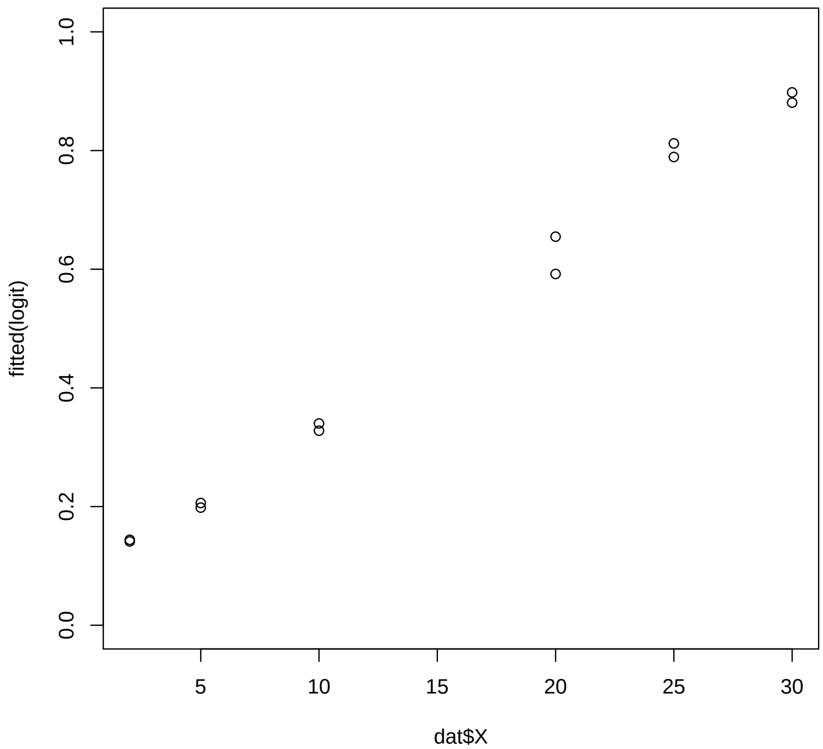
```
## glm(formula = Y/n ~ X, family = "binomial", data = dat)
##
## Deviance Residuals:
##         1          2          3          4          5          6
##  0.007846   0.019363   0.025865  -0.130556   0.056842   0.054601
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.0766     1.8970  -1.095    0.274
## X             0.1359     0.1067   1.273    0.203
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.216343  on 5  degrees of freedom
## Residual deviance: 0.024363  on 4  degrees of freedom
## AIC: 7.1154
##
## Number of Fisher Scoring iterations: 4
```

From the summary, the maximum likelihood estimates of $\hat{\beta}_0 = -2.0766$, $\hat{\beta}_1 = 0.1359$,

$$\hat{\pi} = \frac{exp(\beta_0 + \beta_1 X)}{1 + exp(\beta_0 + \beta_1 X)} = \frac{exp(-2.0766 + 0.1359X)}{1 + exp(-2.0766 + 0.1359X)}$$

(c)

```
plot(dat$X, fitted(logit), ylim = c(0, 1))
points(dat$X, dat$Y/dat$n, lwd = 1)
```



The fitted logistic response function appears to be well.

(d)
```
exp(0.1359)

## [1] 1.145567
```

$exp(\beta_1) = 1.145567$, so that the odds of the bottles being returned is increased by 14.5567% with each one deposit level increased.

(e)
```
newdat = data.frame(X = 15)
predict(logit, newdata = newdat, type = "response")

##         1
## 0.4903005
```

The estimated probability that a bottle will be returned when the deposit is 15 cents is 0.4903005.

(f)
```r
newpi = 0.75
pi_2 = log(newpi/(1-newpi))
(pi_2 - (-2.0766))/0.1359
```

```
## [1] 23.36433
```

Estimate the amount of deposit for which 75% of the bottles are expected to be returned is 23.36433.

