

Arizona State University

From the Selected Works of Joseph M Hilbe

May, 2009

Logistic Regression Using R

Joseph Hilbe, *Arizona State University*



Available at: https://works.bepress.com/joseph_hilbe/22/

LOGISTIC REGRESSION MODELS USING R

Chapman & Hall/CRC Press, 2009

Joseph M. Hilbe

R

Code and Output

from

Logistic Regression Models

Chapter 1

No statistical output in this chapter

Chapter 2

SECTION 2.1: 2x2 TABLES

```
> library('foreign') #foreign import package library
> ds <- read.dta('heart01x.dta') #works for STATA v.5-9

> #drop all other vars and factor levels except 'death' and 'anterior'
> simple<- data.frame(death=ds$death, anterior=ds$anterior[,drop=TRUE])
> head(simple)

> #contingency table information
> tsimple<- ftable(simple) #frequency table
> tsimple #show

> rowSums(tsimple)
> colSums(tsimple)
> sum(tsimple) #total number data

> #spiffier format
> library('gmodels')
> CrossTable(simple$death, simple$anterior, dnn=c('death','anterior'))

> #odds ratio
> tsimple[1,1]*tsimple[2,2]/(tsimple[2,1]*tsimple[1,2]) #compute odds ratio
from 2x2 table

> #logistic command (actually IRLS in R):
> fit <- glm(death ~ anterior, data=simple, family=binomial(link='logit'))
> summary(fit)

> exp(coef(fit)) #odds ratios corresponding to intercept and anterior coef

> confint(fit)
> exp(confint(fit)) #odds ratios

> #predictions of probability
> eta <- coef(fit)[1] + coef(fit)[2]*1 #predictor for anterior=1 (anterior MI)
> eta

> p<- 1/(1+exp(-eta)) #probability
> p
> exp(eta)/(1+exp(eta)) #different formula

> eta<- coef(fit)[1] + coef(fit)[2]*0 #predictor for anterior=0 (inferior MI)
> eta
> p<- 1/(1+exp(-eta)) #probability
```

```

> p

> #get all fitted values
> eta<- predict(fit, data.frame(anterior=simple$anterior), type='link',
na.action=na.omit)
> CrossTable(eta)

> p<- predict(fit, data.frame(anterior=simple$anterior), type='response',
na.action=na.omit)
> CrossTable(p)

> sqrt(1/120 + 1/2005 + 1/67 + 1/2504)
[1] 0.1554246

> fit <- glm(death ~ anterior, data=simple, family=binomial(link='logit'))
> summary(fit)

```

OUTPUT: 2.1

```

> library('foreign') #foreign import package library
> ds <- read.dta('heart01x.dta') #works for STATA v.5-9

> #drop all other vars and factor levels except 'death' and 'anterior'
> simple<- data.frame(death=ds$death, anterior=ds$anterior[,drop=TRUE])
> head(simple)
  death anterior
1      0 Anterior
2      1 Inferior
3      0 Inferior
4      1 Anterior
5      0 Anterior
6      0 Anterior

> #contingency table information
> tsimple<- ftable(simple) #frequency table
> tsimple #show
      anterior Inferior Anterior
death
0      2504      2005
1       67      120

> rowSums(tsimple)
[1] 4509 187
> colSums(tsimple)
[1] 2571 2125
> sum(tsimple) #total number data
[1] 4696

> #spiffier format
> library('gmodels')
> CrossTable(simple$death, simple$anterior, dnn=c('death','anterior'))

```

```

      Cell Contents
|-----|
|              N |
| Chi-square contribution |
|      N / Row Total |
|      N / Col Total |
|      N / Table Total |
|-----|

```

Total Observations in Table: 4696

death	anterior		Row Total
	Inferior	Anterior	
0	2504	2005	4509
	0.507	0.613	
	0.555	0.445	0.960
	0.974	0.944	
	0.533	0.427	
1	67	120	187
	12.227	14.793	
	0.358	0.642	0.040
	0.026	0.056	
	0.014	0.026	
Column Total	2571	2125	4696
	0.547	0.453	

```
> #odds ratio
> tsimple[1,1]*tsimple[2,2]/(tsimple[2,1]*tsimple[1,2]) #compute odds ratio
from 2x2 table
[1] 2.236796
```

```
> #logistic command (actually IRLS in R):
> fit <- glm(death ~ anterior, data=simple, family=binomial(link='logit'))
> summary(fit)
```

```
Call:
glm(formula = death ~ anterior, family = binomial(link = "logit"),
    data = simple)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3410 -0.3410 -0.2298 -0.2298  2.7009
```

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.6210     0.1238  -29.25  < 2e-16 ***
anteriorAnterior  0.8050     0.1554   5.18 2.22e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 1572.0 on 4695 degrees of freedom
Residual deviance: 1543.8 on 4694 degrees of freedom
(692 observations deleted due to missingness)
AIC: 1547.8
```

```
Number of Fisher Scoring iterations: 6
```

```
> exp(coef(fit)) #odds ratios corresponding to intercept and anterior coef
(Intercept) anteriorAnterior
  0.02675719    2.23679607
> confint(fit)
Waiting for profiling to be done...
```

```

                2.5 %    97.5 %
(Intercept)      -3.8732890 -3.387273
anteriorAnterior  0.5039546  1.114299
> exp(confint(fit)) #odds ratios
Waiting for profiling to be done...
                2.5 %    97.5 %
(Intercept)      0.02078988 0.03380074
anteriorAnterior 1.65525429 3.04743118

> #predictions of probability
> eta <- coef(fit)[1] + coef(fit)[2]*1 #predictor for anterior=1 (anterior MI)
> eta
(Intercept)
-2.815908
> p<- 1/(1+exp(-eta)) #probability
> p
(Intercept)
0.05647059
> exp(eta)/(1+exp(eta)) #different formula
(Intercept)
0.05647059
>
> eta<- coef(fit)[1] + coef(fit)[2]*0 #predictor for anterior=0 (inferior MI)
> eta
(Intercept)
-3.620952
> p<- 1/(1+exp(-eta)) #probability
> p
(Intercept)
0.0260599

> #get all fitted values
> eta<- predict(fit, data.frame(anterior=simple$anterior), type='link',
na.action=na.omit)
> CrossTable(eta)

```

```

      Cell Contents
|-----|
|              N |
|      N / Table Total |
|-----|

```

Total Observations in Table: 4696

```

| -3.62095211271667 | -2.81590759695862 |
|-----|-----|
|              2571 |              2125 |
|              0.547 |              0.453 |
|-----|-----|

```

```

> p<- predict(fit, data.frame(anterior=simple$anterior), type='response',
na.action=na.omit)
> CrossTable(p)

```

```

      Cell Contents
|-----|

```

	N
N / Table Total	

Total Observations in Table: 4696

0.0260598988748859	0.0564705882352941

2571	2125
0.547	0.453

```
> sqrt(1/120 + 1/2005 + 1/67 + 1/2504)
[1] 0.1554246
```

SECTION 2.2: 2xk TABLES

```
> #2 x k logistic
> CrossTable(ds$death, ds$killip, dnn=c('death','killip'))

> fit2 <- glm(death ~ kk2 + kk3 + kk4, data=ds, family=binomial(link='logit'))
> summary(fit2)

> exp(coef(fit2)) #odds ratios
> confint(fit2)
> exp(confint(fit2)) #odds ratios

> 97/3671
[1] 0.02642332
> 71/958
[1] 0.07411273
> (71/958)/(97/3671)
[1] 2.804823
> 30/262
[1] 0.1145038
> (30/262)/(97/3671)
[1] 4.333438
> 27/37
[1] 0.7297297
> (27/37)/(97/3671)
[1] 27.61688

> fit3 <- glm(death ~ kk1 + kk3 + kk4, data=ds, family=binomial(link='logit'))
> summary(fit3)

> exp(coef(fit3)) #odds ratios
> exp(confint(fit3)) #odds ratios CI

> (27*958)/(71*37)

> vcov(fit2) #covariance matrix
```

```
> exp(confint(fit2, level=.99)+ )
```

OUTPUT: 2.2

```
> #2 x k logistic
> CrossTable(ds$death, ds$killip, dnn=c('death','killip'))
```

```

Cell Contents
|-----|
|              N |
| Chi-square contribution |
|      N / Row Total |
|      N / Col Total |
|      N / Table Total |
|-----|

```

Total Observations in Table: 5153

death	killip				Row Total
	1	2	3	4	
0	3671	958	262	37	4928
	1.265	0.691	1.066	9.573	
	0.745	0.194	0.053	0.008	0.956
	0.974	0.931	0.897	0.578	
	0.712	0.186	0.051	0.007	
1	97	71	30	27	225
	27.714	15.127	23.339	209.665	
	0.431	0.316	0.133	0.120	0.044
	0.026	0.069	0.103	0.422	
	0.019	0.014	0.006	0.005	
Column Total	3768	1029	292	64	5153
	0.731	0.200	0.057	0.012	

```
> fit2 <- glm(death ~ kk2 + kk3 + kk4, data=ds, family=binomial(link='logit'))
> summary(fit2)
```

Call:

```
glm(formula = death ~ kk2 + kk3 + kk4, family = binomial(link = "logit"),
    data = ds)
```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-1.0469  -0.2284  -0.2284  -0.2284   2.7054

```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6335     0.1029  -35.323  < 2e-16 ***
kk2           1.0313     0.1603   6.432 1.26e-10 ***
kk3           1.4664     0.2185   6.712 1.92e-11 ***
kk4           3.3184     0.2732  12.146  < 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: 1849.1  on 5152  degrees of freedom
Residual deviance: 1698.6  on 5149  degrees of freedom
(235 observations deleted due to missingness)
AIC: 1706.6

Number of Fisher Scoring iterations: 6

> exp(coef(fit2)) #odds ratios
(Intercept)      kk2      kk3      kk4
  0.02642332  2.80482319  4.33343826 27.61688492
> confint(fit2)
              2.5 %      97.5 %
(Intercept) -3.8418009 -3.438108
kk2          0.7142288  1.343891
kk3          1.0232689  1.882675
kk4          2.7761204  3.851398
> exp(confint(fit2)) #odds ratios
              2.5 %      97.5 %
(Intercept)  0.02145493  0.0321254
kk2          2.04261080  3.8339342
kk3          2.78227485  6.5710617
kk4          16.05660682 47.0588070

> 97/3671
[1] 0.02642332
> 71/958
[1] 0.07411273
> (71/958)/(97/3671)
[1] 2.804823
> 30/262
[1] 0.1145038
> (30/262)/(97/3671)
[1] 4.333438
> 27/37
[1] 0.7297297
> (27/37)/(97/3671)
[1] 27.61688

> fit3 <- glm(death ~ kk1 + kk3 + kk4, data=ds, family=binomial(link='logit'))
> summary(fit3)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.6022      0.1230  -21.156  < 2e-16 ***
kk1           -1.0313      0.1603   -6.432 1.26e-10 ***
kk3            0.4350      0.2286    1.903  0.0571 .
kk4            2.2871      0.2814    8.127 4.39e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1849.1  on 5152  degrees of freedom
Residual deviance: 1698.6  on 5149  degrees of freedom
(235 observations deleted due to missingness)
AIC: 1706.6

Number of Fisher Scoring iterations: 6

```

```

> exp(coef(fit3)) #odds ratios
(Intercept)      kk1      kk3      kk4
0.07411273 0.35652871 1.54499516 9.84621241
> exp(confint(fit3)) #odds ratios CI
              2.5 %      97.5 %
(Intercept) 0.05772916 0.09356497
kk1         0.26082868 0.48956952
kk3         0.97452988 2.39556421
kk4         5.63959390 17.06465157

> (27*958)/(71*37)
[1] 9.846212

> vcov(fit2) #covariance matrix
              (Intercept)      kk2      kk3      kk4
(Intercept) 0.01058151 -0.01058151 -0.01058151 -0.01058151
kk2         -0.01058151 0.02570986 0.01058151 0.01058151
kk3         -0.01058151 0.01058151 0.04773164 0.01058151
kk4         -0.01058151 0.01058151 0.01058151 0.07464558

> exp(confint(fit2, level=.99)
+ )
Waiting for profiling to be done...
              0.5 %      99.5 %
(Intercept) 0.02003855 0.03407536
kk2         1.84614083 4.22784454
kk3         2.40388885 7.45792268
kk4         13.48606622 55.64096728

```

SECTION 2.3: MODELING A QUANTITATIVE PREDICTOR

```

> #quantitative predictor
> fit4<- glm(death ~ age, data=ds, family=binomial)
> summary(fit4)

> confint(fit4)
> exp(coef(fit4)) #odds ratios
> exp(confint(fit4)) #odds ratios CI

> #histogram and normal tests for age
> layout(1:2)
> hist(ds$age, col='blue', breaks=15, freq=FALSE)
> lines(density(ds$age), col='red')
> qqnorm(ds$age, col='blue')
> qqline(ds$age, col='red')

> shapiro.test(sample(na.omit(ds$age), 1000, replace=TRUE)) #only handles up to
5000 data

```

OUTPUT: 2.3

```

> #quantitative predictor
> fit4<- glm(death ~ age, data=ds, family=binomial)
> summary(fit4)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6569  -0.3375  -0.2559  -0.1780   3.1060

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.076784   0.372890  -18.98  <2e-16 ***
age          0.056529   0.004819   11.73  <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: 1974.8  on 5387  degrees of freedom
Residual deviance: 1826.6  on 5386  degrees of freedom
AIC: 1830.6

Number of Fisher Scoring iterations: 6

> confint(fit4)
              2.5 %      97.5 %
(Intercept) -7.82133752 -6.35896946
age          0.04716326  0.06606394
> exp(coef(fit4)) #odds ratios
(Intercept)          age
0.00084444846 1.0581574608
> exp(confint(fit4)) #odds ratios CI
              2.5 %      97.5 %
(Intercept) 0.0004010849 0.001731150
age          1.0482931381 1.068295017

```

SECTION 2.4: LOGISTIC MODELING DESIGNS

No statistical output in this section.

Chapter 3

No statistical output in this chapter.

Chapter 4

No example output in Sections 1 and 3.

SECTION 4.2:

```
library('foreign') #foreign import package library
```

```
heart01<- read.dta('heart01.dta') #works for STATA v.5-9
head(heart01) #show a few records
nrow(heart01) #show number of rows

fit1<- glm(death ~ anterior, data=heart01, family=binomial(link='logit'))
summary(fit1)
```

OUTPUT: 4.2:

```
> library('foreign') #foreign import package library
> heart01<- read.dta('heart01.dta') #works for STATA v.5-9
> head(heart01)
  death anterior hcabg kk1 kk2 kk3 kk4 age1 age2 age3 age4 center killip agegrp
1     0 Anterior     0  1  0  0  0     1   0   0   0   1255     1   =<60
2     1 Inferior     0  0  1  0  0     0   0   0   1   1255     2   >=80
3     0 Inferior     0  1  0  0  0     0   0   1   0   1255     1  70-79
4     1 Anterior     0  0  1  0  0     0   0   1   0   1255     2  70-79
5     0 Anterior     0  0  1  0  0     0   1   0   0   1255     2  60-69
6     0 Anterior     0 NA  NA  NA  NA     1   0   0   0   1255    NA   =<60
  age kk23
1  55     0
2  98     1
3  76     0
4  79     1
5  66     1
6  40    NA
> nrow(heart01)
[1] 5388
> fit1<- glm(death ~ anterior, data=heart01, family=binomial(link='logit'))
> summary(fit1)
```

Call:

```
glm(formula = death ~ anterior, family = binomial(link = "logit"),
    data = heart01)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.3410	-0.3410	-0.2298	-0.2298	2.7009

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.6210	0.1238	-29.25	< 2e-16 ***
anteriorAnterior	0.8050	0.1554	5.18	2.22e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1572.0 on 4695 degrees of freedom
 Residual deviance: 1543.8 on 4694 degrees of freedom
 (692 observations deleted due to missingness)
 AIC: 1547.8

Number of Fisher Scoring iterations: 6

Chapter 5

SECTION 5.1:

```
#preliminaries to get dataset
#DATA: 'heart01.dta'
#death: 1=yes, 0=no
#anterior: Anterior, Inferior
#age: in years
#hcabg: History of 1=CABG, 0=PTCA
#killip: Level 1-4 (4 more severe)
#agegrp: 4 age groups
#center: Provider or Center ID
library('foreign')
heart<- read.dta('heart01.dta') #read Stata format file
heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that complicate
tabulations
heart$center<- factor(heart$center) #convert to factor from numeric levels
heart$killip<- factor(heart$killip) #convert to factor from numeric levels

head(heart) #show some records
attach(heart) #make fields available directly as names

#Section 5.1:
t1<-table(hcabg) #frequency table
t1 #show
cat(format(t1/sum(t1), digits=3),'\n') #show fractions in each level
sum(t1) #total number

t2<-table(killip)
t2 #show
cat(format(t2/sum(t2), digits=3),'\n') #show fractions in each level
sum(t2)

t3<-table(agegrp)
t3 #show
cat(format(t3/sum(t3), digits=3),'\n') #show fractions in each level
sum(t3)

#construct age variable from categorized age
#(demo only, to keep 'age' from data intact for comparisons)
ageNew<- ifelse(age1==1, 40+trunc(21*runif(1)),
  ifelse(age2==1, 61+trunc(10*runif(1)),
    ifelse(age3==1, 71+trunc(10*runif(1)),
      ifelse(age4==1, 81+trunc(20*runif(1)),NA)))

#construct dummy variables for killip and age as 4 categories
#(demo only, to keep kk1-kk4 and age1-age4 already in dataset)
#NOTE: Dummy variables are generated automatically by glm().
kk1New<- ifelse(killip=='1', 1, 0)
kk2New<- ifelse(killip=='2', 1, 0)
kk3New<- ifelse(killip=='3', 1, 0)
kk4New<- ifelse(killip=='4', 1, 0)

age1<- ifelse(agegrp=='<-60', 1, 0)
age2<- ifelse(agegrp=='60-69', 1, 0)
age3<- ifelse(agegrp=='70-79', 1, 0)
age4<- ifelse(agegrp=='>=80', 1, 0)
```

```

#describe dataset heart01:
summary(heart)
str(heart) #show fields and encoding

#ulogit function to show univariate fit
ulogit <- function (y,x) {
  f<- glm(y ~ x, family=binomial())
  if (is.factor(x) && length(levels(x))>2) {
    print(anova(f, test='Chisq'))
  } else {
    b<- coef(f)[2]
    ci<- confint(f)[2,]
    cat('NullDeviance Deviance          OR          P-value          LCL          UCL\n')

cat(format(f$null.deviance,nsmall=2,width=10),format(f$deviance,nsmall=2,width=
10),

format(exp(b),nsmall=4,width=10),format(pchisq(b*b/vcov(f)[2,2],1,lower.tail=FA
LSE),

nsmall=4,width=10),format(exp(ci[1]),nsmall=4,width=10),format(exp(ci[2]),
nsmall=4,width=10),'\n')
  }
}

ulogit(death,anterior)
ulogit(death,hcabg)
ulogit(death,age)

#helper function to summarize logistic fit
showLogistic<- function (fit) {
  s<-summary(fit)
  print(s)
  b<- coef(fit) #coefficients
  o<- exp(b) #odds ratios
  CI<- confint(fit) #confidence intervals for coeff
  oCI<- exp(CI) #confidence intervals for odds ratios
  cat('\n')
  cat('Variable          Coefficient    LCL          UCL\n')
  for (i in 1:length(b)) {
    cat(format(names(o)[i],width=15),formatC(b[i],digits=5,width=10),

format(CI[i,1],digits=5,width=10),formatC(CI[i,2],digits=5,width=10),'\n')
  }
  cat('\n')
  cat('Variable          Odds Ratio    LCL          UCL\n')
  for (i in 1:length(o)) {
    cat(format(names(o)[i],width=15),formatC(o[i],digits=5, width=10),
formatC(oCI[i,1],digits=5,width=10),formatC(oCI[i,2],digits=5,
width=10),'\n')
  }
  cat('\n')
}

#killip factor fits:
fitk<- glm(death ~ kk2+kk3+kk4, data=heart, family=binomial())
showLogistic(fitk)
fitk2<- glm(death ~ killip, data=heart, family=binomial()) #note how dummy vars
are created
summary(fitk2) #same fit as above

tk<-table(factor(killip, exclude=NULL)) #include NA's

