Stat 207 HW6

Cheng Luo 912466499
Fan Wu 912538518

February 23, 2015

1 14.9

(a)

```
dat = read.table("CH14PR09.txt")
names(dat) = c("Y", "X")
logit = glm(Y ~ X, data = dat, family = "binomial")
summary(logit)
```

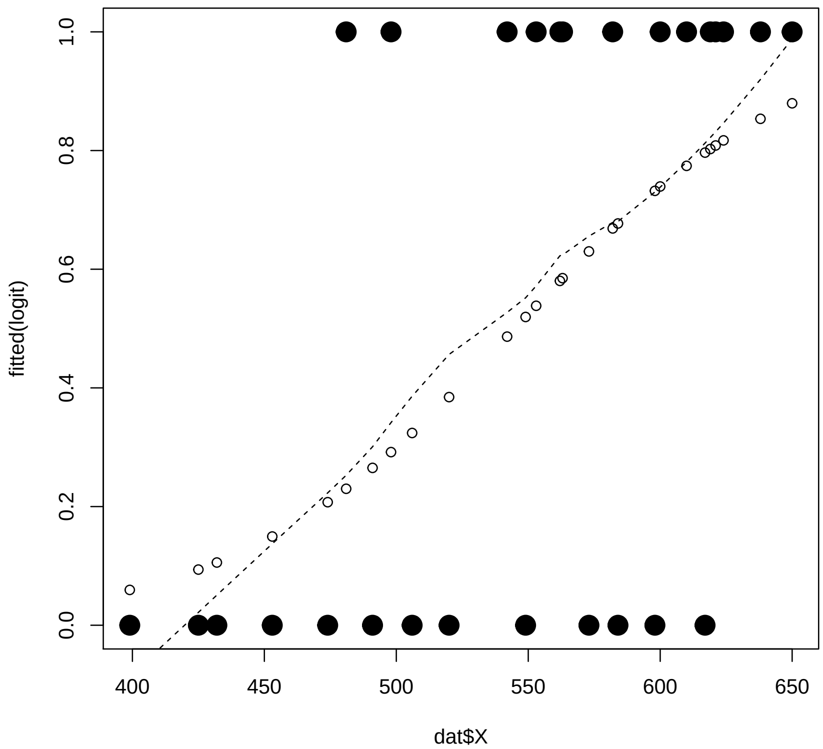
```
##
## Call:
## glm(formula = Y ~ X, family = "binomial", data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7845  -0.8350   0.5065   0.8371   1.7145
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.308925   4.376997  -2.355   0.0185 *
## X             0.018920   0.007877   2.402   0.0163 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 37.393  on 26  degrees of freedom
## Residual deviance: 29.242  on 25  degrees of freedom
## AIC: 33.242
##
## Number of Fisher Scoring iterations: 4
```

From the summary, the maximum likelihood estimates of $\hat{\beta}_0 = -10.308925$, $\hat{\beta}_1 = 0.018920$,

$$\hat{\pi} = \frac{exp(\beta_0 + \beta_1 X)}{1 + exp(\beta_0 + \beta_1 X)} = \frac{exp(-10.308925 + 0.018920X)}{1 + exp(-10.308925 + 0.018920X)}$$

(b)

```
plot(dat$X, fitted(logit), ylim = c(0, 1))
points(dat$X, dat$Y, lwd = 9)
lines(lowess(dat$Y ~ dat$X), lty = 2)
```



The fitted logistic response function appears to be well.

(c)
```
exp(0.018920)
```

```
## [1] 1.0191
```

$exp(\beta_1) = 1.0191$, so that the odds of employee's ability increased by 1.91% with each additional employee's emotional stability.

(d)
```
newdat = data.frame(X = 550)
predict(logit, newdata = newdat, type = "response")
```

```
##         1
## 0.5242263
```

The estimated probability that employees with an emotional stability test score of 550 will be able to perform in a task group is 0.5242263 .