2 14.14

(a)
```r
dat = read.table("CH14PR14.txt")
names(dat) = c("Y", "X1", "X2", "X3")
logit = glm(Y ~ X1 + X2 + X3, data = dat, family = "binomial")
summary(logit)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3, family = "binomial", data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4037  -0.5637  -0.3352  -0.1542   2.9394
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.17716    2.98242  -0.395  0.69307
## X1           0.07279    0.03038   2.396  0.01658 *
## X2          -0.09899    0.03348  -2.957  0.00311 **
## X3           0.43397    0.52179   0.832  0.40558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 134.94  on 158  degrees of freedom
## Residual deviance: 105.09  on 155  degrees of freedom
## AIC: 113.09
##
## Number of Fisher Scoring iterations: 6
```

From the summary, the maximum likelihood estimates $\hat{\beta}_0 = -1.17716$,

$\hat{\beta}_1 = 0.07279$, $\hat{\beta}_2 = -0.09899$, $\hat{\beta}_4 = 0.43397$

$$\hat{\pi} = \frac{exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3)}{1 + exp(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3)} = \frac{exp(-1.17716 + 0.07279X1 - 0.09899X2 + 0.43397X3)}{1 + exp(-1.17716 + 0.07279X1 - 0.09899X2 + 0.43397X3)}$$

(b)

```
exp(0.07279)

## [1] 1.075505

exp(-0.09899 )

## [1] 0.9057518

exp(0.43397)

## [1] 1.543373
```

- $exp(\beta_1) = 1.075505$, so that the odds of getting a flu shot is increased by 7.5% with each one age increased.
- $exp(\beta_2) = 0.9057518$, so that the odds of getting a flu shot is decreased by 9.4% with each one health awareness index increased.
- $exp(\beta_3) = 1.543373$, so that the odds of getting a flu shot is increased by 54.3% from woman to man.

(c)

```
newdat = data.frame(X1 = 55, X2 = 60, X3 = 1)
predict(logit, newdata = newdat, type = "response")

##          1
## 0.06422197
```

The estimated probability with X1=55, X2=60 and X3=1 is 0.06422197

.

3 14.17

(a)

```
dat = read.table("CH14PR11.txt")
names(dat) = c("X", "n", "Y")
logit = glm(Y/n ~ X, data = dat, family = "binomial")

## Warning:  non-integer #successes in a binomial glm!

summary(logit)
```

```
##
## Call:
## glm(formula = Y/n ~ X, family = "binomial", data = dat)
##
## Deviance Residuals:
##         1          2          3          4          5          6
##  0.007846   0.019363   0.025865  -0.130556   0.056842   0.054601
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.0766     1.8970  -1.095    0.274
## X             0.1359     0.1067   1.273    0.203
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.216343  on 5  degrees of freedom
## Residual deviance: 0.024363  on 4  degrees of freedom
## AIC: 7.1154
##
## Number of Fisher Scoring iterations: 4

  b1 = 0.1359
  s1 = 0.1067
  z = qnorm(1-0.05/2)
  c(b1-s1*z, b1+s1*z)

## [1] -0.07322816  0.34502816

  c(exp(b1-s1*z), exp(b1+s1*z))

## [1] 0.9293888 1.4120297
```

From summary(logit), we get $s(b_1) = 0.1067, b_1 = 0.1359$, based on $b_k \pm z(1-\alpha/2)sb_k$, we conclude that we are 95 % confident that β_1 is between -0.07322816 and 0.34502816, and corresponding confidence limits for the odds ratio $exp(\beta_1)$ is between 0.9293888 and 1.4120297.

(b)

```
  qnorm(1-0.05/2)

## [1] 1.959964
```

$$H_0{:}\beta_1 = 0$$
$$\text{VS. } H_1{:}\beta_1 \neq 0$$

$z^* = \frac{b_1}{s(b_1)} = 0.1359/0.1067 = 1.273664$
we can reject H_0 if $|z^*| > Z(1-0.05/2) = 1.959964$,otherwise reject$H_1$
so that reject H_1 because $|z^*| < 1.959964$,
therefore, X1 can be dropped from the regression model, and the P-value is 0.203

(c)
```r
logLik(logit)

## 'log Lik.' -1.557684 (df=2)

logitR = glm(Y/n ~ 1, data = dat, family = "binomial")

## Warning:  non-integer #successes in a binomial glm!

logLik(logitR)

## 'log Lik.' -4.158904 (df=1)

qchisq(1-0.05, 2-1)

## [1] 3.841459

pchisq(5.339856, 1, lower.tail = FALSE)

## [1] 0.02084319
```

H_0:$\beta_1 = 0$
VS. H_1:$\beta_1 \neq 0$
The full model: $\pi = [1 + exp(-(\beta_0 + \beta_1 X1))]^{-1}$
ln(L(F))= -1.557684
The reduced model: $\pi = [1 + exp(-(\beta_0))]^{-1}$
ln(L(R))= -4.158904
G^2 = -2(ln(L(R)-ln(L(F)))) = 5.20244
we can reject H_0 if $G^2 > \chi^2(1-0.05, 2-1) = 3.8415$,otherwise reject$H_1$
so that reject H_0 because $G^2 > 3.8415$,
therefore, X1 cannot be dropped from the regression model, and the P-value is 0.02084319.And the result is different from the result we get in (b).