```

```

tk #show
cat(format(tk/sum(tk), digits=3),'\n') #show fractions in each level
sum(tk)

#agegrp factor fits:
fitage<- glm(death ~ age2+age3+age4, data=heart, family=binomial())
showLogistic(fitage)
fitage2<- glm(death ~ agegrp, data=heart, family=binomial()) #dummy vars auto-
created
summary(fitage2) #slight differences

fitage3<- glm(death ~ age1+age2+age3, data=heart, family=binomial())
showLogistic(fitage3)

#full model
fitFull<- glm(death ~ anterior+hcabg+age+kk2+kk3+kk4, data=heart,
family=binomial())
showLogistic(fitFull)

#Full model with categorical age
fitFull2<- glm(death ~ anterior+hcabg+kk2+kk3+kk4+age2+age3+age4, data=heart,
family=binomial())
showLogistic(fitFull2)

#drop age2 (combine with age1)
fitRed<- glm(death ~ anterior+hcabg+kk2+kk3+kk4+age3+age4, data=heart,
family=binomial())
showLogistic(fitRed)

#likelihood ratio test (deviance difference)
anova(fitRed,fitFull2,test='Chisq')

```

OUTPUT: 5.1

```

> source('ulogit.r') #get univariate logit display program
> source('tab.r') #get improved table display program
> source('showLogistic.r') #display summary of logistic fit
>
> #preliminaries to get dataset
> #DATA: 'heart01.dta'
> #death: 1=yes, 0=no
> #anterior: Anterior, Inferior
> #age: in years
> #hcabg: History of 1=CABG, 0=PTCA
> #killip: Level 1-4 (4 more severe)
> #agegrp: 4 age groups
> #center: Provider or Center ID
> library('foreign')
> heart<- read.dta('heart01.dta') #read Stata format file
> heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that
complicate tabulations
> heart$center<- factor(heart$center) #convert to factor from numeric levels
> heart$killip<- factor(heart$killip) #convert to factor from numeric levels
>
> head(heart) #show some records
  death anterior hcabg kk1 kk2 kk3 kk4 age1 age2 age3 age4 center killip agegrp
1     0 Anterior     0   1   0   0   0     1     0     0     0  1255     1  =<60
2     1 Inferior     0   0   1   0   0     0     0     0     1  1255     2  >=80
3     0 Inferior     0   1   0   0   0     0     0     1     0  1255     1  70-79
4     1 Anterior     0   0   1   0   0     0     0     1     0  1255     2  70-79
5     0 Anterior     0   0   1   0   0     0     1     0     0  1255     2  60-69
6     0 Anterior     0  NA  NA  NA  NA     1     0     0     0  1255    <NA>  =<60

```

```

    age kk23
1  55    0
2  98    1
3  76    0
4  79    1
5  66    1
6  40   NA
> attach(heart) #make fields available directly as names
>
> #Section 5.1:
> t1<-table(hcabg) #frequency table
> t1 #show
hcabg
  0    1
5207 181
> cat(format(t1/sum(t1), digits=3),'\n') #show fractions in each level
0.9664 0.0336
> sum(t1) #total number
[1] 5388
>
> t2<-table(killip)
> t2 #show
killip
  1    2    3    4
3768 1029 292   64
> cat(format(t2/sum(t2), digits=3),'\n') #show fractions in each level
0.7312 0.1997 0.0567 0.0124
> sum(t2)
[1] 5153
>
> t3<-table(agegrp)
> t3 #show
agegrp
=<60 60-69 70-79 >=80
1907 1390 1425   666
> cat(format(t3/sum(t3), digits=3),'\n') #show fractions in each level
0.354 0.258 0.264 0.124
> sum(t3)
[1] 5388
>
> #construct age variable from categorized age
> #(demo only, to keep 'age' from data intact for comparisons)
> ageNew<- ifelse(age1==1, 40+trunc(21*runif(1)),
+   ifelse(age2==1, 61+trunc(10*runif(1)),
+   ifelse(age3==1, 71+trunc(10*runif(1)),
+   ifelse(age4==1, 81+trunc(20*runif(1)),NA)))
>
> #construct dummy variables for killip and age as 4 categories
> #(demo only, to keep kk1-kk4 and age1-age4 already in dataset)
> #NOTE: Dummy variables are generated automatically by glm().
> kk1New<- ifelse(killip=='1', 1, 0)
> kk2New<- ifelse(killip=='2', 1, 0)
> kk3New<- ifelse(killip=='3', 1, 0)
> kk4New<- ifelse(killip=='4', 1, 0)
>
> age1<- ifelse(agegrp=='<=60', 1, 0)
> age2<- ifelse(agegrp=='60-69', 1, 0)
> age3<- ifelse(agegrp=='70-79', 1, 0)
> age4<- ifelse(agegrp=='>=80', 1, 0)
>
> #describe dataset heart01:
> summary(heart)
      death      anterior      hcabg      kk1

```



```

Min.      :0.00000    Inferior:2571    Min.      :0.00000    Min.      : 0.0000
1st Qu.:0.00000    Anterior:2125    1st Qu.:0.00000    1st Qu.: 0.0000
Median :0.00000    NA's      : 692    Median :0.00000    Median : 1.0000
Mean    :0.04491                    Mean    :0.03359    Mean    : 0.7312
3rd Qu.:0.00000                    3rd Qu.:0.00000    3rd Qu.: 1.0000
Max.    :1.00000                    Max.    :1.00000    Max.    : 1.0000
                                     NA's     :235.0000

      kk2      kk3      kk4      age1
Min.      : 0.0000    Min.      : 0.00000    Min.      : 0.00000    Min.      :0.0000
1st Qu.: 0.0000    1st Qu.: 0.00000    1st Qu.: 0.00000    1st Qu.:0.0000
Median : 0.0000    Median : 0.00000    Median : 0.00000    Median :0.0000
Mean    : 0.1997    Mean    : 0.05667    Mean    : 0.01242    Mean    :0.3539
3rd Qu.: 0.0000    3rd Qu.: 0.00000    3rd Qu.: 0.00000    3rd Qu.:1.0000
Max.    : 1.0000    Max.    : 1.00000    Max.    : 1.00000    Max.    :1.0000
NA's     :235.0000    NA's     :235.00000    NA's     :235.00000

      age2      age3      age4      center
Min.      : 0.0000    Min.      : 0.0000    Min.      : 0.0000    1255      : 295
1st Qu.: 0.0000    1st Qu.: 0.0000    1st Qu.: 0.0000    3749      : 255
Median : 0.0000    Median : 0.0000    Median : 0.0000    8332      : 251
Mean    : 0.2591    Mean    : 0.2656    Mean    : 0.1241    5863      : 245
3rd Qu.: 1.0000    3rd Qu.: 1.0000    3rd Qu.: 0.0000    6331      : 245
Max.    : 1.0000    Max.    : 1.0000    Max.    : 1.0000    4382      : 244
NA's     :23.0000    NA's     :23.0000    NA's     :23.0000    (Other):3853

      killip      agegrp      age      kk23
1      :3768    =<60 :1907    Min.      : 40.00    Min.      : 0.0000
2      :1029    60-69:1390    1st Qu.: 54.00    1st Qu.: 0.0000
3      : 292    70-79:1425    Median : 66.00    Median : 0.0000
4      : 64     >=80 : 666    Mean    : 65.81    Mean    : 0.2564
NA's: 235                    3rd Qu.: 76.00    3rd Qu.: 1.0000
                                     Max.    :100.00    Max.    : 1.0000
                                     NA's     :235.0000

> str(heart) #show fields and encoding
'data.frame': 5388 obs. of 16 variables:
 $ death : int 0 1 0 1 0 0 0 0 0 0 ...
 $ anterior: Factor w/ 2 levels "Inferior","Anterior": 2 1 1 2 2 2 1 NA NA 1
 ...
 $ hcabg : int 0 0 0 0 0 0 0 0 0 0 ...
 $ kk1 : int 1 0 1 0 0 NA 1 0 1 1 ...
 $ kk2 : int 0 1 0 1 1 NA 0 0 0 0 ...
 $ kk3 : int 0 0 0 0 0 NA 0 1 0 0 ...
 $ kk4 : int 0 0 0 0 0 NA 0 0 0 0 ...
 $ age1 : int 1 0 0 0 0 1 1 1 1 0 ...
 $ age2 : int 0 0 0 0 1 0 0 0 0 1 ...
 $ age3 : int 0 0 1 1 0 0 0 0 0 0 ...
 $ age4 : int 0 1 0 0 0 0 0 0 0 0 ...
 $ center : Factor w/ 47 levels "1255","1287",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ killip : Factor w/ 4 levels "1","2","3","4": 1 2 1 2 2 NA 1 3 1 1 ...
 $ agegrp : Factor w/ 4 levels "<=60","60-69",...: 1 4 3 3 2 1 1 1 1 2 ...
 $ age : int 55 98 76 79 66 40 48 57 40 70 ...
 $ kk23 : int 0 1 0 1 1 NA 0 1 0 0 ...
 - attr(*, "datalabel")= chr ""
 - attr(*, "time.stamp")= chr "25 Oct 2007 19:59"
 - attr(*, "formats")= chr "%4.0g" "%9.0g" "%4.0g" "%8.0g" ...
 - attr(*, "types")= int 251 251 251 251 251 251 251 251 251 251 ...
 - attr(*, "val.labels")= chr "" "anterior" "" "" ...
 - attr(*, "var.labels")= chr "Death within 48 hrs onset" "1=anterior;
0=inferior" "Hist of 1=CABG;0=PTCA" "killip==" 1.0000" ...
 - attr(*, "version")= int 10
 - attr(*, "label.table")=List of 2
 ..$ age : Named num 1 2 3 4
 ..$ ..- attr(*, "names")= chr "<=60" "60-69" "70-79" ">=80"
 ..$ anterior: Named num 0 1 2 3 4 5 6 7
 ..$ ..- attr(*, "names")= chr "Inferior" "Anterior" "Inferior" "LBBB" ...

```

```

>
> #ulogit function to show univariate fit
> ulogit <- function (y,x) {
+   f<- glm(y ~ x, family=binomial())
+   if (is.factor(x) && length(levels(x))>2) {
+     print(anova(f, test='Chisq'))
+   } else {
+     b<- coef(f)[2]
+     ci<- confint(f)[2,]
+     cat('NullDeviance Deviance          OR          P-value          LCL          UCL\n')
+
+     cat(format(f$null.deviance,nsmall=2,width=10),format(f$deviance,nsmall=2,width=
+ 10),
+
+     format(exp(b),nsmall=4,width=10),format(pchisq(b*b/vcov(f)[2,2],1,lower.tail=FA
+ LSE),
+
+     format(exp(ci[1]),nsmall=4,width=10),format(exp(ci[2]),
+     format(exp(ci[2]),nsmall=4,width=10),'\n')
+   }
+ }
>
> ulogit(death,anterior)
Waiting for profiling to be done...
NullDeviance Deviance          OR          P-value          LCL          UCL
1571.988 1543.845 2.236796 2.222798e-07 1.655254 3.047431
> ulogit(death,hcabg)
Waiting for profiling to be done...
NullDeviance Deviance          OR          P-value          LCL          UCL
1974.812 1972.102 1.682106 0.07863687 0.8979673 2.889201
> ulogit(death,age)
Waiting for profiling to be done...
NullDeviance Deviance          OR          P-value          LCL          UCL
1974.812 1826.571 1.058157 8.956398e-32 1.048293 1.068295
>
> #helper function to summarize logistic fit
> showLogistic<- function (fit) {
+   s<-summary(fit)
+   print(s)
+   b<- coef(fit) #coefficients
+   o<- exp(b) #odds ratios
+   CI<- confint(fit) #confidence intervals for coeff
+   oCI<- exp(CI) #confidence intervals for odds ratios
+   cat('\n')
+   cat('Variable          Coefficient          LCL          UCL\n')
+   for (i in 1:length(b)) {
+     cat(format(names(o)[i],width=15),formatC(b[i],digits=5,width=10),
+
+     format(CI[i,1],digits=5,width=10),formatC(CI[i,2],digits=5,width=10),'\n')
+   }
+   cat('\n')
+   cat('Variable          Odds Ratio          LCL          UCL\n')
+   for (i in 1:length(o)) {
+     cat(format(names(o)[i],width=15),formatC(o[i],digits=5, width=10),
+     formatC(oCI[i,1],digits=5,width=10),formatC(oCI[i,2],digits=5,
+ width=10),'\n')
+   }
+   cat('\n')
+ }
>
> #killip factor fits:
> fitk<- glm(death ~ kk2+kk3+kk4, data=heart, family=binomial())
> showLogistic(fitk)

```

```

Call:
glm(formula = death ~ kk2 + kk3 + kk4, family = binomial(), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0469  -0.2284  -0.2284  -0.2284   2.7054

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6335     0.1029 -35.323  < 2e-16 ***
kk2           1.0313     0.1603   6.432 1.26e-10 ***
kk3           1.4664     0.2185   6.712 1.92e-11 ***
kk4           3.3184     0.2732  12.146  < 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: 1849.1  on 5152  degrees of freedom
Residual deviance: 1698.6  on 5149  degrees of freedom
(235 observations deleted due to missingness)
AIC: 1706.6

Number of Fisher Scoring iterations: 6

Waiting for profiling to be done...

Variable      Coefficient      LCL      UCL
(Intercept)   -3.6335      -3.8418   -3.4381
kk2            1.0313       0.71423   1.3439
kk3            1.4664       1.0233     1.8827
kk4            3.3184       2.7761     3.8514

Variable      Odds Ratio      LCL      UCL
(Intercept)   0.026423    0.021455  0.032125
kk2            2.8048       2.0426     3.8339
kk3            4.3334       2.7823     6.5711
kk4           27.617       16.057    47.059

> fitk2<- glm(death ~ killip, data=heart, family=binomial()) #note how dummy
vars are created
> summary(fitk2) #same fit as above

Call:
glm(formula = death ~ killip, family = binomial(), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0469  -0.2284  -0.2284  -0.2284   2.7054

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6335     0.1029 -35.323  < 2e-16 ***
killip2       1.0313     0.1603   6.432 1.26e-10 ***
killip3       1.4664     0.2185   6.712 1.92e-11 ***
killip4       3.3184     0.2732  12.146  < 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: 1849.1  on 5152  degrees of freedom

```

Residual deviance: 1698.6 on 5149 degrees of freedom
 (235 observations deleted due to missingness)
 AIC: 1706.6

Number of Fisher Scoring iterations: 6

```
>
> tk<-table(factor(killip, exclude=NULL)) #include NA's
> tk #show

  1    2    3    4 <NA>
3768 1029 292   64  235
> cat(format(tk/sum(tk), digits=3),'\n') #show fractions in each level
0.6993 0.1910 0.0542 0.0119 0.0436
> sum(tk)
[1] 5388
>
> #agegrp factor fits:
> fitage<- glm(death ~ age2+age3+age4, data=heart, family=binomial())
> showLogistic(fitage)
```

Call:
 glm(formula = death ~ age2 + age3 + age4, family = binomial(),
 data = heart)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.5291	-0.3715	-0.2291	-0.1601	2.9540

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.3503	0.2054	-21.176	< 2e-16 ***
age2	0.7230	0.2659	2.719	0.00655 **
age3	1.7112	0.2313	7.399	1.37e-13 ***
age4	2.4549	0.2354	10.427	< 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: 1972.7 on 5364 degrees of freedom
 Residual deviance: 1805.5 on 5361 degrees of freedom
 (23 observations deleted due to missingness)
 AIC: 1813.5

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable	Coefficient	LCL	UCL
(Intercept)	-4.3503	-4.7812	-3.9723
age2	0.72298	0.20771	1.2559
age3	1.7112	1.2755	2.1862
age4	2.4549	2.0101	2.9371

Variable	Odds Ratio	LCL	UCL
(Intercept)	0.012903	0.0083858	0.01883
age2	2.0606	1.2309	3.511
age3	5.5357	3.5803	8.9012
age4	11.645	7.4638	18.861

```
> fitage2<- glm(death ~ agegrp, data=heart, family=binomial()) #dummy vars
auto-created
```

```

> summary(fitage2) #slight differences

Call:
glm(formula = death ~ agegrp, family = binomial(), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5291 -0.3715 -0.2291 -0.1592  2.9581

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.3626     0.2054 -21.237  < 2e-16 ***
agegrp60-69   0.7353     0.2659   2.765  0.00569 **
agegrp70-79   1.7235     0.2312   7.453  9.12e-14 ***
agegrp>=80    2.4672     0.2354  10.480  < 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: 1974.8  on 5387  degrees of freedom
Residual deviance: 1806.1  on 5384  degrees of freedom
AIC: 1814.1

Number of Fisher Scoring iterations: 7

>
> fitage3<- glm(death ~ age1+age2+age3, data=heart, family=binomial())
> showLogistic(fitage3)

Call:
glm(formula = death ~ age1 + age2 + age3, family = binomial(),
    data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5291 -0.3715 -0.2291 -0.1601  2.9540

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.8954     0.1150 -16.484  < 2e-16 ***
age1         -2.4549     0.2354 -10.427  < 2e-16 ***
age2         -1.7319     0.2043  -8.477  < 2e-16 ***
age3         -0.7437     0.1565  -4.751  2.02e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1972.7  on 5364  degrees of freedom
Residual deviance: 1805.5  on 5361  degrees of freedom
(23 observations deleted due to missingness)
AIC: 1813.5

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable      Coefficient      LCL      UCL
(Intercept)   -1.8954      -2.1273    -1.676
age1          -2.4549      -2.9371    -2.0101
age2          -1.7319      -2.1435    -1.3402
age3          -0.74366     -1.0504    -0.43594