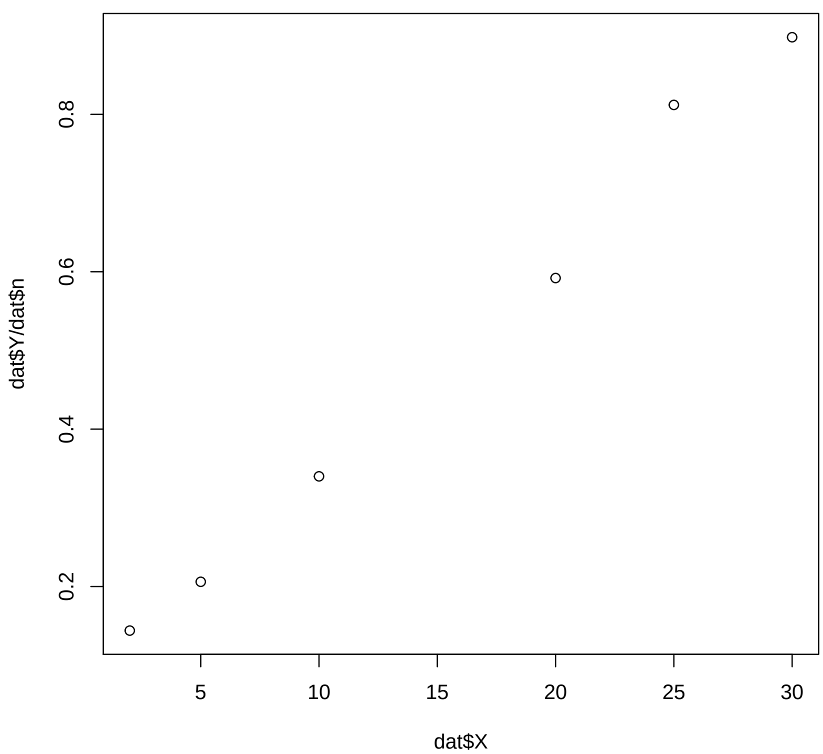
(e)
```r
newpi = 0.7
pi_2 = log(newpi/(1-newpi))
(pi_2 - (-10.308925))/0.018920
```

```
## [1] 589.6524
```

The emotional stability test score for which 70 percents of the employees with this test score are expected to be able to perform in a task group is 598.6524.

2 14.11

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



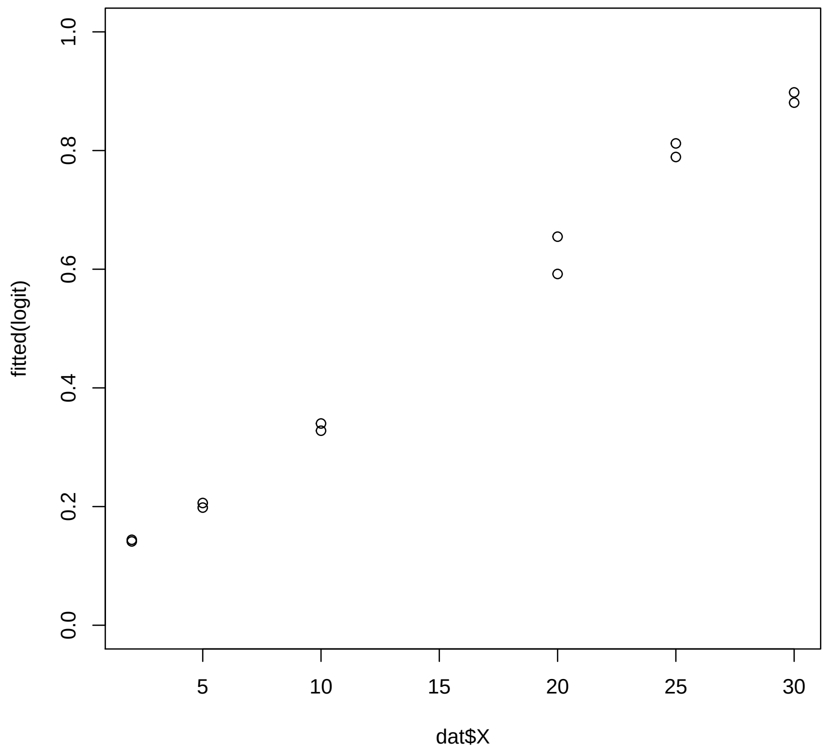
The plot support the analyst's belief that the logistic response functiion is appropriate.