4 14.20

(a)

```r
dat = read.table("CH14PR14.txt")
names(dat) = c("Y", "X1", "X2", "X3")
logit = glm(Y ~ X1 + X2 + X3, data = dat, family = "binomial")
summary(logit)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3, family = "binomial", data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4037  -0.5637  -0.3352  -0.1542   2.9394
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.17716    2.98242  -0.395  0.69307
## X1           0.07279    0.03038   2.396  0.01658 *
## X2          -0.09899    0.03348  -2.957  0.00311 **
## X3           0.43397    0.52179   0.832  0.40558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 134.94  on 158  degrees of freedom
## Residual deviance: 105.09  on 155  degrees of freedom
## AIC: 113.09
##
## Number of Fisher Scoring iterations: 6
```

$z(1-\frac{0.1}{2*2}) = 0.975, s(b_1) = 0.03038, s(b_2) = 0.03348$

$$exp(30(0.07280.030381.96)) < exp(30\beta_1) < exp(30(0.0728 + 0.030381.96))$$

$$1.4878 < exp(30\beta_1) < 52.9837$$

$$exp(25(0.0990.033481.96)) < exp(25\beta_2) < exp(25(0.0728 + 0.033481.96))$$

$$0.0163 < exp(25\beta_2) < 31.824$$

(b)

```r
qnorm(1-0.05/2)
```

```
## [1] 1.959964
```

H_0:$\beta_3 = 0$

VS. H_1:$\beta_3 \neq 0$

$z^* = \frac{b_3}{s(b_3)} = 0.43397/0.52179 = 0.8316947$

we can reject H_0 if $|z^*| > Z(1-0.05/2) = 1.959964$,otherwise reject$H_1$

so that reject H_1 because $|z^*| < 1.959964$,

therefore, X3 can be dropped from the regression model, and the P-value is 0.40558

(c)

```r
logLik(logit)

## 'log Lik.' -52.54659 (df=4)

logitR = glm(Y ~ X1+X2, data = dat, family = "binomial")
logLik(logitR)

## 'log Lik.' -52.89769 (df=3)

qchisq(1-0.05, 4-3)

## [1] 3.841459

pchisq(0.70236, 1, lower.tail = FALSE)

## [1] 0.4019918
```

H_0:$\beta_3 = 0$

VS. H_1:$\beta_3 \neq 0$

The full model: $\pi = [1 + exp(-(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3))]^{-1}$

ln(L(F))= -52.54659

The reduced model: $\pi = [1 + exp(-(\beta_0 + \beta_1 X1 + \beta_2 X2))]^{-1}$

ln(L(R))= -52.89769

G^2 = -2(ln(L(R)-ln(L(F)))) = 0.70236

we can reject H_0 if $G^2 > \chi^2(1-0.05, 4-3) = 3.8415$,otherwise reject$H_1$

so that reject H_1 because $G^2 < 3.8415$,

therefore, X3 can be dropped from the regression model, and the P-value is 0.4019918.And the result is the same as the result we get in (b).

(d)

```
  logitF = glm(Y ~ X1+X2+I(X1^2)+I(X2^2)+I(X1*X2), data = dat, family = "binomial")
  logLik(logitF)

## 'log Lik.' -52.13072 (df=6)

  logitR = glm(Y ~ X1+X2, data = dat, family = "binomial")
  logLik(logitR)

## 'log Lik.' -52.89769 (df=3)

  qchisq(1-0.05, 6-3)

## [1] 7.814728

  pchisq(1.53394, 3, lower.tail = FALSE)

## [1] 0.6744594
```

H_0:$\beta_3 = \beta_4 = \beta_5 = 0$

VS. H_1:$not all \beta_3, \beta_4, \beta_5 equal 0$

The full model:

$\pi = [1 + exp(-(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X1^2 + \beta_4 X2^2 + \beta_5 X1 * X2))]^{-1}$

ln(L(F))= -52.13072

The reduced model: $\pi = [1 + exp(-(\beta_0 + \beta_1 X1 + \beta_2 X2))]^{-1}$

ln(L(R))= -52.89769

G^2 = -2(ln(L(R)-ln(L(F)))) = 1.53394

we can reject H_0 if $G^2 > \chi^2(1 - 0.05, 6 - 3)$ = 7.814728,otherwise rejectH_1

so that reject H_1 because $G^2 <$ 7.814728,

therefore, $X1^2, X2^2, I(X1 * X2)$ can be dropped from the regression model, and the P-value is 0.6744594.And the result is the same as the result we get in (b).