```

```

Variable      Odds Ratio      LCL      UCL
(Intercept)    0.15026      0.11916      0.18712
age1           0.085873    0.053019    0.13398
age2           0.17695      0.11725      0.2618
age3           0.47537      0.3498      0.64665

>
> #full model
> fitFull<- glm(death ~ anterior+hcabg+age+kk2+kk3+kk4, data=heart,
family=binomial())
> showLogistic(fitFull)

Call:
glm(formula = death ~ anterior + hcabg + age + kk2 + kk3 + kk4,
     family = binomial(), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5019  -0.2892  -0.2078  -0.1467   3.3356

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.625430   0.453186  -16.826 < 2e-16 ***
anteriorAnterior  0.649676   0.167181   3.886 0.000102 ***
hcabg          0.696126   0.349818   1.990 0.046595 *
age            0.051655   0.005784   8.931 < 2e-16 ***
kk2            0.824753   0.179973   4.583 4.59e-06 ***
kk3            0.831972   0.268521   3.098 0.001946 **
kk4            2.701840   0.352744   7.659 1.87e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1486.2 on 4502 degrees of freedom
Residual deviance: 1287.6 on 4496 degrees of freedom
(885 observations deleted due to missingness)
AIC: 1301.6

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable      Coefficient      LCL      UCL
(Intercept)   -7.6254      -8.534      -6.7563
anteriorAnterior  0.64968    0.32536    0.98189
hcabg          0.69613   -0.049294    1.3358
age            0.051655   0.040423    0.06311
kk2            0.82475    0.46816     1.175
kk3            0.83197    0.2817     1.3392
kk4            2.7018     1.9920     3.3819

Variable      Odds Ratio      LCL      UCL
(Intercept)    0.00048789 0.00019667 0.0011635
anteriorAnterior  1.9149      1.3845      2.6695
hcabg           2.006      0.9519      3.803
age             1.053      1.0413      1.0651
kk2             2.2813      1.5971      3.2383
kk3             2.2978      1.3254      3.8159
kk4            14.907      7.3299     29.426

>

```

```
> #Full model with categorical age
> fitFull2<- glm(death ~ anterior+hcabg+kk2+kk3+kk4+age2+age3+age4, data=heart,
family=binomial())
> showLogistic(fitFull2)
```

Call:

```
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age2 +
    age3 + age4, family = binomial(), data = heart)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-1.3670  -0.3276  -0.1986  -0.1443   3.1807
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.0521	0.2586	-19.536	< 2e-16 ***
anteriorAnterior	0.6426	0.1676	3.835	0.000126 ***
hcabg	0.7444	0.3530	2.109	0.034929 *
kk2	0.8117	0.1805	4.497	6.90e-06 ***
kk3	0.7757	0.2691	2.883	0.003939 **
kk4	2.6597	0.3560	7.471	7.96e-14 ***
age2	0.4930	0.3102	1.589	0.111961
age3	1.5112	0.2662	5.676	1.38e-08 ***
age4	2.1853	0.2718	8.039	9.06e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1484.6 on 4482 degrees of freedom
Residual deviance: 1273.3 on 4474 degrees of freedom
(905 observations deleted due to missingness)
AIC: 1291.3

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable	Coefficient	LCL	UCL
(Intercept)	-5.0521	-5.5909	-4.5731
anteriorAnterior	0.64256	0.31745	0.97546
hcabg	0.74445	-0.0058172	1.3916
kk2	0.8117	0.45407	1.163
kk3	0.7757	0.22456	1.2842
kk4	2.6597	1.9445	3.3472
age2	0.49302	-0.11156	1.113
age3	1.5112	1.0096	2.0591
age4	2.1853	1.6715	2.743

Variable	Odds Ratio	LCL	UCL
(Intercept)	0.0063961	0.0037317	0.010326
anteriorAnterior	1.9013	1.3736	2.6524
hcabg	2.1053	0.9942	4.0211
kk2	2.2517	1.5747	3.1997
kk3	2.1721	1.2518	3.6116
kk4	14.291	6.9904	28.422
age2	1.6373	0.89444	3.0435
age3	4.532	2.7445	7.8391
age4	8.8932	5.3199	15.533

```
>
> #drop age2 (combine with age1)
```

```
> fitRed<- glm(death ~ anterior+hcabg+kk2+kk3+kk4+age3+age4, data=heart,
family=binomial())
> showLogistic(fitRed)
```

Call:

```
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
age4, family = binomial(), data = heart)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-1.3716  -0.3258  -0.1751  -0.1275   3.1039
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.8091	0.1935	-24.853	< 2e-16 ***
anteriorAnterior	0.6386	0.1675	3.811	0.000138 ***
hcabg	0.7867	0.3528	2.230	0.025747 *
kk2	0.8231	0.1804	4.562	5.07e-06 ***
kk3	0.7949	0.2692	2.953	0.003147 **
kk4	2.6811	0.3563	7.524	5.29e-14 ***
age3	1.2610	0.2006	6.287	3.23e-10 ***
age4	1.9351	0.2080	9.303	< 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: 1484.6 on 4482 degrees of freedom
Residual deviance: 1275.8 on 4475 degrees of freedom
(905 observations deleted due to missingness)
AIC: 1291.8

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable	Coefficient	LCL	UCL
(Intercept)	-4.8091	-5.2033	-4.4439
anteriorAnterior	0.63856	0.31348	0.97143
hcabg	0.78666	0.036613	1.433
kk2	0.82314	0.46563	1.1743
kk3	0.79489	0.24359	1.3035
kk4	2.6811	1.9653	3.369
age3	1.261	0.87146	1.6598
age4	1.9351	1.5302	2.3478

Variable	Odds Ratio	LCL	UCL
(Intercept)	0.0081555	0.0054985	0.01175
anteriorAnterior	1.8938	1.3682	2.6417
hcabg	2.1961	1.0373	4.1914
kk2	2.2776	1.593	3.236
kk3	2.2142	1.2758	3.6823
kk4	14.601	7.1368	29.05
age3	3.5289	2.3904	5.2584
age4	6.9251	4.6191	10.462

>

> #likelihood ratio test (deviance difference)

> anova(fitRed,fitFull2,test='Chisq')

Analysis of Deviance Table

Model 1: death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4

Model 2: death ~ anterior + hcabg + kk2 + kk3 + kk4 + age2 + age3 + age4

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	4475	1275.81			
2	4474	1273.25	1	2.55	0.11

SECTION 5.2:

```
#Section 5.2:

#Box-Tidwell
summary(age) #show some descriptive stats for age
sd(age)
length(age)

fitage<- glm(death ~ age, family=binomial(link=logit))
summary(fitage)

lnageage<- age*log(age) #Box-Tidwell interaction variable
fitlnageage<- glm(death ~ age + lnageage, family=binomial)
summary(fitlnageage)

#Tukey-Pregibon
hatage<- hatvalues(fitage) #get hat values for death ~ age fit
summary(glm(death ~ age + hatage + I(hatage^2), family=binomial))

#Partial Residuals
lpartr<- function (glmModel, iVar)
{
  xlab <- attributes(glmModel$terms)$term.labels[iVar] #construct x-axis label
  for var
  x<- model.matrix(glmModel)[, iVar + 1] #get values for var
  y<- coef(glmModel)[iVar + 1]*x + residuals(glmModel, type='pearson') #partial
  residual
  plot(x, y, col='blue', xlab = xlab, ylab = 'Partial Residual')
  lines(lowess(x,y), col='red')
  invisible()
}
fitPartial<- glm(death ~ age + anterior + hcabg + kk2 + kk3 + kk4,
family=binomial)
lpartr(fitPartial, 1)

#Linearity of slopes test

#construct dummy 'agegrp' factor as 'agegrpNew' to avoid changing 'agegrp' in
heart01
agegrpNew<- rep(1,length(age)) #start with all level 1 (age <= 60)
agegrpNew<- ifelse(age>60 & age<=70, 2, agegrpNew)
agegrpNew<- ifelse(age>70 & age<=80, 3, agegrpNew)
agegrpNew<- ifelse(age>80, 4, agegrpNew)
agegrpNew<- factor(agegrpNew, labels=c('<=60', '61-70', '71-80', '>80'))

tagegrp<-table(agegrp) #frequency table for agegrp
tagegrp #show
cat(format(tagegrp/sum(tagegrp), digits=3),'\n') #show fractions in each level
sum(tagegrp) #total number

age23<- as.numeric(age2==1 | age3==1) #middle ages (numeric, not factor)
tage23<-table(age23) #frequency table for agegrp
tage23 #show
```

```

cat(format(tage23/sum(tage23), digits=3),'\n') #show fractions in each level
sum(tage23) #total number

fit1<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age23 + age4,
           family=binomial)
summary(fit1)

fit2<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age2 +
           age3 + age4, family=binomial)
summary(fit2)

fit3<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
           age3 + age4, family=binomial)
summary(fit3)
2.1958 - 1.5218
1.9409 - 1.2668

#GAM
#NOTE: 'heart01' dataset already loaded as 'heart'.
library('mgcv') #get package for GAM functions

heartgam<- gam(death ~ s(age) + kk2 + kk3 + kk4, family=binomial)
summary(heartgam)
plot(heartgam, residuals=TRUE, se=TRUE, col='blue', pch='.')

heartgam2<- gam(death ~ s(age) + anterior + hcabg +
               kk2 + kk3 + kk4, family=binomial)
summary(heartgam2)
plot(heartgam2, col='blue')

#Fractional polynomials for GLM not available yet in a R package!
#But see: Amber G. (1999) Fractional polynomials in S-plus,
# http://lib.stat.cmu.edu/S/fracpoly.

```

OUTPUT: 5.2

```
> #Box-Tidwell
> summary(age) #show some descriptive stats for age
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  40.00  54.00   66.00   65.81  76.00  100.00
> sd(age)
[1] 14.42962
> length(age)
[1] 5388

> fitage<- glm(death ~ age, family=binomial(link=logit))
> summary(fitage)

Call:
glm(formula = death ~ age, family = binomial(link = logit))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6569 -0.3375 -0.2559 -0.1780  3.1060

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.076784    0.372890  -18.98  <2e-16 ***
age          0.056529    0.004819   11.73  <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: 1974.8  on 5387  degrees of freedom
Residual deviance: 1826.6  on 5386  degrees of freedom
AIC: 1830.6

Number of Fisher Scoring iterations: 6

> lnageage<- age*log(age) #Box-Tidwell interaction variable
> fitlnageage<- glm(death ~ age + lnageage, family=binomial)
> summary(fitlnageage)

Call:
glm(formula = death ~ age + lnageage, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6268 -0.3450 -0.2594 -0.1712  3.1940

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.62618    3.03522  -3.171  0.00152 **
age          0.24332    0.21997   1.106  0.26866
lnageage     -0.03525    0.04148  -0.850  0.39541
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1974.8  on 5387  degrees of freedom
Residual deviance: 1825.8  on 5385  degrees of freedom
AIC: 1831.8
```

Number of Fisher Scoring iterations: 7

```
> #Tukey-Pregibon
> hatage<- hatvalues(fitage) #get hat values for death ~ age fit
> summary(glm(death ~ age + hatage + I(hatage^2), family=binomial))
```

Call:

```
glm(formula = death ~ age + hatage + I(hatage^2), family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.5530	-0.3527	-0.2542	-0.1640	3.1979

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.727e+00	5.365e-01	-14.403	< 2e-16 ***
age	6.716e-02	8.426e-03	7.971	1.58e-15 ***
hatage	-2.346e+02	5.333e+02	-0.440	0.66
I(hatage^2)	-2.594e+04	1.888e+05	-0.137	0.89

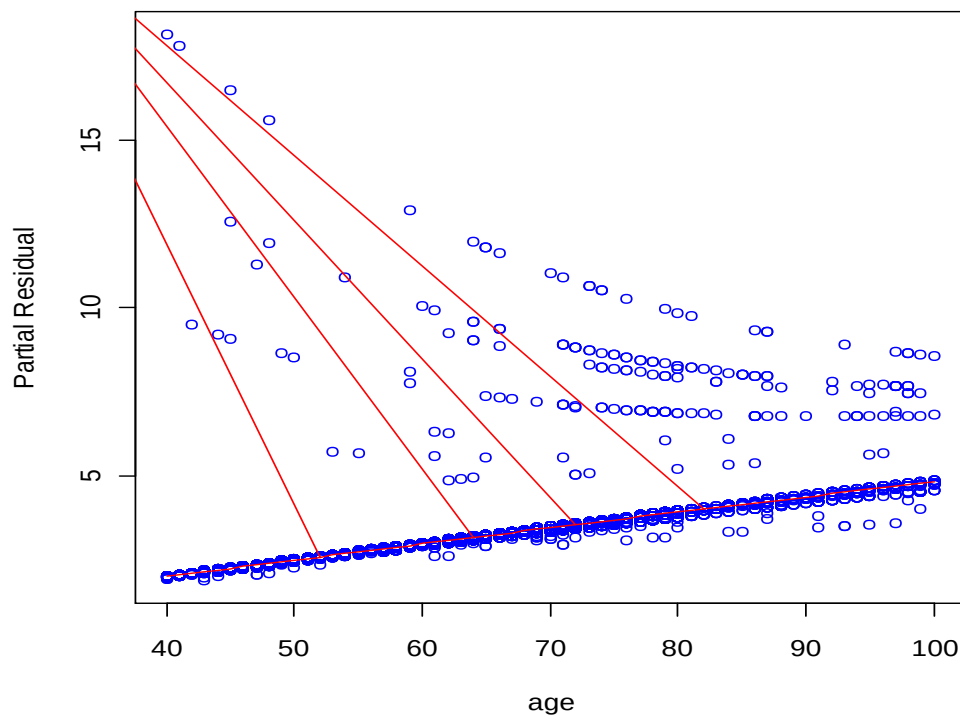
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1974.8 on 5387 degrees of freedom
Residual deviance: 1822.7 on 5384 degrees of freedom
AIC: 1830.7

Number of Fisher Scoring iterations: 7

```
> #Partial Residuals
> lpartr<- function (glmModel, iVar)
+ {
+   xlab <- attributes(glmModel$terms)$term.labels[iVar] #construct x-axis
+   label for var
+   x<- model.matrix(glmModel)[, iVar + 1] #get values for var
+   y<- coef(glmModel)[iVar + 1]*x + residuals(glmModel, type='pearson')
+   #partial residual
+   plot(x, y, col='blue', xlab = xlab, ylab = 'Partial Residual')
+   lines(lowess(x,y), col='red')
+   invisible()
+ }
> fitPartial<- glm(death ~ age + anterior + hcabg + kk2 + kk3 + kk4,
family=binomial)
> lpartr(fitPartial, 1)
```



```
> #Linearity of slopes test

> #construct dummy 'agegrp' factor as 'agegrpNew' to avoid changing 'agegrp' in
heart01
> agegrpNew<- rep(1,length(age)) #start with all level 1 (age <= 60)
> agegrpNew<- ifelse(age>60 & age<=70, 2, agegrpNew)
> agegrpNew<- ifelse(age>70 & age<=80, 3, agegrpNew)
> agegrpNew<- ifelse(age>80, 4, agegrpNew)
> agegrpNew<- factor(agegrpNew, labels=c('<=60', '61-70', '71-80', '>80'))
> tagegrp<-table(agegrp) #frequency table for agegrp
> tagegrp #show
agegrp
=<60 61-70 71-80 >80
 1907  1390  1425   666
> cat(format(tagegrp/sum(tagegrp), digits=3),'\n') #show fractions in each
level
0.354 0.258 0.264 0.124
> sum(tagegrp) #total number
[1] 5388

> age23<- as.numeric(age2==1 | age3==1) #middle ages (numeric, not factor)
> tage23 #show
age23
  0    1
2573 2815
> cat(format(tage23/sum(tage23), digits=3),'\n') #show fractions in each level
0.478 0.522
> sum(tage23) #total number
[1] 5388
```

```
> fit1<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age23 + age4,
+           family=binomial)
> summary(fit1)
```

Call:

```
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age23 +
    age4, family = binomial)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.3605	-0.2722	-0.1955	-0.1552	3.1886

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-5.0772	0.2588	-19.617	< 2e-16	***
anteriorAnterior	0.6635	0.1670	3.973	7.09e-05	***
hcabg	0.6706	0.3507	1.912	0.05588	.
kk2	0.8209	0.1798	4.566	4.98e-06	***
kk3	0.8243	0.2690	3.064	0.00218	**
kk4	2.6447	0.3498	7.560	4.04e-14	***
age23	1.1297	0.2569	4.397	1.10e-05	***
age4	2.1898	0.2719	8.054	8.03e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1486.2 on 4502 degrees of freedom
Residual deviance: 1293.4 on 4495 degrees of freedom
(885 observations deleted due to missingness)
AIC: 1309.4

Number of Fisher Scoring iterations: 7

```
> fit2<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age2 +
+           age3 + age4, family=binomial)
> summary(fit2)
```

Call:

```
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age2 +
    age3 + age4, family = binomial)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.3675	-0.3275	-0.1985	-0.1443	3.1842

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-5.0634	0.2585	-19.588	< 2e-16	***
anteriorAnterior	0.6428	0.1676	3.836	0.000125	***
hcabg	0.7436	0.3529	2.107	0.035131	*
kk2	0.8129	0.1805	4.503	6.69e-06	***
kk3	0.7768	0.2691	2.887	0.003893	**
kk4	2.6614	0.3561	7.474	7.80e-14	***
age2	0.5037	0.3102	1.624	0.104405	
age3	1.5218	0.2662	5.716	1.09e-08	***
age4	2.1958	0.2719	8.077	6.62e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1486.2 on 4502 degrees of freedom

```
Residual deviance: 1273.7 on 4494 degrees of freedom
(885 observations deleted due to missingness)
AIC: 1291.7
```

```
Number of Fisher Scoring iterations: 7
```

```
> fit3<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
+           age3 + age4, family=binomial)
> summary(fit3)
```

```
Call:
```

```
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial)
```

```
Deviance Residuals:
```

```
      Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061
```

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.8159	0.1934	-24.896	< 2e-16 ***
anteriorAnterior	0.6387	0.1675	3.812	0.000138 ***
hcabg	0.7864	0.3527	2.229	0.025786 *
kk2	0.8249	0.1804	4.572	4.84e-06 ***
kk3	0.7967	0.2692	2.959	0.003083 **
kk4	2.6837	0.3565	7.529	5.11e-14 ***
age3	1.2668	0.2006	6.316	2.68e-10 ***
age4	1.9409	0.2080	9.329	< 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: 1486.2 on 4502 degrees of freedom
Residual deviance: 1276.3 on 4495 degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3
```

```
Number of Fisher Scoring iterations: 7
```

```
> 2.1958 - 1.5218
[1] 0.674
> 1.9409 - 1.2668
[1] 0.6741
```

```
> #GAM
```

```
> library('mgcv') #get package for GAM functions
```

```
This is mgcv 1.4-0
```

```
> heartgam<- gam(death ~ s(age) + kk2 + kk3 + kk4, family=binomial)
```

```
> summary(heartgam)
```

```
Family: binomial
```

```
Link function: logit
```

```
Formula:
```

```
death ~ s(age) + kk2 + kk3 + kk4
```

```
Parametric coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.8072	0.1121	-33.948	< 2e-16 ***
kk2	0.8744	0.1630	5.364	8.15e-08 ***
kk3	1.0988	0.2244	4.897	9.74e-07 ***
kk4	3.0820	0.2873	10.726	< 2e-16 ***