(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.

3 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

.

4 14.19

(a)

```
dat = read.table("CH14PR13.txt")
names(dat) = c("Y", "X1", "X2")
logit = glm(Y ~ X1 + X2, data = dat, family = "binomial")
summary(logit)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2, family = "binomial", data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6189  -0.8949  -0.5880   0.9653   2.0846
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.73931    2.10195  -2.255   0.0242 *
## X1           0.06773    0.02806   2.414   0.0158 *
## X2           0.59863    0.39007   1.535   0.1249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 44.987  on 32  degrees of freedom
## Residual deviance: 36.690  on 30  degrees of freedom
## AIC: 42.69
##
## Number of Fisher Scoring iterations: 4
```

(b)

```
qnorm(1-0.05/2)

## [1] 1.959964
```

$$H_0 : \beta_2 = 0$$
$$\text{VS. } H_1 : \beta_2 \neq 0$$
$$z^* = \frac{b2}{s(b_2)} = 0.59863/0.39007 = 1.535$$

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, X2 can be dropped from the regression model, and the P-value is 0.1249

(c)

```
logLik(logit)

## 'log Lik.' -18.34482 (df=3)
```

```r
logitR = glm(Y ~ X1, data = dat, family = "binomial")
logLik(logitR)
```

```
## 'log Lik.' -19.65227 (df=2)
```

```r
qchisq(1-0.05, 3-2)
```

```
## [1] 3.841459
```

```r
pchisq(2.614, 1, lower.tail = FALSE)
```

```
## [1] 0.1059243
```

H_0:$\beta_2 = 0$
VS. H_1:$\beta_2 \neq 0$
The full model: $\pi = [1 + exp(-(\beta_0 + \beta_1 X1 + \beta_2 X2))]^{-1}$
L(F)= -18.34482
The reduced model: $\pi = [1 + exp(-(\beta_0 + \beta_1 X1))]^{-1}$
L(R)= -19.65227
G^2 = -2(ln(L(R)-ln(L(F)))) = 2.614
we can reject H_0 if $G^2 > \chi^2(1 - 0.05, 3 - 2) = 3.8415$,otherwise reject$H_1$
so that reject H_1 because $G^2 < 3.8415$,
therefore, X2 can be dropped from the regression model, and the P-value is 0.1059.And the result is same as the result we get in (b).

(d)

```r
logLik(logit)
```

```
## 'log Lik.' -18.34482 (df=3)
```

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

```
## 'log Lik.' -17.12634 (df=6)
```

```r
qchisq(1-0.05, 6-3)
```

```
## [1] 7.814728
```

```r
pchisq(2.436953, 3, lower.tail = FALSE)
```

```
## [1] 0.4867929
```

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

VS. H_1:not all $\beta_k = 0$,k=3,4,5

The full model:

$\pi = [1 + exp(-(\beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_4 X4 + \beta_5 X5))]^{-1}$

L(F)= -17.12634

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

L(R)= -18.34482

G^2 = -2(ln(L(R)-ln(L(F)))) = 2.436953

we can reject H_0 if $G^2 > \chi^2(1 - 0.05, 6 - 3) = 7.81$,otherwise reject$H_1$

so that reject H_1 because $G^2 < 7.81$,

therefore, X3,X4,X5 can be dropped from the regression model, and the P-value is 0.4867929.