5 14.22

(a)

(b)

(c)

(d)

6 14.23

```
  dat = read.table("CH14PR11.txt")
  names(dat) = c("X", "n", "Y")
  logit = glm(Y/n ~ X, data = dat, family = "binomial")

## Warning:  non-integer #successes in a binomial glm!

  Oj1 = dat$Y
  Ej1 = round(dat$n*fitted(logit), 1)
  Oj0 = dat$n-dat$Y
  Ej0 = dat$n-Ej1
  rbind(Oj1, Oj0)

##      [,1] [,2] [,3] [,4] [,5] [,6]
## Oj1   72  103  170  296  406  449
## Oj0  428  397  330  204   94   51

  rbind(Ej1, Ej0)

##          1     2     3     4     5     6
## Ej1  70.6  99.1 163.9 327.4 394.6 440.3
## Ej0 429.4 400.9 336.1 172.6 105.4  59.7

  X.squ = sum((rbind(Oj1, Oj0)-rbind(Ej1, Ej0))^2/rbind(Ej1, Ej0));X.squ

## [1] 12.28748
```

H_0:$E(Y) = [1 + exp(-\beta_0 - \beta_1 X1)]^{-1}$
VS. H_1:$E(Y) \neq [1 + exp(-\beta_0 - \beta_1 X1)]^{-1}$
$X^2 = \sum_j \sum_k \frac{(O_{jk} - E_{jk})^2}{E_{jk}} = 12.287$
we can reject H_0 if $X^2 > \chi^2(0.99, 3) = 13.2767$,otherwise reject$H_1$
so that reject H_1 because $X^2 < 13.2767$,

7 14.40

$$
\begin{aligned}
\frac{exp(\beta_0 + \beta_1 Xi)}{1 + exp(\beta_0 + \beta_1 Xi)} &= \frac{1}{\frac{1}{exp(\beta_0 + \beta_1 Xi)} + 1} \\
&= \frac{1}{exp(0 - \beta_0 - \beta_1 Xi) + 1} \\
&= [1 + exp(-\beta_0 - \beta_1 Xi)]^{-1}
\end{aligned}
$$

8 14.41

For given observations $Y_1, Y_2 ..., Y_n$, all terms with a given X value, X_j, we get

$$Y_{\cdot j}(\beta_0 + \beta_1 X_j) - n_j ln(1 + exp(\beta_0 + \beta_1 Xj)$$

Because there are $\binom{n_j}{Y_{\cdot j}}$ ways of getting these, we must add $ln\binom{n_j}{Y_{\cdot j}}$, hence we get

$$ln\binom{n_j}{Y_{\cdot j}} + Y_{\cdot j}(\beta_0 + \beta_1 X_j) - n_j ln(1 + exp(\beta_0 + \beta_1 Xj)$$

9 14.42

$$\pi_i = \frac{exp(\pi_i^{'})}{1 + exp(\pi_i^{'})}$$
$$1 - \pi_i = \frac{1}{1 + exp(\pi_i^{'})}$$
$$\frac{\pi_i}{1 - \pi_i} = exp(\pi_i^{'})$$
$$F_L^{-1}(\pi_i) = \pi_i^{'} = log_e(\frac{\pi_i}{1 - \pi_i})$$

10 14.43

$$lnL(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 X_i) - \sum_{i=1}^{n}(1 + exp(\beta_0 + \beta_1 X_i))$$
$$\frac{\partial^2 lnL}{\partial \beta_0^2} = -\sum_{i=1}^{n} \frac{exp(\beta_0 + \beta_1 X_i)}{[1 + exp(\beta_0 + \beta_1 X_i)]^2}$$
$$\frac{\partial^2 lnL}{\partial \beta_1^2} = -\sum_{i=1}^{n} \frac{X_i^2 exp(\beta_0 + \beta_1 X_i)}{[1 + exp(\beta_0 + \beta_1 X_i)]^2}$$
$$\frac{\partial^2 lnL}{\partial \beta_0 \partial \beta_1} = -\sum_{i=1}^{n} \frac{X_i exp(\beta_0 + \beta_1 X_i)}{[1 + exp(\beta_0 + \beta_1 X_i)]^2}$$
$$-E\{\frac{\partial^2 lnL}{\partial \beta_0 \partial \beta_1}\} = -g_{01} = -g_{10} \text{which is reduced to (14.51)}$$