```

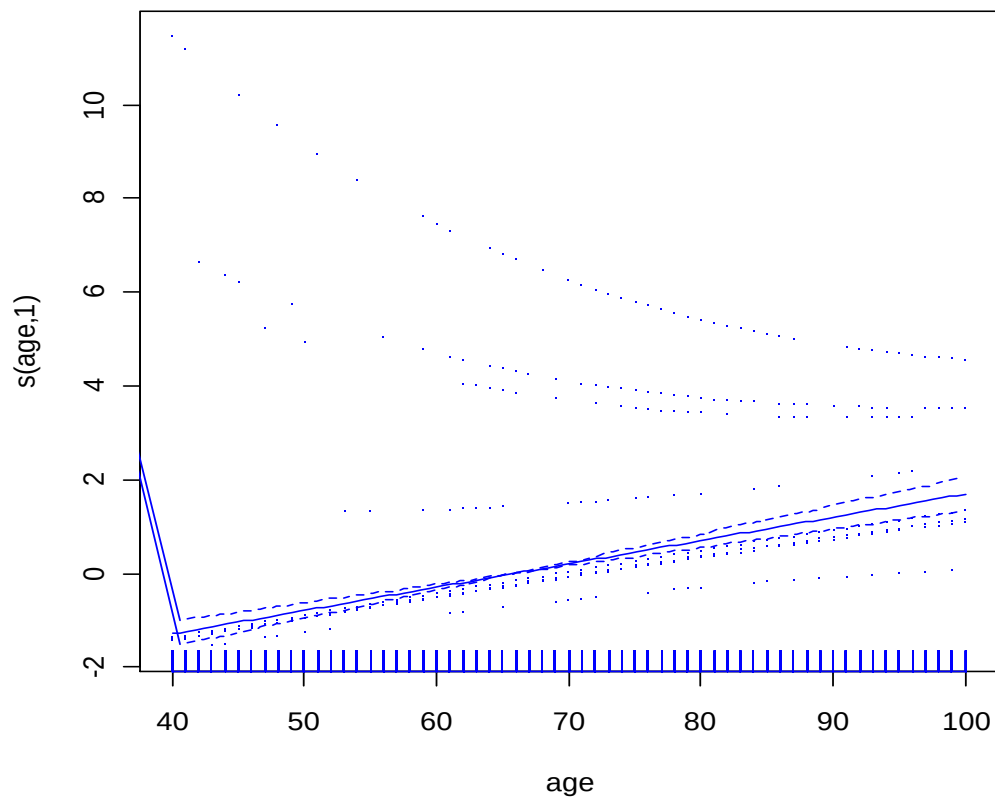
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df Chi.sq p-value
s(age)   1     1.5  91.86 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0687   Deviance explained = 13.4%
UBRE score = -0.68742  Scale est. = 1          n = 5153

> plot(heartgam, residuals=TRUE, se=TRUE, col='blue', pch='.')

```



```

> heartgam2<- gam(death ~ s(age) + anterior + hcabg +
+   kk2 + kk3 + kk4, family=binomial)
> summary(heartgam2)

Family: binomial
Link function: logit

Formula:
death ~ s(age) + anterior + hcabg + kk2 + kk3 + kk4

Parametric coefficients:
      Estimate Std. Error z value Pr(>|z|)

```



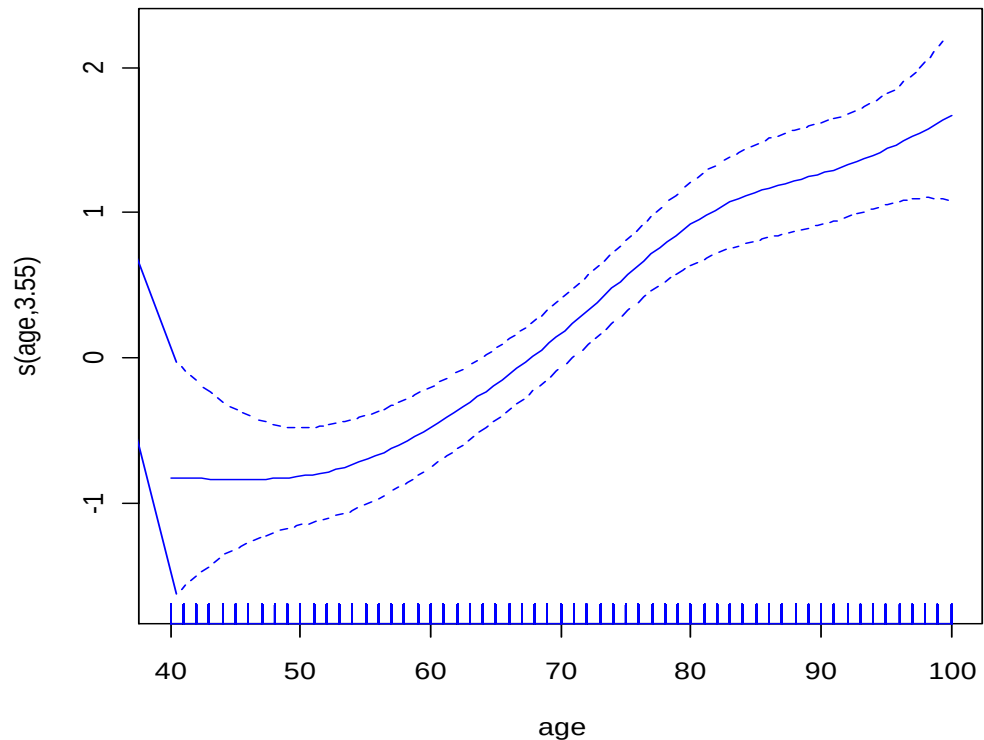
```

(Intercept)      -4.2431      0.1593 -26.630 < 2e-16 ***
anteriorAnterior  0.6503      0.1673   3.886 0.000102 ***
hcabg            0.7203      0.3513   2.051 0.040310 *
kk2              0.8195      0.1803   4.546 5.47e-06 ***
kk3              0.8228      0.2686   3.063 0.002191 **
kk4              2.7089      0.3540   7.651 1.99e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df Chi.sq p-value
s(age) 3.553  4.053  82.47 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0563   Deviance explained = 13.8%
UBRE score = -0.71114   Scale est. = 1         n = 4503
> plot(heartgam2, col='blue')

```



```

> #Fractional polynomials for GLM not available yet in a R package!
> #But see: Amber G. (1999) Fractional polynomials in S-plus,
> # http://lib.stat.cmu.edu/S/fracpoly.

```

SECTION 5.3

```

#SECTION 5.3: STANDARDIZED COEFFICIENTS

sqrt((pi*pi)/3)
sd(as.numeric(anterior), na.rm=TRUE) #std. dev. dropping missing data
.6387*.4977929/1.8137994

lstand<- function (lfit) {
#compute standardized coefficients per Hilber
  f<- sqrt(pi*pi/3) #factor for logistic
  b<- coef(lfit)
  o<- exp(b)
  v<- as.character(attributes(lfit$terms)$variables) #get names of vars
  cat('Seq. Variable   Coef      OR      Std. Coef.','\n')
  cat(0, format(v[2], width=10), format(b[1], width=8, digits=4),
      format(o[1], width=8, digits=4), '\n')
  for (i in 1:(length(lfit$model)-1)) {
    x<- lfit$model[i+1][,1] #get values of i-th variable
    if (is.numeric(x)) {
      s<- sd(x, na.rm=TRUE) #get std. dev.
    } else { #factor
      s<- sd(as.integer(x), na.rm=TRUE) #std. dev. for factor
    }
    bs<- b[i+1]*s/f #std. coef.
    cat(i, format(v[i+2], width=10), format(b[i+1], width=8, digits=4),
        format(o[i+1], width=8, digits=4), format(bs, width=8, digits=4), '\n')
  }
  invisible() #suppress return print
}

lstand(fit3)

```

OUTPUT: 5.3

```

> #SECTION 5.3: STANDARDIZED COEFFICIENTS
> sqrt((pi*pi)/3)
[1] 1.813799

> sd(as.numeric(anterior), na.rm=TRUE) #std. dev. dropping missing data
[1] 0.4977929

> .6387*.4977929/1.8137994
[1] 0.1752897

> lstand<- function (lfit) {
+ #compute standardized coefficients per Hilber
+   f<- sqrt(pi*pi/3) #factor for logistic
+   b<- coef(lfit)
+   o<- exp(b)
+   v<- as.character(attributes(lfit$terms)$variables) #get names of vars
+   cat('Seq. Variable   Coef      OR      Std. Coef.','\n')
+   cat(0, format(v[2], width=10), format(b[1], width=8, digits=4),
+       format(o[1], width=8, digits=4), '\n')
+   for (i in 1:(length(lfit$model)-1)) {
+     x<- lfit$model[i+1][,1] #get values of i-th variable
+     if (is.numeric(x)) {
+       s<- sd(x, na.rm=TRUE) #get std. dev.
+     } else { #factor
+       s<- sd(as.integer(x), na.rm=TRUE) #std. dev. for factor
+     }
+   }
+ }

```

```

+     bs<- b[i+1]*s/f #std. coef.
+     cat(i, format(v[i+2], width=10), format(b[i+1], width=8, digits=4),
+       format(o[i+1], width=8, digits=4), format(bs, width=8, digits=4), '\n')
+   }
+   invisible() #suppress return print
+ }
> lstand(fit3)
Seq. Variable   Coef      OR      Std. Coef.
0 death        -4.816    0.0081
1 anterior      0.6387    1.894    0.1753
2 hcabg         0.7864    2.196    0.07552
3 kk2           0.8249    2.282    0.1811
4 kk3           0.7967    2.218    0.09691
5 kk4           2.684     14.64    0.1520
6 age3          1.267     3.550    0.3053
7 age4          1.941     6.965    0.3479

```

[NOTE: Differences in standardized coefficients from Stata are due to the use here of the standard deviation of the actual variable values used in the fit, which omits rows dropped because of missing data. - RAL]

SECTION 5.4

```

#SECTION 5.4: STANDARD ERRORS
fit4<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4, family=binomial)
summary(fit4)
v<- vcov(fit4) #variance-covariance matrix
v
sqrt(v[2,2]) #anterior
sqrt(0.027113125) #and again
sqrt(v[3,3]) #hcabg
sqrt(v[4,4]) #kk2

0.7227 / 0.1646606 #z-value for anterior
0.7185 / 0.3460558 #z-value for hcabg

se<- sqrt(diag(v)) #S.E.s of coefficients in fit4 model
se
coef(fit4)[2]/se[2] #could also use
coef(fit4)['anteriorAnterior']/se['anteriorAnterior']
coef(fit4)[3]/se[3] #also coef(fit4)['hcabg']/se['hcabg']

coef(fit4)[2]/se[2] #anterior value in table is 4.389
2*pnorm(4.38874, lower.tail=FALSE) #complementary normal probability x 2
coef(fit4)[3]/se[3] #hcabg z value in table is 2.076
2*pnorm(2.076255, lower.tail=FALSE) #complementary normal probability x 2

qnorm(0.05/2) #normal quantile for p = 0.025
qnorm(1-0.05/2) #normal quantile for p= 0.975

matrix(c(coef(fit4)-1.96*se, coef(fit4)+ 1.96*se),ncol=2) #Wald 95% C.I.s for
coefficients
confint(fit4) #more accurate 95% C.I.s
0.7227 - 1.96*0.1646606 #95% Wald LCL for anterior
0.7227 + 1.96*0.1646606 #95% Wald UCL for anterior
0.7185 - 1.96*0.3460558 #95% Wald LCL for hcabg
0.7185 + 1.96*0.3460558 #95% Wald UCL for hcabg

exp(coef(fit4)) #odds ratios for coefficients

```

```
exp(matrix(c(coef(fit4)-1.96*se, coef(fit4)+ 1.96*se),ncol=2)) #Wald 95% C.I.s
for odds ratios
exp(confint(fit4)) #more accurate 95% C.I.s for odds ratios
exp(0.7227 - 1.96*0.1646606) #95% Wald LCL for anterior odds ratio
exp(0.7227 + 1.96*0.1646606) #95% Wald UCL for anterior odds ratio
```

OUTPUT: 5.4

```
> #SECTION 5.4: STANDARD ERRORS
> fit4<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4, family=binomial)
> summary(fit4)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0015  -0.2627  -0.2627  -0.1838   2.8598

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.0722     0.1495 -27.244 < 2e-16 ***
anteriorAnterior  0.7227     0.1647   4.389 1.14e-05 ***
hcabg          0.7185     0.3461   2.076  0.0379 *
kk2            0.9453     0.1771   5.337 9.42e-08 ***
kk3            1.1769     0.2621   4.490 7.14e-06 ***
kk4            2.9206     0.3300   8.849 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1372.6  on 4497  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1384.6

Number of Fisher Scoring iterations: 6

> v<- vcov(fit4) #variance-covariance matrix
> v
              (Intercept) anteriorAnterior          hcabg          kk2
(Intercept)   0.022341470      -0.016046780 -0.007830591 -0.011149164
anteriorAnterior -0.016046780      0.027113125  0.003167079 -0.002213358
hcabg          -0.007830591      0.003167079  0.119754634 -0.002021795
kk2            -0.011149164      -0.002213358 -0.002021795  0.031369347
kk3            -0.009678818      -0.004682757 -0.002487497  0.012958033
kk4            -0.011836924      -0.001470963  0.003043371  0.012619693
              kk3          kk4
(Intercept)   -0.009678818 -0.011836924
anteriorAnterior -0.004682757 -0.001470963
hcabg          -0.002487497  0.003043371
kk2            0.012958033  0.012619693
kk3            0.068717491  0.012748892
kk4            0.012748892  0.108923216
> sqrt(v[2,2]) #anterior
[1] 0.1646606
> sqrt(0.027113125) #and again
```

```

[1] 0.1646606
>> sqrt(v[3,3]) #hcabg
[1] 0.3460558
> sqrt(v[4,4]) #kk2
[1] 0.1771139

> 0.7227 / 0.1646606 #z-value for anterior
[1] 4.389028
> 0.7185 / 0.3460558 #z-value for hcabg
[1] 2.076255

> se<- sqrt(diag(v)) #S.E.s of coefficients in fit4 model
> se
      (Intercept) anteriorAnterior      hcabg      kk2
      0.1494706      0.1646606      0.3460558      0.1771139
      kk3      kk4
      0.2621402      0.3300352

> coef(fit4)[2]/se[2] #could also use coef(fit4)['anteriorAnterior'] /
se['anteriorAnterior']
anteriorAnterior
      4.38874

> coef(fit4)[3]/se[3] #also coef(fit4)['hcabg']/se['hcabg']
      hcabg
      2.076276

> coef(fit4)[2]/se[2] #value in table is 4.389
anteriorAnterior
      4.38874
> 2*pnorm(4.38874, lower.tail=FALSE) #complementary normal probability x 2
[1] 1.140093e-05

> coef(fit4)[3]/se[3] #hcabg z value in table is 2.076
      hcabg
      2.076276
> 2*pnorm(2.076255, lower.tail=FALSE) #complementary normal probability x 2
[1] 0.03787037

> qnorm(0.05/2) #normal quantile for p = 0.025
[1] -1.959964
> qnorm(1-0.05/2) #normal quantile for p= 0.975
[1] 1.959964

> matrix(c(coef(fit4)-1.96*se, coef(fit4)+ 1.96*se),ncol=2) #Wald 95% C.I.s for
coefficients
      [,1]      [,2]
[1,] -4.36516489 -3.779240
[2,]  0.39991781  1.045387
[3,]  0.04023802  1.396777
[4,]  0.59819803  1.292485
[5,]  0.66308983  1.690679
[6,]  2.27376174  3.567500
> confint(fit4) #more accurate 95% C.I.s
Waiting for profiling to be done...
      2.5 %      97.5 %
(Intercept) -4.37620420 -3.789572
anteriorAnterior 0.40359365 1.050274
hcabg -0.02064490 1.349783
kk2 0.59435604 1.290122
kk3 0.63838746 1.671180
kk4 2.25301463 3.554909

```

```

> 0.7227 - 1.96*0.1646606 #95% Wald LCL for anterior
[1] 0.3999652
> 0.7227 + 1.96*0.1646606 #95% Wald UCL for anterior
[1] 1.045435
> 0.7185 - 1.96*0.3460558 #95% Wald LCL for hcabg
[1] 0.04023063
> 0.7185 + 1.96*0.3460558 #95% Wald UCL for hcabg
[1] 1.396769

> exp(coef(fit4)) #odds ratios for coefficients
      (Intercept) anteriorAnterior      hcabg      kk2
      0.01703982      2.05989014      2.05136913      2.57369176
           kk3           kk4
      3.24425144      18.55298477

> exp(matrix(c(coef(fit4)-1.96*se, coef(fit4)+ 1.96*se),ncol=2)) #Wald 95%
C.I.s for odds ratios
      [,1] [,2]
[1,] 0.01271256 0.02284004
[2,] 1.49170208 2.84450054
[3,] 1.04105854 4.04215051
[4,] 1.81883835 3.64182406
[5,] 1.94077975 5.42316427
[6,] 9.71588078 35.42789910

> exp(confint(fit4)) #more accurate 95% C.I.s for odds ratios
Waiting for profiling to be done...
           2.5 %      97.5 %
(Intercept) 0.01257299 0.02260528
anteriorAnterior 1.49719545 2.85843441
hcabg 0.97956674 3.85658932
kk2 1.81186381 3.63323146
kk3 1.89342519 5.31843869
kk4 9.51638105 34.98463815

> exp(0.7227 - 1.96*0.1646606) #95% Wald LCL for anterior odds ratio
[1] 1.491773
> exp(0.7227 + 1.96*0.1646606) #95% Wald UCL for anterior odds ratio
[1] 2.844635

```

SECTION 5.5:

```

#SECTION 5.5: ODDS RATIOS

fit5<- glm(death ~ anterior, data=heart, family=binomial(link=logit))
fit5 #show fit
exp(coef(fit5)) #show odds ratios for coefficients (unit change in covariates)

library('gmodels') #package for CrossTable()
CrossTable(death, anterior, prop.r=FALSE, prop.c=FALSE,
  prop.t=FALSE, prop.chisq=FALSE)
(2504*120)/(2005*67) #calc. odds ratio
(120/2125)/(67/2571) #calc. risk ratio

fit6<- glm(death ~ anterior, data=heart, family=poisson)
fit6
exp(coef(fit6)) #show the IRR

```

```

fit7<- glm(death ~ anterior, data=heart, family=binomial(link=log))
fit7
exp(coef(fit7)) #show RR

mean(death[anterior=='Inferior'], na.rm=TRUE) #get risk for unexposed
2.2368/(1-0.02606 + 0.02606*2.2368)

library('epitools') #package for oddsratio() and riskratio()
oddsratio(anterior, death, method='wald')
riskratio(anterior, death, method='wald')