5 Problem 5

(a)

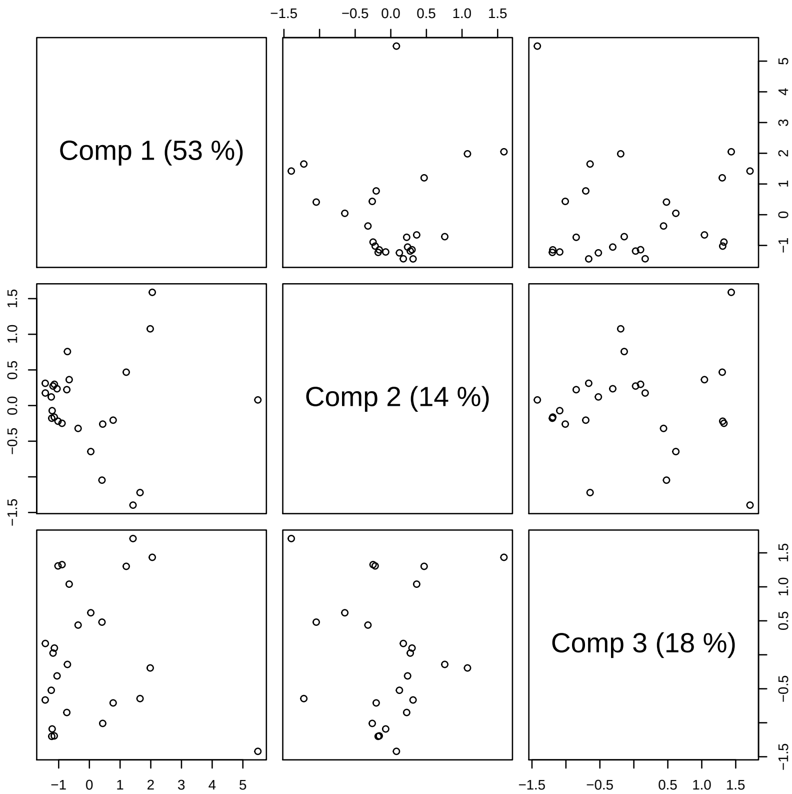
```
dat = read.table("apartment.txt", header = TRUE)
require("pls")

## Loading required package:  pls

## Warning:  package 'pls' was built under R version 3.1.2

##
## Attaching package:  'pls'
##
## The following object is masked from 'package:stats':
##
##     loadings

dat.stan = dat
for(j in 1:ncol(dat))
  dat.stan[,j] = (dat[,j] - mean(dat[,j]))/sd(dat[,j])
dat = dat.stan
fit = plsr(Y ~ X1 + X2 + X3 + X4 + X5, data = dat, 5, validation="CV")
summary(fit)

## Data:  X dimension: 25 5
##  Y dimension: 25 1
## Fit method: kernelpls
## Number of components considered: 5
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
```

```
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps
## CV           1.021   0.3139   0.2964   0.2419   0.1853   0.2259
## adjCV        1.021   0.3094   0.2901   0.2340   0.1823   0.2213
##
## TRAINING: % variance explained
##    1 comps  2 comps  3 comps  4 comps  5 comps
## X    52.58    66.94    85.17    91.35   100.00
## Y    92.14    96.53    97.92    98.01    98.05
```

```
loadings(fit)[, 1:3]
```

```
##         Comp 1     Comp 2      Comp 3
## X1 -0.07983743  0.6903977 -0.76418606
## X2  0.59805736  0.1008482 -0.08164449
## X3  0.54164949 -0.5440785 -0.25282151
## X4  0.16890957 -0.7416964  0.59414237
## X5  0.56738864  0.5744813  0.04300711
```

```
plot(fit, plottype = "scores", comps = 1:3)
```



(b)
```
Radj = c(0, 95.1 ,   95.21   ,  96.61   ,  98.05 , 98.05)/100
n = 25
ans = integer(5)
for (i in 2:6)
{
  ans[i-1] = (n-i-2)*(Radj[i]-Radj[i-1])/(1-Radj[i])
}
```

(c)

6 Problem 6

7 Problem 7

8 Problem 8