CrossTable(death, anterior, prop.r=FALSE, prop.c=FALSE,
  prop.t=FALSE, prop.chisq=FALSE) #show 2x2 table again

library('reshape')
dfcount<- data.frame(death=death, anterior=anterior) #new, simple data frame
dfcounts<- data.frame(cast(melt(dfcount, measure="death"),
  anterior~ .,
  function(x) { c(Lived=sum(x==0), Died=sum(x==1)) }))) #create data frame of
counts
dfcounts #show

fit8<- glm(cbind(Died,Lived)~ anterior, data=dfcounts, family=binomial)
summary(fit8) #show fit based on counts instead of binary data
confint(fit8) #Confidence intervals on coefs
exp(confint(fit8)) #C.I. for odds ratios

(120*2504)/(2005*67) #compute OR
log(2.2368) #coefficient
sqrt(1/120 + 1/67 + 1/2005 + 1/2504) #S.E. of log(OR)=coefficient
0.8050463 - 1.96*0.1554246 #Wald 95% LCL
0.8050463 + 1.96*0.1554246 #Wald 95% UCL
exp(0.5004141) #OR LCL
exp(1.109679) #OR UCL
#NOTE: confint() returns more accurate C.I.'s than the simple Wald C.I.s
exp(coef(fit8)) #get ORs

120/2125 #risk for anterior
67/2571 #risk for inferior
0.05647059/0.0260599 #risk ratio
log(2.166953) #ln(RR)

0.05647059 - 1.96*sqrt(0.05647059*(1-0.05647059)/4696) #Wald 95% LCL for
proportion
0.05647059 + 1.96*sqrt(0.05647059*(1-0.05647059)/4696) #Wald 95% UCL for
proportion
0.0260599 - 1.96*sqrt(0.0260599*(1-0.0260599)/4696) #Wald 95% LCL for
proportion
0.0260599 + 1.96*sqrt(0.0260599*(1-0.0260599)/4696) #Wald 95% UCL for
proportion

sqrt(1/120 - 1/2125 + 1/67 -1/2571) #S.E. for ln(RR)
0.773322 - 1.96*0.1496635 #Wald 95% LCL for ln(RR)
0.773322 + 1.96*0.1496635 #Wald 95% UCL for ln(RR)
exp(0.4799815) #Wald 95% LCL for RR
exp(1.066662) #Wald 95% UCL for RR
riskratio(anterior, death, method='wald')$measure #show RR C.I. again
riskratio(anterior, death, method='boot')$measure #show RR bootstrap C.I.
exp(confint(fit6)) #95% C.I. for Poisson RR fit above
exp(confint(fit7)) #95% C.I. from binomial RR fit above

oddsratio(death ~ anterior, death, method='wald')

```

```

riskratio(anterior, death, method='wald')

#oddsrisk() not available in R
fit9<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
  data=heart, family=poisson) #poisson loglinear fit
summary(fit9)
exp(coef(fit9)) #show the IRRs
exp(confint(fit9)) #95% C.I.s for IRRs

fit10<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
  data=heart, family=binomial(link=logit)) #logistic fit
summary(fit10)
exp(coef(fit10)) #show the ORs
exp(confint(fit10)) #95% C.I.s for ORs

1.36842/(1-0.02606 + 0.02606*1.36842) #OR->RR C.I.
2.6422/(1-0.02606 + 0.02606*2.26422) #OR->RR C.I.

```


OUTPUT: 5.5

```
> fit5<- glm(death ~ anterior, data=heart, family=binomial(link=logit))
> fit5 #show fit
```

```
Call: glm(formula = death ~ anterior, family = binomial(link = logit),
data = heart)
```

Coefficients:

```
(Intercept)  anteriorAnterior
      -3.621           0.805
```

```
Degrees of Freedom: 4695 Total (i.e. Null);  4694 Residual
(692 observations deleted due to missingness)
```

```
Null Deviance:      1572
```

```
Residual Deviance: 1544      AIC: 1548
```

```
> exp(coef(fit5)) #show odds ratios for coefficients (unit change in
covariates)
```

```
(Intercept)  anteriorAnterior
  0.02675719    2.23679607
```

```
> library('gmodels') #package for CrossTable()
```

```
> CrossTable(death, anterior, prop.r=FALSE, prop.c=FALSE,
+   prop.t=FALSE, prop.chisq=FALSE)
```

```
      Cell Contents
|-----|
|              N |
|-----|
```

```
Total Observations in Table:  4696
```

	anterior	
death	Inferior	Anterior
0	2504	2005
1	67	120
Column Total	2571	2125

```
> (2504*120)/(2005*67) #calc. odds ratio
```

```
[1] 2.236796
```

```
> (120/2125)/(67/2571) #calc. risk ratio
```

```
[1] 2.166953
```

```
> fit6<- glm(death ~ anterior, data=heart, family=poisson)
```

```
> fit6
```

```
Call: glm(formula = death ~ anterior, family = poisson, data = heart)
```

Coefficients:

```
(Intercept)  anteriorAnterior
      -3.6474           0.7733
```

```
Degrees of Freedom: 4695 Total (i.e. Null);  4694 Residual
```

```

(692 observations deleted due to missingness)
Null Deviance:      1206
Residual Deviance: 1179      AIC: 1557
> exp(coef(fit6)) #show the IRR
      (Intercept) anteriorAnterior
      0.0260599      2.1669535

> fit7<- glm(death ~ anterior, data=heart, family=binomial(link=log))
> fit7

Call:  glm(formula = death ~ anterior, family = binomial(link = log),
           data
           = heart)

Coefficients:
      (Intercept)  anteriorAnterior
      -3.6474      0.7733

Degrees of Freedom: 4695 Total (i.e. Null);  4694 Residual
(692 observations deleted due to missingness)
Null Deviance:      1572
Residual Deviance: 1544      AIC: 1548
> exp(coef(fit7)) #show RR
      (Intercept) anteriorAnterior
      0.0260599      2.1669535

> mean(death[anterior=='Inferior'], na.rm=TRUE) #get risk for unexposed
[1] 0.0260599
> 2.2368/(1-0.02606 + 0.02606*2.2368)
[1] 2.166957

> library('epitools') #package for oddsratio() and riskratio()
> oddsratio(anterior, death, method='wald')
$data
      Outcome
Predictor    0    1 Total
  Inferior 2504  67  2571
  Anterior 2005 120  2125
  Total   4509 187  4696

$measure
      odds ratio with 95% C.I.
Predictor estimate    lower    upper
  Inferior 1.000000      NA      NA
  Anterior 2.236796 1.649410 3.033361

$p.value
      two-sided
Predictor midp.exact fisher.exact  chi.square
  Inferior      NA      NA      NA
  Anterior 1.222038e-07 1.199299e-07 1.128655e-07

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"

> riskratio(anterior, death, method='wald')
$data
      Outcome
Predictor    0    1 Total
  Inferior 2504  67  2571
  Anterior 2005 120  2125

```

```

Total      4509 187  4696

$measure
      risk ratio with 95% C.I.
Predictor estimate      lower      upper
Inferior 1.000000      NA      NA
Anterior 2.166953 1.616054 2.905651

$p.value
      two-sided
Predictor midp.exact fisher.exact chi.square
Inferior      NA      NA      NA
Anterior 1.222038e-07 1.199299e-07 1.128655e-07

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"

> CrossTable(death, anterior, prop.r=FALSE, prop.c=FALSE,
+   prop.t=FALSE, prop.chisq=FALSE) #show 2x2 table again

```

```

      Cell Contents
|-----|
|              N |
|-----|

```

Total Observations in Table: 4696

	anterior		
death	Inferior	Anterior	Row Total
0	2504	2005	4509
1	67	120	187
Column Total	2571	2125	4696

```

> library('reshape')
> dframe<- data.frame(death=death, anterior=anterior) #new, simple data frame
> dfcounts<- data.frame(cast(melt(dframe, measure="death"),
+   anterior~.,
+   function(x) { c(Lived=sum(x==0), Died=sum(x==1)) }))) #create data frame of
counts
> dfcounts #show
      anterior Lived Died
1 Inferior 2504    67
2 Anterior 2005   120
3    <NA>   637    55

> fit8<- glm(cbind(Died,Lived)~ anterior, data=dfcounts, family=binomial)
> summary(fit8) #show fit based on counts instead of binary data

```

```

Call:
glm(formula = cbind(Died, Lived) ~ anterior, family = binomial,
    data = dfcounts)

```

Deviance Residuals:

```

[1] 0 0

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.6210     0.1238  -29.25 < 2e-16 ***
anteriorAnterior  0.8050     0.1554   5.18 2.22e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2.8143e+01  on 1  degrees of freedom
Residual deviance: 1.4877e-13  on 0  degrees of freedom
(1 observation deleted due to missingness)
AIC: 16.587

Number of Fisher Scoring iterations: 3

> confint(fit8) #Confidence intervals on coefs
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)   -3.8732890 -3.387273
anteriorAnterior 0.5039546 1.114299

> exp(confint(fit8)) #C.I. for odds ratios
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)    0.02078988 0.03380074
anteriorAnterior 1.65525429 3.04743119

> (120*2504)/(2005*67) #compute OR
[1] 2.236796
> log(2.2368) #coefficient
[1] 0.8050463
> sqrt(1/120 + 1/67 + 1/2005 + 1/2504) #S.E. of log(OR)=coefficient
[1] 0.1554246

> 0.8050463 - 1.96*0.1554246 #Wald 95% LCL
[1] 0.5004141
> 0.8050463 + 1.96*0.1554246 #Wald 95% UCL
[1] 1.109679

> #NOTE: confint() returns more accurate C.I.'s than the simple Wald C.I.s
> exp(coef(fit8)) #get ORs
      (Intercept) anteriorAnterior
      0.02675719      2.23679607

> 120/2125 #risk for anterior
[1] 0.05647059
> 67/2571 #risk for inferior
[1] 0.0260599
> 0.05647059/0.0260599 #risk ratio
[1] 2.166953
> log(2.166953) #ln(RR)
[1] 0.773322

> 0.05647059 - 1.96*sqrt(0.05647059*(1-0.05647059)/4696) #Wald 95% LCL for
proportion
[1] 0.04986851
> 0.05647059 + 1.96*sqrt(0.05647059*(1-0.05647059)/4696) #Wald 95% UCL for
proportion
[1] 0.06307267

```

```

> 0.0260599 - 1.96*sqrt(0.0260599*(1-0.0260599)/4696) #Wald 95% LCL for
proportion
[1] 0.02150326
> 0.0260599 + 1.96*sqrt(0.0260599*(1-0.0260599)/4696) #Wald 95% UCL for
proportion
[1] 0.03061654

> sqrt(1/120 - 1/2125 + 1/67 -1/2571) #S.E. for ln(RR)
[1] 0.1496635
> 0.773322 - 1.96*0.1496635 #Wald 95% LCL for ln(RR)
[1] 0.4799815
> 0.773322 + 1.96*0.1496635 #Wald 95% UCL for ln(RR)
[1] 1.066662
> exp(0.4799815) #Wald 95% LCL for RR
[1] 1.616045
> exp(1.066662) #Wald 95% UCL for RR
[1] 2.905664

> riskratio(anterior, death, method='wald')$measure #show RR C.I. again
risk ratio with 95% C.I.
Predictor estimate lower upper
Inferior 1.000000 NA NA
Anterior 2.166953 1.616054 2.905651
> riskratio(anterior, death, method='boot')$measure #show RR bootstrap C.I.
risk ratio with 95% C.I.
Predictor estimate lower upper
Inferior 1.000000 NA NA
Anterior 2.166953 1.6281 2.963094

> exp(confint(fit6)) #95% C.I. for Poisson RR fit above
Waiting for profiling to be done...
2.5 % 97.5 %
(Intercept) 0.02030753 0.03280671
anteriorAnterior 1.61296705 2.93597994

> exp(confint(fit7)) #95% C.I. from binomial RR fit above
Waiting for profiling to be done...
2.5 % 97.5 %
(Intercept) 0.02036641 0.03269551
anteriorAnterior 1.62220381 2.92021380

> #oddsrisk() not available in R
> fit9<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
+ data=heart, family=poisson) #poisson loglinear fit
> summary(fit9)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
age4, family = poisson, data = heart)

Deviance Residuals:
Min 1Q Median 3Q Max
-1.2202 -0.3201 -0.1756 -0.1318 2.7406

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.7468 0.1864 -25.472 < 2e-16 ***
anteriorAnterior 0.5741 0.1609 3.568 0.00036 ***
hcabg 0.7059 0.3272 2.158 0.03096 *
kk2 0.7681 0.1731 4.438 9.08e-06 ***
kk3 0.7437 0.2541 2.927 0.00342 **
kk4 2.1015 0.2800 7.504 6.19e-14 ***
age3 1.2012 0.1950 6.160 7.29e-10 ***

```

```

age4                1.7761      0.1998    8.888 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1141.2  on 4502  degrees of freedom
Residual deviance:  947.9  on 4495  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1315.9

Number of Fisher Scoring iterations: 6

> exp(coef(fit9)) #show the IRRs
      (Intercept) anteriorAnterior      hcabg      kk2
      0.008679835      1.775474733      2.025651052      2.155711140
      kk3      kk4      age3      age4
      2.103696912      8.178385137      3.324076713      5.906523270
> exp(confint(fit9)) #95% C.I.s for IRRs
Waiting for profiling to be done...
      2.5 %      97.5 %
(Intercept)      0.005936803      0.01233640
anteriorAnterior 1.300263995      2.44611547
hcabg            0.997418512      3.64585852
kk2              1.529438194      3.01903680
kk3              1.245563955      3.39029289
kk4              4.563273615     13.77615448
age3             2.276791879      4.90150623
age4             4.005251095      8.78609393

> fit10<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
+   data=heart, family=binomial(link=logit)) #logistic fit
> summary(fit10)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial(link = logit), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.8159     0.1934  -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg          0.7864     0.3527   2.229 0.025786 *
kk2            0.8249     0.1804   4.572 4.84e-06 ***
kk3            0.7967     0.2692   2.959 0.003083 **
kk4            2.6837     0.3565   7.529 5.11e-14 ***
age3           1.2668     0.2006   6.316 2.68e-10 ***
age4           1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3

Number of Fisher Scoring iterations: 7

```

```

> exp(coef(fit10)) #show the ORs
      (Intercept) anteriorAnterior      hcabg      kk2
0.008099558      1.894095884      2.195519114      2.281692192
      kk3      kk4      age3      age4
2.218198595      14.639836908      3.549576517      6.964846706
> exp(confint(fit10)) #95% C.I.s for ORs
Waiting for profiling to be done...
      2.5 %      97.5 %
(Intercept)      0.005461512      0.01166807
anteriorAnterior 1.368418117      2.64220438
hcabg      1.037061789      4.19021301
kk2      1.595839214      3.24176424
kk3      1.278043511      3.68924231
kk4      7.154252753      29.13630155
age3      2.404385923      5.28930691
age4      4.645436584      10.52267115

> 1.36842/(1-0.02606 + 0.02606*1.36842) #OR->RR C.I.
[1] 1.355407
> 2.6422/(1-0.02606 + 0.02606*2.26422) #OR->RR C.I.
[1] 2.557928

```

SECTION 5.6:

```

#SECTION 5.6: SCALING OF STANDARD ERRORS
x1<- rnorm(50000, mean=0, sd=1) #unit normal covariate
x2<- rnorm(50000, 0, 1) #unit normal covariate
xb<- 1 + 0.25*x1 - 1.5*x2 #model logit in x1 and x2
p<- 1/(1+exp(-xb)) #response probabilities
y<- rbinom(50000, size=100, prob=p) #random binomial counts
fit11<- glm(cbind(y, 100-y) ~ x1 + x2, family=binomial) #logistic fit on
counts
summary(fit11)

fit11$deviance/fit11$df.residual #residual deviance per d.f.
sum(residuals(fit11, type='pearson')^2)/fit11$df.residual #Pearson X2 per d.f.
sum(residuals(fit11, type='deviance')^2)/fit11$df.residual #residual deviance
per d.f.
rm(x1, x2, xb, p, y) #clean up large arrays

fit12<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
  data=heart, family=binomial(link=logit)) #ordinary logistic fit
showLogistic(fit12)
sum(residuals(fit12, type='pearson')^2)/fit12$df.residual #Pearson X2 per d.f.
sum(residuals(fit12, type='deviance')^2)/fit12$df.residual #residual deviance
per d.f.

fit13<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
  data=heart, family=quasibinomial) #quasi-binomial logistic fit
showLogistic(fit13)

```

OUTPUT: 5.6

```

> #SECTION 5.6: SCALING OF STANDARD ERRORS
> x1<- rnorm(50000, mean=0, sd=1) #unit normal covariate
> x2<- rnorm(50000, 0, 1) #unit normal covariate
> xb<- 1 + 0.25*x1 - 1.5*x2 #model logit in x1 and x2
> p<- 1/(1+exp(-xb)) #response probabilities
> y<- rbinom(50000, size=100, prob=p) #random binomial counts
> fit11<- glm(cbind(y, 100-y) ~ x1 + x2, family=binomial) #logistic fit on
counts
> summary(fit11)

Call:
glm(formula = cbind(y, 100 - y) ~ x1 + x2, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.12582  -0.65928   0.02306   0.70585   4.16923

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.996865   0.001224   814.5  <2e-16 ***
x1           0.247502   0.001135   218.0  <2e-16 ***
x2          -1.498330   0.001522  -984.3  <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: 1596198  on 49999  degrees of freedom
Residual deviance:  50626  on 49997  degrees of freedom
AIC: 269509

Number of Fisher Scoring iterations: 4

> fit11$deviance/fit11$df.residual #residual deviance per d.f.
[1] 1.012576
> sum(residuals(fit11, type='pearson')^2)/fit11$df.residual #Pearson X2 per
d.f.
[1] 0.9938317
> sum(residuals(fit11, type='deviance')^2)/fit11$df.residual #residual deviance
per d.f.
[1] 1.012576

> rm(x1, x2, xb, p, y) #clean up large arrays

> fit12<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
+ data=heart, family=binomial(link=logit)) #ordinary logistic fit
> showLogistic(fit12)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
age4, family = binomial(link = logit), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.8159    0.1934  -24.896  < 2e-16 ***
anteriorAnterior  0.6387    0.1675   3.812 0.000138 ***
hcabg         0.7864    0.3527   2.229 0.025786 *
kk2           0.8249    0.1804   4.572 4.84e-06 ***
kk3           0.7967    0.2692   2.959 0.003083 **

```



```

kk4          2.6837      0.3565      7.529 5.11e-14 ***
age3          1.2668      0.2006      6.316 2.68e-10 ***
age4          1.9409      0.2080      9.329 < 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: 1486.2 on 4502 degrees of freedom
Residual deviance: 1276.3 on 4495 degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3

```

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable	Coefficient	LCL	UCL
(Intercept)	-4.8159	-5.21	-4.4509
anteriorAnterior	0.63874	0.31366	0.97161
hcabg	0.78642	0.036392	1.4328
kk2	0.82492	0.4674	1.1761
kk3	0.7967	0.24533	1.3054
kk4	2.6837	1.9677	3.372
age3	1.2668	0.8773	1.6657
age4	1.9409	1.5359	2.3535

Variable	Odds Ratio	LCL	UCL
(Intercept)	0.0080996	0.0054615	0.011668
anteriorAnterior	1.8941	1.3684	2.6422
hcabg	2.1955	1.0371	4.1902
kk2	2.2817	1.5958	3.2418
kk3	2.2182	1.278	3.6892
kk4	14.64	7.1543	29.136
age3	3.5496	2.4044	5.2893
age4	6.9648	4.6454	10.523

```

> sum(residuals(fit12, type='pearson')^2)/fit12$df.residual #Pearson X2 per
d.f.
[1] 0.9371817
> sum(residuals(fit12, type='deviance')^2)/fit12$df.residual #residual deviance
per d.f.
[1] 0.2839420

```

```

> fit13<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
+   data=heart, family=quasibinomial) #quasi-binomial logistic fit
> showLogistic(fit13)

```

Call:

```

glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = quasibinomial, data = heart)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3724	-0.3257	-0.1745	-0.1270	3.1061

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.8159	0.1873	-25.717	< 2e-16 ***
anteriorAnterior	0.6387	0.1622	3.938	8.34e-05 ***
hcabg	0.7864	0.3415	2.303	0.02133 *
kk2	0.8249	0.1747	4.723	2.40e-06 ***
kk3	0.7967	0.2606	3.057	0.00225 **

```

kk4                2.6837      0.3451    7.777 9.12e-15 ***
age3                1.2668      0.1942    6.525 7.58e-11 ***
age4                1.9409      0.2014    9.637 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.937182)

Null deviance: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
(885 observations deleted due to missingness)
AIC: NA

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable            Coefficient      LCL      UCL
(Intercept)        -4.8159        -5.197    -4.4621
anteriorAnterior    0.63874         0.32395    0.96083
hcabg               0.78642         0.062411   1.4133
kk2                 0.82492         0.47895    1.165
kk3                 0.7967         0.26374    1.2897
kk4                 2.6837         1.9912     3.3502
age3                1.2668         0.88965    1.6527
age4                1.9409         1.5488     2.3402

Variable            Odds Ratio      LCL      UCL
(Intercept)         0.0080996   0.0055333  0.011538
anteriorAnterior     1.8941         1.3826     2.6139
hcabg                2.1955         1.0644     4.1097
kk2                  2.2817         1.6144     3.2058
kk3                  2.2182         1.3018     3.6317
kk4                  14.64         7.3244     28.51
age3                  3.5496         2.4343     5.2213
age4                  6.9648         4.7056    10.383

```

SECTION 5.7:

#SECTION 5.7: ROBUST VARIANCE ESTIMATORS

```

h<- na.omit(heart) #drop incomplete rows
nrow(h)
fit16<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
  data=h, family=binomial(link=logit)) #ordinary logistic fit
summary(fit16) #show ordinary logistic fit
exp(coef(fit16)) #odds ratios

library('sandwich') #package for variance sandwich estimation
vcovHC(fit16)
sqrt(diag(vcovHC(fit16)))
sqrt(diag(vcovHC(fit16, type="HC0"))))

library('haplo.ccs') #package for sandcov()
sandcov(fit16, h$center)
sqrt(diag(sandcov(fit16, h$center)))

```

OUTPUT: 5.7

```
> #SECTION 5.7: ROBUST VARIANCE ESTIMATORS
> h<- na.omit(heart) #drop incomplete rows
> nrow(h)
[1] 4503
> fitl6<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
+   data=h, family=binomial(link=logit)) #ordinary logistic fit
> summary(fitl6) #show ordinary logistic fit

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial(link = logit), data = h)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8159     0.1934 -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg           0.7864     0.3527   2.229 0.025786 *
kk2             0.8249     0.1804   4.572 4.84e-06 ***
kk3             0.7967     0.2692   2.959 0.003083 **
kk4             2.6837     0.3565   7.529 5.11e-14 ***
age3            1.2668     0.2006   6.316 2.68e-10 ***
age4            1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
AIC: 1292.3

Number of Fisher Scoring iterations: 7

> exp(coef(fitl6)) #odds ratios
              (Intercept) anteriorAnterior          hcabg          kk2
0.008099558      1.894095884      2.195519114      2.281692192
              kk3          kk4          age3          age4
2.218198595     14.639836908     3.549576517     6.964846706
> library('sandwich') #package for variance sandwich estimation
Loading required package: zoo

Attaching package: 'zoo'

The following object(s) are masked from package:base :

    as.Date.numeric

> vcovHC(fitl6)
              (Intercept) anteriorAnterior          hcabg          kk2
(Intercept)    0.036651043      -0.017588499 -0.0079820171 -0.009992288
anteriorAnterior -0.017588499      0.029984711  0.0051107613 -0.002368952
hcabg          -0.007982017      0.005110761  0.1299467502 -0.002404603
```

```

kk2          -0.009992288      -0.002368952 -0.0024046026  0.032447260
kk3           0.003881687      -0.014417278 -0.0007920807  0.014851781
kk4          -0.002002682      -0.008601965  0.0091263197  0.013934291
age3         -0.023155611       0.002551110 -0.0046954479 -0.000618709
age4         -0.021311638       0.001407039  0.0001376313 -0.003879869

              kk3          kk4          age3          age4
(Intercept)  0.0038816866 -0.002002682 -0.023155611 -0.0213116378
anteriorAnterior -0.0144172784 -0.008601965  0.002551110  0.0014070393
hcabg        -0.0007920807  0.009126320 -0.004695448  0.0001376313
kk2          0.0148517811  0.013934291 -0.000618709 -0.0038798693
kk3          0.0858093674  0.017663676 -0.009547979 -0.0173695742
kk4          0.0176636756  0.181092841 -0.007698816 -0.0117285524
age3         -0.0095479787 -0.007698816  0.040086891  0.0240430007
age4         -0.0173695742 -0.011728552  0.024043001  0.0464017536

> sqrt(diag(vcovHC(fit16)))
              (Intercept) anteriorAnterior          hcabg          kk2
0.1914446          0.1731609          0.3604813          0.1801312
              kk3          kk4          age3          age4
0.2929324          0.4255500          0.2002171          0.2154107
> sqrt(diag(vcovHC(fit16, type="HC0")))
              (Intercept) anteriorAnterior          hcabg          kk2
0.1907744          0.1723218          0.3543301          0.1795156
              kk3          kk4          age3          age4
0.2912805          0.4154713          0.1993996          0.2143083

> library('haplo.ccs') #package for sandcov()
> sandcov(fit16, h$center)
              (Intercept) anteriorAnterior          hcabg          kk2
(Intercept)  0.037013237      -0.0241977557 -0.0122623400 -0.0099555058
anteriorAnterior -0.024197756  0.0407907068  0.0046607637 -0.0002984985
hcabg        -0.012262340  0.0046607637  0.0958409186 -0.0145605642
kk2          -0.009955506  -0.0002984985 -0.0145605642  0.0310858125
kk3          0.018328445  -0.0252363931  0.0008422075  0.0229412893
kk4          -0.018483720  0.0065156016  0.0210195568  0.0167299684
age3         -0.018862126  0.0083740628 -0.0030455500 -0.0015100806
age4         -0.030654037  0.0069987660  0.0111307094  0.0008987864

              kk3          kk4          age3          age4
(Intercept)  0.0183284447 -0.018483720 -0.018862126 -0.0306540373
anteriorAnterior -0.0252363931  0.006515602  0.008374063  0.0069987660
hcabg        0.0008422075  0.021019557 -0.003045550  0.0111307094
kk2          0.0229412893  0.016729968 -0.001510081  0.0008987864
kk3          0.1045252164 -0.006511285 -0.019174047 -0.0261808438
kk4          -0.0065112849  0.088281436  0.007225885  0.0091487239
age3         -0.0191740467  0.007225885  0.027737823  0.0272444860
age4         -0.0261808438  0.009148724  0.027244486  0.0561450054

> sqrt(diag(sandcov(fit16, h$center)))
              (Intercept) anteriorAnterior          hcabg          kk2
0.1923882          0.2019671          0.3095818          0.1763117
              kk3          kk4          age3          age4
0.3233036          0.2971219          0.1665468          0.2369494

```

SECTION 5.8:

```

#SECTION 5.8: BOOTSTRAPPED & JACKKNIFED STANDARD ERRORS
summary(fit16) #show ordinary logistic fit again

```

```

exp(coef(fit16)) #odds ratios
sqrt(diag(vcov(fit16))) #show SEs of coefs

library('boot')
ffit<- function (x, i) {
  xx<- x[i,]
  fglm<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
    data=xx, family=binomial(link=logit)) #ordinary logistic fit on resample
  return(sqrt(diag(vcov(fglm)))) #return SE of coeffs
}
bse<- boot(h, ffit, R=50) #do 50 resamples by cases (rows)
sqrt(diag(vcov(fit16))) #show SEs of coefs
apply(bse$t, 2, mean) #mean estimate of SEs

#jackknife is preferred for small numbers of data, bootstrap for large numbers
#Here, jackknife will require 5,403 model fits, but the bootstrap above
#only required 50 (although 100 might be surer).
#If you must, the following code will generate jackknife estimates of the SEs:
library('bootstrap')
theta<- function (i) {
  xx<- h[i,] #resample
  fglm<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
    data=xx, family=binomial(link=logit)) #ordinary logistic fit on resample
  return(sqrt(diag(vcov(fglm))[iCoef])) #SE of coef[iCoef]
}
sqrt(diag(vcov(fit16))) #show SEs of coefs again
for (iCoef in 1:8) { #repeat these time-consuming calculations for each coef
  jSE<- jackknife(1:nrow(h), theta) #perform jackknife for coef[iCoef]
  cat(iCoef, jSE$jack.se, '\n')
}

kk23New<- as.numeric(kk2==1 | kk3==1) #new variable equiv to merging kk2 and
kk3
sum(kk23New!=kk23, na.rm=TRUE) #Equal to kk23 in dataset?

fit17<- glm(death ~ anterior + hcabg + kk23 + kk4 + age3 + age4,
  data=heart, family=binomial(link=logit))
summary(fit17) #show fit
exp(coef(fit17)) #odds ratios

abic<- function (glmfit) {
  d <- glmfit$deviance
  nobs <- length(glmfit$y)
  k <- glmfit$rank #number predictors, incl. intercept
  aic <- glmfit$aic
  bic <- d + 2*k*log(k)
  cat('AIC per datum: ',format(aic/nobs, digits=8),' AIC total:
',format(aic,digits=8),'\n')
  cat('BIC per datum: ',format(bic/nobs, digits=8),' BIC total:
',format(bic,digits=8),'\n')
  return(c(aic,bic))
}

abic(fit17) #show AIC and BIC for kk23 fit
abic(fit16) #show AIC and BIC for kk2 and kk3 fit

```

OUTPUT: 5.8

```

> #SECTION 5.8: BOOTSTRAPPED & JACKKNIFED STANDARD ERRORS
> summary(fit16) #show ordinary logistic fit again

```

```

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
     age4, family = binomial(link = logit), data = h)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8159     0.1934  -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg           0.7864     0.3527   2.229 0.025786 *
kk2             0.8249     0.1804   4.572 4.84e-06 ***
kk3             0.7967     0.2692   2.959 0.003083 **
kk4             2.6837     0.3565   7.529 5.11e-14 ***
age3            1.2668     0.2006   6.316 2.68e-10 ***
age4            1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
AIC: 1292.3

Number of Fisher Scoring iterations: 7

> exp(coef(fit16)) #odds ratios
              (Intercept) anteriorAnterior          hcabg          kk2
0.008099558      1.894095884      2.195519114      2.281692192
              kk3          kk4          age3          age4
2.218198595      14.639836908      3.549576517      6.964846706

> sqrt(diag(vcov(fit16))) #show SEs of coefs
              (Intercept) anteriorAnterior          hcabg          kk2
0.1934407      0.1675426      0.3527468      0.1804368
              kk3          kk4          age3          age4
0.2692168      0.3564503      0.2005657      0.2080462

> library('boot')
> ffit<- function (x, i) {
+   xx<- x[i,]
+   fglm<- glm(death ~ anterior + hcabg +kk2 + kk3 + kk4 + age3 + age4,
+     data=xx, family=binomial(link=logit)) #ordinary logistic fit on resample
+   return(sqrt(diag(vcov(fglm)))) #return SE of coeffs
+ }
> bse<- boot(h, ffit, R=50) #do 50 resamples by cases (rows)
> sqrt(diag(vcov(fit16))) #show SEs of coefs
              (Intercept) anteriorAnterior          hcabg          kk2
0.1934407      0.1675426      0.3527468      0.1804368
              kk3          kk4          age3          age4
0.2692168      0.3564503      0.2005657      0.2080462

> apply(bse$t, 2, mean) #mean estimate of SEs
[1] 0.1943857 0.1682196 0.3703508 0.1810382 0.2766298 0.3656162 0.2017321
[8] 0.2092239

NOTE: [jackknife omitted]

```

```

> kk23New<- as.numeric(kk2==1 | kk3==1) #new variable equiv to merging kk2 and
kk3
> sum(kk23New!=kk23, na.rm=TRUE) #Equal to kk23 in dataset?
[1] 0

> fit17<- glm(death ~ anterior + hcabg + kk23 + kk4 + age3 + age4,
+ data=heart, family=binomial(link=logit))
> summary(fit17) #show fit

Call:
glm(formula = death ~ anterior + hcabg + kk23 + kk4 + age3 +
age4, family = binomial(link = logit), data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3719  -0.3256  -0.1745  -0.1271   3.1057

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8145     0.1928  -24.965  < 2e-16 ***
anteriorAnterior  0.6377     0.1672   3.813  0.000137 ***
hcabg           0.7870     0.3527   2.232  0.025634 *
kk23            0.8180     0.1671   4.895  9.85e-07 ***
kk4             2.6840     0.3564   7.531  5.04e-14 ***
age3            1.2659     0.2003   6.318  2.64e-10 ***
age4            1.9392     0.2074   9.351  < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4496  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1290.3

Number of Fisher Scoring iterations: 7

> exp(coef(fit17)) #odds ratios
      (Intercept) anteriorAnterior      hcabg      kk23
      0.008111542      1.892058505      2.196883057      2.265916179
           kk4      age3      age4
      14.643046276      3.546141441      6.952852621

> abic<- function (glmfit) {
+   d <- glmfit$deviance
+   nobs <- length(glmfit$y)
+   k <- glmfit$rank #number predictors, incl. intercept
+   aic <- glmfit$aic
+   bic <- d + 2*k*log(k)
+   cat('AIC per datum: ',format(aic/nobs, digits=8),' AIC total:
',format(aic,digits=8),'\n')
+   cat('BIC per datum: ',format(bic/nobs, digits=8),' BIC total:
',format(bic,digits=8),'\n')
+   return(c(aic,bic))
+ }

> abic(fit17) #show AIC and BIC for kk23 fit
AIC per datum: 0.28654883 AIC total: 1290.3294
BIC per datum: 0.2894897 BIC total: 1303.5721
[1] 1290.329 1303.572

> abic(fit16) #show AIC and BIC for kk2 and kk3 fit

```

```
AIC per datum: 0.2869907    AIC total: 1292.3191
BIC per datum: 0.29082616   BIC total: 1309.5902
[1] 1292.319 1309.590
```

SECTION 5.9:

```
#SECTION 5.9: STEPWISE METHODS
detach(heart) #remove this data frame from easy access

library('foreign')
lrdata<- read.dta('lr_model.dta') #read Stata format file
head(lrdata) #show some rows

fit18<- glm(heartatk ~ reg2 + reg3 + reg4 + race2 + race3 + gender +
  height + weight + bpsystol + tcresult + tgresult + hdbresult + hgb +
  diabetes + albumin + hsi2 + hsi3 + hsi4 + hsi5,
  data=lrdata, family=binomial) #set up overall model
showLogistic(fit18)

fit19<- step(fit18, direction='both') #stepwise fits
showLogistic(fit19) #final fit

fit20<- glm(heartatk ~ reg4 + gender + height + tcresult + tgresult +
  hdbresult + diabetes + albumin + hsi3 + hsi4 + hsi5,
  data=lrdata, family=binomial) #set up overall model
showLogistic(fit20)

lbw<- read.dta('lbw.dta') #read Stata format file
head(lbw) #show some rows
lbw$age<- ifelse(lbwdata$age>34, NA, lbwdata$age) #mark age>34 values as
missing

fit21<- glm(low ~ age + smoke + race2 + race3, data=lbw, family=binomial)
fit22<- step(fit21, direction='both') #stepwise fits have problem due to
missing data

lbw2<- na.omit(lbw) #remove missing data rows
fit23<- glm(low ~ age + smoke + race2 + race3, data=lbw2, family=binomial)
fit24<- step(fit23, direction='both') #stepwise fits
summary(fit24) #show final fit

fit25<- glm(low ~ smoke + race2 + race3, data=lbw, family=binomial)
summary(fit25) #show fit on original dataset lbw
```

OUTPUT: 5.9

```
> library('foreign')
> lrdata<- read.dta('lr_model.dta') #read Stata format file
> head(lrdata) #show some rows
```

	heartatk	strata	region	sex	race	age	height	weight	bpsystol	tcresult	tgresult
1	0	16	3	2	2	72	156.598	71.67	204	215	NA
2	1	6	4	1	2	64	162.199	47.85	140	211	160
3	0	30	2	1	1	64	169.598	87.43	140	187	NA
4	1	14	3	1	1	49	186.199	87.32	156	122	NA
5	0	10	4	1	1	73	163.500	66.45	144	166	NA

	6	0	16	3	2	1	65	157.199	67.59	130	323	NA
	hdresult	hgb	diabetes	albumin	hsizgp	race1	race2	race3	reg1	reg2	reg3	reg4
1	48	11.9	0	3.0	1	0	1	0	0	0	1	0
2	74	11.9	0	3.2	2	0	1	0	0	0	0	1
3	36	14.0	0	3.2	2	1	0	0	0	1	0	0
4	58	16.9	0	3.3	2	1	0	0	0	0	1	0
5	41	15.9	0	3.3	5	1	0	0	0	0	0	1
6	NA	13.6	1	3.3	3	1	0	0	0	0	1	0

	gender	hsi1	hsi2	hsi3	hsi4	hsi5
1	0	1	0	0	0	0
2	1	0	1	0	0	0
3	1	0	1	0	0	0
4	1	0	1	0	0	0
5	1	0	0	0	0	1
6	0	0	0	1	0	0

```
> fit18<- glm(heartatk ~ reg2 + reg3 + reg4 + race2 + race3 + gender +
+ height + weight + bpsystol + tcresult + tgresult + hdresult + hgb +
+ diabetes + albumin + hsi2 + hsi3 + hsi4 + hsi5,
+ data=lrdata, family=binomial) #set up overall model
> showLogistic(fit18)
```

Call:

```
glm(formula = heartatk ~ reg2 + reg3 + reg4 + race2 + race3 +
gender + height + weight + bpsystol + tcresult + tgresult +
hdresult + hgb + diabetes + albumin + hsi2 + hsi3 + hsi4 +
hsi5, family = binomial, data = lrdata)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.1120	-0.3077	-0.2091	-0.1392	3.3401

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	7.7035634	2.4666442	3.123	0.00179 **
reg2	0.3037289	0.2616729	1.161	0.24576
reg3	0.3394275	0.2668867	1.272	0.20344
reg4	0.6387156	0.2542255	2.512	0.01199 *
race2	-0.1891324	0.3217525	-0.588	0.55665
race3	0.1469744	0.6401134	0.230	0.81840
gender	1.6754980	0.2569962	6.520	7.05e-11 ***
height	-0.0417328	0.0133172	-3.134	0.00173 **
weight	-0.0036861	0.0068047	-0.542	0.58802
bpsystol	0.0065695	0.0035833	1.833	0.06674 .
tcresult	0.0039379	0.0018935	2.080	0.03755 *
tgresult	0.0022578	0.0009825	2.298	0.02156 *
hdresult	-0.0181952	0.0076063	-2.392	0.01675 *
hgb	-0.0153563	0.0703747	-0.218	0.82727
diabetes	0.7036702	0.2922886	2.407	0.01606 *
albumin	-1.2779252	0.2773385	-4.608	4.07e-06 ***
hsi2	0.3043360	0.2384840	1.276	0.20191
hsi3	-0.5579459	0.3246219	-1.719	0.08566 .
hsi4	-0.5878732	0.3508653	-1.675	0.09384 .
hsi5	-0.9331491	0.3783298	-2.466	0.01364 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1410.4 on 4118 degrees of freedom
Residual deviance: 1233.8 on 4099 degrees of freedom
(6232 observations deleted due to missingness)
AIC: 1273.8

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable	Coefficient	LCL	UCL
(Intercept)	7.7036	2.8698	12.546
reg2	0.30373	-0.20083	0.82939
reg3	0.33943	-0.17650	0.87423
reg4	0.63872	0.15143	1.1521
race2	-0.18913	-0.86258	0.40742
race3	0.14697	-1.3325	1.2586
gender	1.6755	1.1753	2.1837
height	-0.041733	-0.067832	-0.0156
weight	-0.0036861	-0.017274	0.0094159
bpsystol	0.0065695	-0.00057963	0.013483
tcresult	0.0039379	0.00020236	0.0076307
tgresult	0.0022578	0.00022412	0.00412
hdresult	-0.018195	-0.033468	-0.0036325
hgb	-0.015356	-0.15239	0.12372
diabetes	0.70367	0.098073	1.2505
albumin	-1.2779	-1.8226	-0.73503
hsi2	0.30434	-0.14828	0.79033
hsi3	-0.55795	-1.2079	0.072892
hsi4	-0.58787	-1.2996	0.086536
hsi5	-0.93315	-1.7126	-0.21529

Variable	Odds Ratio	LCL	UCL
(Intercept)	2216.2	17.633	2.8088e+05
reg2	1.3549	0.81806	2.2919
reg3	1.4041	0.8382	2.397
reg4	1.894	1.1635	3.1647
race2	0.82768	0.42207	1.5029
race3	1.1583	0.26382	3.5206
gender	5.3415	3.2392	8.8794
height	0.95913	0.93442	0.98452
weight	0.99632	0.98287	1.0095
bpsystol	1.0066	0.99942	1.0136
tcresult	1.0039	1.0002	1.0077
tgresult	1.0023	1.0002	1.0041
hdresult	0.98197	0.96709	0.99637
hgb	0.98476	0.85865	1.1317
diabetes	2.0212	1.103	3.4921
albumin	0.27861	0.1616	0.47949
hsi2	1.3557	0.86219	2.2041
hsi3	0.57238	0.29883	1.0756
hsi4	0.55551	0.27265	1.0904
hsi5	0.39331	0.18039	0.80631

```
> fit19<- step(fit18, direction='both') #stepwise fits
```

```
Start: AIC=1273.85
```

```
heartatk ~ reg2 + reg3 + reg4 + race2 + race3 + gender + height +
  weight + bpsystol + tcresult + tgresult + hdresult + hgb +
  diabetes + albumin + hsi2 + hsi3 + hsi4 + hsi5
```

	Df	Deviance	AIC
- hgb	1	1233.9	1271.9
- race3	1	1233.9	1271.9
- weight	1	1234.1	1272.1
- race2	1	1234.2	1272.2
- reg2	1	1235.2	1273.2
- reg3	1	1235.5	1273.5
- hsi2	1	1235.5	1273.5

<none>		1233.8	1273.8
- hsi4	1	1236.8	1274.8
- hsi3	1	1236.8	1274.8
- bpsystol	1	1237.1	1275.1
- tcresult	1	1238.1	1276.1
- tgresult	1	1238.5	1276.5
- diabetes	1	1238.9	1276.9
- hdresult	1	1239.9	1277.9
- hsi5	1	1240.4	1278.4
- reg4	1	1240.5	1278.5
- height	1	1243.6	1281.6
- albumin	1	1255.2	1293.2
- gender	1	1277.8	1315.8

Step: AIC=1271.89

heartatk ~ reg2 + reg3 + reg4 + race2 + race3 + gender + height +
weight + bpsystol + tcresult + tgresult + hdresult + diabetes +
albumin + hsi2 + hsi3 + hsi4 + hsi5

	Df	Deviance	AIC
- race3	1	1234.0	1270.0
- race2	1	1234.2	1270.2
- weight	1	1234.2	1270.2
- reg2	1	1235.2	1271.2
- reg3	1	1235.5	1271.5
- hsi2	1	1235.6	1271.6
<none>		1233.9	1271.9
- hsi4	1	1236.8	1272.8
- hsi3	1	1236.9	1272.9
- bpsystol	1	1237.2	1273.2
+ hgb	1	1233.8	1273.8
- tcresult	1	1238.1	1274.1
- tgresult	1	1238.6	1274.6
- diabetes	1	1239.0	1275.0
- hdresult	1	1239.9	1275.9
- hsi5	1	1240.5	1276.5
- reg4	1	1240.5	1276.5
- height	1	1243.7	1279.7
- albumin	1	1257.5	1293.5
- gender	1	1280.8	1316.8

Step: AIC=1269.95

heartatk ~ reg2 + reg3 + reg4 + race2 + gender + height + weight +
bpsystol + tcresult + tgresult + hdresult + diabetes + albumin +
hsi2 + hsi3 + hsi4 + hsi5

	Df	Deviance	AIC
- race2	1	1234.3	1268.3
- weight	1	1234.3	1268.3
- reg2	1	1235.3	1269.3
- reg3	1	1235.6	1269.6
- hsi2	1	1235.6	1269.6
<none>		1234.0	1270.0
- hsi4	1	1236.8	1270.8
- hsi3	1	1236.9	1270.9
- bpsystol	1	1237.3	1271.3
+ race3	1	1233.9	1271.9
+ hgb	1	1233.9	1271.9
- tcresult	1	1238.1	1272.1
- tgresult	1	1238.7	1272.7
- diabetes	1	1239.0	1273.0
- hdresult	1	1240.0	1274.0
- hsi5	1	1240.5	1274.5

```

- reg4      1    1240.8 1274.8
- height    1    1244.0 1278.0
- albumin   1    1257.5 1291.5
- gender    1    1281.2 1315.2

```

Step: AIC=1268.28

```

heartatk ~ reg2 + reg3 + reg4 + gender + height + weight + bpsystol +
          tcresult + tresult + hdresult + diabetes + albumin + hsi2 +
          hsi3 + hsi4 + hsi5

```

	Df	Deviance	AIC
- weight	1	1234.7	1266.7
- reg2	1	1235.6	1267.6
- reg3	1	1235.8	1267.8
- hsi2	1	1236.2	1268.2
<none>		1234.3	1268.3
- hsi4	1	1237.1	1269.1
- hsi3	1	1237.1	1269.1
- bpsystol	1	1237.6	1269.6
+ race2	1	1234.0	1270.0
+ race3	1	1234.2	1270.2
+ hgb	1	1234.3	1270.3
- tcresult	1	1238.5	1270.5
- tresult	1	1239.2	1271.2
- diabetes	1	1239.2	1271.2
- hsi5	1	1240.8	1272.8
- hdresult	1	1240.8	1272.8
- reg4	1	1241.1	1273.1
- height	1	1244.3	1276.3
- albumin	1	1257.5	1289.5
- gender	1	1281.3	1313.3

Step: AIC=1266.68

```

heartatk ~ reg2 + reg3 + reg4 + gender + height + bpsystol +
          tcresult + tresult + hdresult + diabetes + albumin + hsi2 +
          hsi3 + hsi4 + hsi5

```

	Df	Deviance	AIC
- reg2	1	1236.0	1266.0
- reg3	1	1236.2	1266.2
- hsi2	1	1236.5	1266.5
<none>		1234.7	1266.7
- hsi4	1	1237.6	1267.6
- bpsystol	1	1237.6	1267.6
- hsi3	1	1237.7	1267.7
+ weight	1	1234.3	1268.3
+ race2	1	1234.3	1268.3
+ race3	1	1234.6	1268.6
+ hgb	1	1234.7	1268.7
- tcresult	1	1238.8	1268.8
- tresult	1	1239.4	1269.4
- diabetes	1	1239.5	1269.5
- hdresult	1	1240.8	1270.8
- hsi5	1	1241.3	1271.3
- reg4	1	1241.5	1271.5
- height	1	1249.0	1279.0
- albumin	1	1257.9	1287.9
- gender	1	1283.0	1313.0

Step: AIC=1265.99

```

heartatk ~ reg3 + reg4 + gender + height + bpsystol + tcresult +
          tresult + hdresult + diabetes + albumin + hsi2 + hsi3 +
          hsi4 + hsi5

```

	Df	Deviance	AIC
- reg3	1	1236.5	1264.5
- hsi2	1	1237.9	1265.9
<none>		1236.0	1266.0
+ reg2	1	1234.7	1266.7
- hsi4	1	1238.9	1266.9
- bpsystol	1	1238.9	1266.9
- hsi3	1	1239.0	1267.0
+ weight	1	1235.6	1267.6
+ race2	1	1235.7	1267.7
- tcresult	1	1239.9	1267.9
+ race3	1	1235.9	1267.9
+ hgb	1	1236.0	1268.0
- diabetes	1	1240.8	1268.8
- tgresult	1	1240.9	1268.9
- reg4	1	1241.8	1269.8
- hdresult	1	1242.3	1270.3
- hsi5	1	1242.7	1270.7
- height	1	1250.1	1278.1
- albumin	1	1259.6	1287.6
- gender	1	1283.6	1311.6

Step: AIC=1264.48

heartatk ~ reg4 + gender + height + bpsystol + tcresult + tgresult +
hdresult + diabetes + albumin + hsi2 + hsi3 + hsi4 + hsi5

	Df	Deviance	AIC
- hsi2	1	1238.4	1264.4
<none>		1236.5	1264.5
- hsi4	1	1239.3	1265.3
- bpsystol	1	1239.3	1265.3
- hsi3	1	1239.5	1265.5
+ reg3	1	1236.0	1266.0
+ weight	1	1236.1	1266.1
+ reg2	1	1236.2	1266.2
+ race2	1	1236.2	1266.2
+ race3	1	1236.4	1266.4
+ hgb	1	1236.5	1266.5
- tcresult	1	1240.5	1266.5
- tgresult	1	1241.3	1267.3
- diabetes	1	1241.3	1267.3
- reg4	1	1241.9	1267.9
- hdresult	1	1243.0	1269.0
- hsi5	1	1243.1	1269.1
- height	1	1250.3	1276.3
- albumin	1	1260.4	1286.4
- gender	1	1283.8	1309.8

Step: AIC=1264.39

heartatk ~ reg4 + gender + height + bpsystol + tcresult + tgresult +
hdresult + diabetes + albumin + hsi3 + hsi4 + hsi5

	Df	Deviance	AIC
<none>		1238.4	1264.4
+ hsi2	1	1236.5	1264.5
- bpsystol	1	1241.3	1265.3
+ reg3	1	1237.9	1265.9
+ race2	1	1238.0	1266.0
+ weight	1	1238.0	1266.0
+ reg2	1	1238.1	1266.1
+ race3	1	1238.3	1266.3
+ hgb	1	1238.4	1266.4

```

- tcresult 1 1242.7 1266.7
- diabetes 1 1243.0 1267.0
- tgresult 1 1243.3 1267.3
- reg4 1 1243.9 1267.9
- hdresult 1 1245.2 1269.2
- hsi4 1 1247.7 1271.7
- hsi3 1 1249.5 1273.5
- height 1 1252.1 1276.1
- hsi5 1 1255.7 1279.7
- albumin 1 1263.6 1287.6
- gender 1 1288.4 1312.4
> showLogistic(fit19) #final fit

Call:
glm(formula = heartatk ~ reg4 + gender + height + bpsystol +
    tcresult + tgresult + hdresult + diabetes + albumin + hsi3 +
    hsi4 + hsi5, family = binomial, data = lrdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0686  -0.3123  -0.2106  -0.1399   3.3913

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  8.3195984  2.2991869   3.618 0.000296 ***
reg4         0.4122683  0.1728121   2.386 0.017049 *
gender       1.6888905  0.2441555   6.917 4.60e-12 ***
height      -0.0441108  0.0119067  -3.705 0.000212 ***
bpsystol     0.0059747  0.0034471   1.733 0.083046 .
tcresult     0.0039218  0.0018730   2.094 0.036270 *
tgresult     0.0022670  0.0009632   2.354 0.018590 *
hdresult     -0.0185166  0.0073188  -2.530 0.011406 *
diabetes     0.6649855  0.2906585   2.288 0.022146 *
albumin     -1.3132822  0.2611929  -5.028 4.96e-07 ***
hsi3        -0.8033996  0.2620314  -3.066 0.002169 **
hsi4        -0.8222432  0.2939774  -2.797 0.005159 **
hsi5        -1.1749473  0.3245545  -3.620 0.000294 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1410.4  on 4118  degrees of freedom
Residual deviance: 1238.4  on 4106  degrees of freedom
(6232 observations deleted due to missingness)
AIC: 1264.4

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

```

Variable	Coefficient	LCL	UCL
(Intercept)	8.3196	3.8169	12.835
reg4	0.41227	0.068089	0.74679
gender	1.6889	1.2145	2.1725
height	-0.044111	-0.067468	-0.020772
bpsystol	0.0059747	-0.0009063	0.012618
tcresult	0.0039218	0.00022625	0.0075742
tgresult	0.002267	0.00027602	0.004104
hdresult	-0.018517	-0.033247	-0.0045316
diabetes	0.66499	0.062258	1.2081
albumin	-1.3133	-1.8259	-0.80145
hsi3	-0.8034	-1.3485	-0.31566

hsi4	-0.82224	-1.4404	-0.27982
hsi5	-1.1749	-1.8675	-0.58349

Variable	Odds Ratio	LCL	UCL
(Intercept)	4103.5	45.464	3.7511e+05
reg4	1.5102	1.0705	2.1102
gender	5.4135	3.3684	8.7804
height	0.95685	0.93476	0.97944
bpsystol	1.006	0.99909	1.0127
tcresult	1.0039	1.0002	1.0076
tgresult	1.0023	1.0003	1.0041
hdresult	0.98165	0.9673	0.99548
diabetes	1.9445	1.0642	3.3472
albumin	0.26894	0.16107	0.44868
hsi3	0.4478	0.25963	0.7293
hsi4	0.43944	0.23684	0.75592
hsi5	0.30884	0.1545	0.55795

```
> fit20<- glm(heartatk ~ reg4 + gender + height + tcresult + tgresult +
+   hdresult + diabetes + albumin + hsi3 + hsi4 + hsi5,
+   data=lrdata, family=binomial) #set up overall model
> showLogistic(fit20)
```

Call:

```
glm(formula = heartatk ~ reg4 + gender + height + tcresult +
    tgresult + hdresult + diabetes + albumin + hsi3 + hsi4 +
    hsi5, family = binomial, data = lrdata)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.0129	-0.3104	-0.2127	-0.1409	3.3861

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	9.3960972	2.2204404	4.232	2.32e-05	***
reg4	0.4123500	0.1727902	2.386	0.017013	*
gender	1.7231801	0.2436441	7.073	1.52e-12	***
height	-0.0455791	0.0118942	-3.832	0.000127	***
tcresult	0.0042332	0.0018590	2.277	0.022779	*
tgresult	0.0024510	0.0009426	2.600	0.009315	**
hdresult	-0.0182625	0.0073117	-2.498	0.012499	*
diabetes	0.7264639	0.2882642	2.520	0.011731	*
albumin	-1.3454067	0.2600680	-5.173	2.30e-07	***
hsi3	-0.8266638	0.2615931	-3.160	0.001577	**
hsi4	-0.8752568	0.2924404	-2.993	0.002763	**
hsi5	-1.2252713	0.3231113	-3.792	0.000149	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1410.4 on 4118 degrees of freedom
Residual deviance: 1241.3 on 4107 degrees of freedom
(6232 observations deleted due to missingness)
AIC: 1265.3

Number of Fisher Scoring iterations: 7

Waiting for profiling to be done...

Variable	Coefficient	LCL	UCL
(Intercept)	9.3961	5.0428	13.752
reg4	0.41235	0.068218	0.74682

gender	1.7232	1.2497	2.2058
height	-0.045579	-0.068909	-0.022262
tcresult	0.0042332	0.00056548	0.007859
tgresult	0.002451	0.0005137	0.0042663
hdresult	-0.018262	-0.032965	-0.0042755
diabetes	0.72646	0.12802	1.2647
albumin	-1.3454	-1.8558	-0.83568
hsi3	-0.82666	-1.3711	-0.33988
hsi4	-0.87526	-1.4910	-0.33626
hsi5	-1.2253	-1.9157	-0.63711

Variable	Odds Ratio	LCL	UCL
(Intercept)	12041	154.91	9.3873e+05
reg4	1.5104	1.0706	2.1103
gender	5.6023	3.4895	9.0774
height	0.95544	0.93341	0.97798
tcresult	1.0042	1.0006	1.0079
tgresult	1.0025	1.0005	1.0043
hdresult	0.9819	0.96757	0.99573
diabetes	2.0678	1.1366	3.5422
albumin	0.26043	0.15633	0.43358
hsi3	0.43751	0.25383	0.71185
hsi4	0.41675	0.22515	0.71443
hsi5	0.29368	0.14724	0.52882

```
> lbw<- read.dta('lbw.dta') #read Stata format file
```

```
> head(lbw) #show some rows
```

	id	low	age	lwt	race	smoke	ptl	ht	ui	ftv	bwt	race1	race2	race3
1	85	0	19	182	black	0	0	0	1	0	2523	0	1	0
2	86	0	33	155	other	0	0	0	0	3	2551	0	0	1
3	87	0	20	105	white	1	0	0	0	1	2557	1	0	0
4	88	0	21	108	white	1	0	0	1	2	2594	1	0	0
5	89	0	18	107	white	1	0	0	1	0	2600	1	0	0
6	91	0	21	124	other	0	0	0	0	0	2622	0	0	1

```
> lbw$age<- ifelse(lbwdata$age>34, NA, lbwdata$age) #mark age>34 values as missing
```

```
> fit21<- glm(low ~ age + smoke + race2 + race3, data=lbw, family=binomial)
```

```
> fit22<- step(fit21, direction='both') #stepwise fits
```

```
Start: AIC=226.62
```

```
low ~ age + smoke + race2 + race3
```

	Df	Deviance	AIC
- age	1	216.94	224.94
<none>		216.62	226.62
- race2	1	221.36	229.36
- race3	1	223.59	231.59
- smoke	1	225.82	233.82

```
Error in step(fit21, direction = "both") :
```

```
number of rows in use has changed: remove missing values?
```

```
> lbw2<- na.omit(lbw) #remove missing data rows
```

```
> fit23<- glm(low ~ age + smoke + race2 + race3, data=lbw2, family=binomial)
```

```
> fit24<- step(fit23, direction='both') #stepwise fits
```

```
Start: AIC=226.62
```

```
low ~ age + smoke + race2 + race3
```

	Df	Deviance	AIC
- age	1	216.94	224.94
<none>		216.62	226.62
- race2	1	221.36	229.36
- race3	1	223.59	231.59


```
- smoke 1 225.82 233.82
```

```
Step: AIC=224.94  
low ~ smoke + race2 + race3
```

```
      Df Deviance   AIC  
<none>      216.94 224.94  
+ age  1      216.62 226.62  
- race2 1      222.21 228.21  
- race3 1      224.34 230.34  
- smoke 1      226.29 232.29  
> summary(fit24) #show final fit
```

```
Call:
```

```
glm(formula = low ~ smoke + race2 + race3, family = binomial,  
     data = lbw2)
```

```
Deviance Residuals:
```

```
      Min       1Q   Median       3Q      Max  
-1.3784 -0.9010 -0.5519  1.4638  1.9784
```

```
Coefficients:
```

```
      Estimate Std. Error z value Pr(>|z|)  
(Intercept) -1.8048      0.3618  -4.989 6.08e-07 ***  
smoke         1.1131      0.3755   2.964 0.00304 **  
race2         1.1528      0.5008   2.302 0.02133 *  
race3         1.0734      0.4057   2.646 0.00815 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 230.87 on 183 degrees of freedom  
Residual deviance: 216.94 on 180 degrees of freedom  
AIC: 224.94
```

```
Number of Fisher Scoring iterations: 4
```

```
> fit25<- glm(low ~ smoke + race2 + race3, data=lbw, family=binomial)  
> summary(fit25) #show fit on original dataset lbw
```

```
Call:
```

```
glm(formula = low ~ smoke + race2 + race3, family = binomial,  
     data = lbw)
```

```
Deviance Residuals:
```

```
      Min       1Q   Median       3Q      Max  
-1.3442 -0.8862 -0.5428  1.4964  1.9939
```

```
Coefficients:
```

```
      Estimate Std. Error z value Pr(>|z|)  
(Intercept) -1.8405      0.3529  -5.216 1.83e-07 ***  
smoke         1.1160      0.3692   3.023 0.00251 **  
race2         1.0841      0.4900   2.212 0.02693 *  
race3         1.1086      0.4003   2.769 0.00562 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 234.67 on 188 degrees of freedom  
Residual deviance: 219.97 on 185 degrees of freedom  
AIC: 227.97
```

Number of Fisher Scoring iterations: 4

SECTION 5.10:

```
#SECTION 5.10: HANDLING MISSING VALUES
library('foreign')
felon<- read.dta('felonx.dta') #read Stata format file
head(felon) #show some rows
str(felon) #show structure
nrow(felon) #show number of records
summary(felon) #descriptive stats on variables

fit26<- glm(hired ~ felony + gender + natamer + mo6 + educ2 + educ3 + educ4,
  data=felon, family=binomial)
summary(fit26)

fit27<- glm(hired ~ felony + gender + mo6 + educ2 + educ3 + educ4,
  data=felon, family=binomial)
summary(fit27)

nomiss<- !with(felon, is.na(hired)|is.na(felony)|is.na(gender)|
  is.na(natamer)|is.na(educ)|is.na(educ2)|is.na(educ3)|is.na(educ4))
table(nomiss)

fit28<- glm(hired ~ felony + gender + natamer + mo6 + educ2 + educ3 + educ4,
  data=felon, subset=nomiss, family=binomial) #note use of subset=
summary(fit28)
fit29<- glm(hired ~ felony + gender + mo6 + educ2 + educ3 + educ4,
  data=felon, subset=nomiss, family=binomial)
summary(fit29)
anova(fit29, fit28, test='Chisq') #LRT for natamer

fit30<- glm(hired ~ gender + natamer + mo6 + educ2 + educ3 + educ4,
  data=felon, subset=nomiss, family=binomial)
anova(fit30, fit28, test='Chisq') #LRT for felony

library('mice') #missing data imputation library for md.pattern(), mice(),
complete()
names(felon) #show variable names
md.pattern(felon[,1:4]) #show patterns for missing data in 1st 4 vars

library('Hmisc') #package for na.pattern() and impute()
na.pattern(felon[,1:4]) #show patterns for missing data in 1st 4 vars

fit31<- glm(hired ~ felony + gender + natamer, data=felon, family=binomial)
summary(fit31)

#no '.ice' equivalent in R that will give exactly the same imputed values as
Stata.

#simple imputation can be done by
felon2<- felon #make copy
felon2$gender<- factor(felon2$gender) #convert to factor
felon2$felony<- impute(felon2$felony) #impute NAs (most frequent)
felon2$gender<- impute(felon2$gender) #impute NAs
```

```

felon2$natamer<- impute(felon2$natamer) #impute NAs
na.pattern(felon2[,1:4]) #show no NAs left in these vars
fit32<- glm(hired ~ felony + gender + natamer, data=felon2, family=binomial)
summary(fit32)

#better, multiple imputation can be done via mice():
imp<- mice(felon[,1:4]) #do multiple imputation (default is 5 realizations)
for (iSet in 1:2) { #show results from 1st two imputation datasets
  fit<- glm(hired ~ felony + gender + natamer,
    data=complete(imp, iSet), family=binomial) #fit to iSet-th realization
  print(summary(fit))
}

felon3<- read.dta('imputedx.dta') #read Stata format file
head(felon3) #show some rows
str(felon3) #show structure
nrow(felon3) #show number of records
summary(felon3) #descriptive stats on variables
fit33<- glm(hired ~ felony + gender + natamer, data=felon3[1:733,],
family=binomial)
summary(fit33)

c(table(felon$felony), sum(is.na(felon$felony))) #frequency in original
c(table(felon3$felony[1:733]), sum(is.na(felon3$felony[1:733]))) #frequency in
imputed
c(table(felon$gender), sum(is.na(felon$gender))) #frequency in original
c(table(felon3$gender[1:733]), sum(is.na(felon3$gender[1:733]))) #frequency in
imputed
c(table(felon$natamer), sum(is.na(felon$natamer))) #frequency in original
c(table(felon3$natamer[1:733]), sum(is.na(felon3$natamer[1:733]))) #frequency
in imputed

```

OUTPUT: 5.10

```

> library('foreign')
> felon<- read.dta('felonx.dta') #read Stata format file
> head(felon) #show some rows
  hired  felony gender  natamer      mo6      educ educ1
1     0 No Felony    0 non Nat Amer  Worked >6Mo prev job Not HS Grad    1
2     0 No Felony   NA      <NA>    Worked >6Mo prev job Not HS Grad    1
3     0 No Felony    0 non Nat Amer      <NA> Not HS Grad    1
4     0 No Felony    0  Native Amer  Worked >6Mo prev job    HS Grad    0
5     0   Felony    0  Native Amer  Worked <6 Mo prev job    HS Grad    0
6     0 No Felony    0 non Nat Amer  Worked >6Mo prev job Not HS Grad    1
  educ2 educ3 educ4
1     0     0     0
2     0     0     0
3     0     0     0
4     1     0     0
5     1     0     0
6     0     0     0

> str(felon) #show structure
'data.frame':   733 obs. of  10 variables:
 $ hired  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ felony : Factor w/ 2 levels "No Felony","Felony": 1 1 1 1 2 1 1 1 1 1 ...
 $ gender : int  0 NA 0 0 0 0 0 0 0 0 ...
 $ natamer: Factor w/ 2 levels "non Nat Amer",...: 1 NA 1 2 2 1 1 1 1 2 ...
 $ mo6    : Factor w/ 2 levels "Worked <6 Mo prev job ",...: 2 2 NA 2 1 2 2 1 2
1 ...
 $ educ   : Factor w/ 4 levels "Not HS Grad",...: 1 1 1 2 2 1 3 2 2 2 ...

```

```

$ educ1 : int 1 1 1 0 0 1 0 0 0 0 ...
$ educ2 : int 0 0 0 1 1 0 0 1 1 1 ...
$ educ3 : int 0 0 0 0 0 0 1 0 0 0 ...
$ educ4 : int 0 0 0 0 0 0 0 0 0 0 ...
- attr(*, "datalabel")= chr ""
- attr(*, "time.stamp")= chr "29 Jun 2008 11:21"
- attr(*, "formats")= chr "%8.0g" "%9.0g" "%9.0g" "%12.0g" ...
- attr(*, "types")= int 251 251 251 251 251 251 251 251 251 251
- attr(*, "val.labels")= chr "" "felony" "" "natamer" ...
- attr(*, "var.labels")= chr "1=hired; 0=not hired" "1=felon; 0=not felon"
"1=male; 0=female" "1=Native Amer; 0= Not N. Amer" ...
- attr(*, "version")= int 8
- attr(*, "label.table")=List of 4
..$ mo6 : Named num 0 1
.. ..- attr(*, "names")= chr "Worked <6 Mo prev job " "Worked >6Mo prev job"
..$ natamer: Named num 0 1
.. ..- attr(*, "names")= chr "non Nat Amer" "Native Amer"
..$ felony : Named num 0 1
.. ..- attr(*, "names")= chr "No Felony" "Felony"
..$ educ : Named num 1 2 3 4
.. ..- attr(*, "names")= chr "Not HS Grad" "HS Grad" "JC" "2+ yrs College"

```

```
> nrow(felon) #show number of records
```

```
[1] 733
```

```
> summary(felon) #descriptive stats on variables
```

hired		felony		gender		natamer	
Min.	:0.0000	No Felony:	604	Min.	:0.0000	non Nat Amer:	640
1st Qu.:	0.0000	Felony	: 78	1st Qu.:	0.0000	Native Amer	: 81
Median	:0.0000	NA's	: 51	Median	:1.0000	NA's	: 12
Mean	:0.3124			Mean	:0.6585		
3rd Qu.:	1.0000			3rd Qu.:	1.0000		
Max.	:1.0000			Max.	:1.0000		
				NA's	:1.0000		

mo6		educ		educ1	
Worked <6 Mo prev job	:222	Not HS Grad	:220	Min.	: 0.0000
Worked >6Mo prev job	:394	HS Grad	:261	1st Qu.:	0.0000
NA's	:117	JC	:117	Median	: 0.0000
		2+ yrs College:	41	Mean	: 0.3443
		NA's	: 94	3rd Qu.:	1.0000
				Max.	: 1.0000
				NA's	:94.0000

educ2		educ3		educ4	
Min.	: 0.0000	Min.	: 0.0000	Min.	: 0.00000
1st Qu.:	0.0000	1st Qu.:	0.0000	1st Qu.:	0.00000
Median	: 0.0000	Median	: 0.0000	Median	: 0.00000
Mean	: 0.4085	Mean	: 0.1831	Mean	: 0.06416
3rd Qu.:	1.0000	3rd Qu.:	0.0000	3rd Qu.:	0.00000
Max.	: 1.0000	Max.	: 1.0000	Max.	: 1.00000
NA's	:94.0000	NA's	:94.0000	NA's	:94.00000

```

> fit26<- glm(hired ~ felony + gender + natamer + mo6 + educ2 + educ3 + educ4,
+ data=felon, family=binomial)
> summary(fit26)

```

```
Call:
```

```

glm(formula = hired ~ felony + gender + natamer + mo6 + educ2 +
educ3 + educ4, family = binomial, data = felon)

```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.0430	-0.8640	-0.6673	1.3180	2.1633

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.1747    0.2687  -4.371 1.24e-05 ***
felonyFelony    -0.7244    0.3653  -1.983 0.047343 *
gender          0.7212    0.2136   3.376 0.000736 ***
natamerNative Amer -0.8098    0.3629  -2.232 0.025648 *
mo6Worked >6Mo prev job  0.1288    0.2023   0.637 0.524274
educ2          -0.3395    0.2249  -1.509 0.131213
educ3          -0.4258    0.2850  -1.494 0.135203
educ4          -0.1195    0.3807  -0.314 0.753482
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 661.31 on 563 degrees of freedom
Residual deviance: 634.26 on 556 degrees of freedom
(169 observations deleted due to missingness)
AIC: 650.26

```

Number of Fisher Scoring iterations: 4

```

> fit27<- glm(hired ~ felony + gender + mo6 + educ2 + educ3 + educ4,
+   data=felon, family=binomial)
> summary(fit27)

```

```

Call:
glm(formula = hired ~ felony + gender + mo6 + educ2 + educ3 +
    educ4, family = binomial, data = felon)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0046  -0.8553  -0.6408   1.3606   2.2252

```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.27519    0.26531  -4.806 1.54e-06 ***
felonyFelony    -0.75485    0.36322  -2.078 0.037690 *
gender          0.69999    0.21249   3.294 0.000987 ***
mo6Worked >6Mo prev job  0.15421    0.20041   0.769 0.441634
educ2          -0.35791    0.22286  -1.606 0.108280
educ3          -0.39631    0.28281  -1.401 0.161126
educ4          -0.05307    0.37863  -0.140 0.888533
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 666.37 on 571 degrees of freedom
Residual deviance: 645.21 on 565 degrees of freedom
(161 observations deleted due to missingness)
AIC: 659.21

```

Number of Fisher Scoring iterations: 4

```

> nomiss<- !with(felon, is.na(hired)|is.na(felony)|is.na(gender)|
+   is.na(natamer)|is.na(educ)|is.na(educ2)|is.na(educ3)|is.na(educ4))
> table(nomiss)
nomiss
FALSE  TRUE
  111   622

```

```

> fit28<- glm(hired ~ felony + gender + natamer + mo6 + educ2 + educ3 + educ4,

```

```
+ data=felon, subset=nomiss, family=binomial) #note use of subset=
> summary(fit28)
```

Call:

```
glm(formula = hired ~ felony + gender + natamer + mo6 + educ2 +
     educ3 + educ4, family = binomial, data = felon, subset = nomiss)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0430	-0.8640	-0.6673	1.3180	2.1633

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.1747	0.2687	-4.371	1.24e-05	***
felonyFelony	-0.7244	0.3653	-1.983	0.047343	*
gender	0.7212	0.2136	3.376	0.000736	***
natamerNative Amer	-0.8098	0.3629	-2.232	0.025648	*
mo6Worked >6Mo prev job	0.1288	0.2023	0.637	0.524274	
educ2	-0.3395	0.2249	-1.509	0.131213	
educ3	-0.4258	0.2850	-1.494	0.135203	
educ4	-0.1195	0.3807	-0.314	0.753482	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 661.31 on 563 degrees of freedom
 Residual deviance: 634.26 on 556 degrees of freedom
 (58 observations deleted due to missingness)
 AIC: 650.26

Number of Fisher Scoring iterations: 4

```
> fit29<- glm(hired ~ felony + gender + mo6 + educ2 + educ3 + educ4,
+ data=felon, subset=nomiss, family=binomial)
> summary(fit29)
```

Call:

```
glm(formula = hired ~ felony + gender + mo6 + educ2 + educ3 +
     educ4, family = binomial, data = felon, subset = nomiss)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0132	-0.8612	-0.6482	1.3510	2.2162

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.27396	0.26586	-4.792	1.65e-06	***
felonyFelony	-0.75201	0.36390	-2.067	0.038777	*
gender	0.71393	0.21288	3.354	0.000797	***
mo6Worked >6Mo prev job	0.16067	0.20102	0.799	0.424112	
educ2	-0.34019	0.22371	-1.521	0.128333	
educ3	-0.40143	0.28345	-1.416	0.156713	
educ4	-0.06936	0.37893	-0.183	0.854774	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 661.31 on 563 degrees of freedom
 Residual deviance: 639.98 on 557 degrees of freedom
 (58 observations deleted due to missingness)
 AIC: 653.98

```

Number of Fisher Scoring iterations: 4

> anova(fit29, fit28, test='Chisq') #LRT for natamer
Analysis of Deviance Table

Model 1: hired ~ felony + gender + mo6 + educ2 + educ3 + educ4
Model 2: hired ~ felony + gender + natamer + mo6 + educ2 + educ3 + educ4
  Resid. Df Resid. Dev  Df Deviance P(>|Chi|)
1       557       639.98
2       556       634.26   1     5.72     0.02

> fit30<- glm(hired ~ gender + natamer + mo6 + educ2 + educ3 + educ4,
+   data=felon, subset=nomiss, family=binomial)
> anova(fit30, fit28, test='Chisq') #LRT for felony
Analysis of Deviance Table

Model 1: hired ~ gender + natamer + mo6 + educ2 + educ3 + educ4
Model 2: hired ~ felony + gender + natamer + mo6 + educ2 + educ3 + educ4
  Resid. Df Resid. Dev  Df Deviance P(>|Chi|)
1       557       638.70
2       556       634.26   1     4.44     0.04

> library('mice') #missing data imputation library for md.pattern(), mice(),
complete()
> names(felon) #show variable names
[1] "hired"  "felony" "gender" "natamer" "mo6"    "educ"   "educ1"
[8] "educ2"  "educ3"  "educ4"

> md.pattern(felon[,1:4]) #show patterns for missing data in 1st 4 vars
  hired gender natamer felony
671    1      1      1      1    0
 50     1      1      1      0    1
 10     1      1      0      1    1
  1     1      1      0      0    2
  1     1      0      0      1    2
    0      0      1     12    51 64

Warning message:
In data.frame.to.matrix(x) : NAs introduced by coercion

> library('Hmisc') #package for na.pattern() and impute()
> na.pattern(felon[,1:4]) #show patterns for missing data in 1st 4 vars
pattern
0000 0001 0011 0100 0101
671  10   1   50   1

> fit31<- glm(hired ~ felony + gender + natamer, data=felon, family=binomial)
> summary(fit31)

Call:
glm(formula = hired ~ felony + gender + natamer, family = binomial,
    data = felon)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9298 -0.9298 -0.6968  1.4471  2.1089

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.2919    0.1678  -7.698 1.38e-14 ***
felonyFelony   -0.5875    0.3133  -1.876 0.060723 .
gender         0.6771    0.1954   3.465 0.000531 ***
natamerNative Amer -0.8173    0.3294  -2.481 0.013100 *

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 790.24  on 670  degrees of freedom
Residual deviance: 766.94  on 667  degrees of freedom
(62 observations deleted due to missingness)
AIC: 774.94

Number of Fisher Scoring iterations: 4

> #no '.ice' equivalent in R that will give exactly the same imputed values as
Stata.
>
> #simple imputation can be done by
> felon2<- felon  #make copy
> felon2$gender<- factor(felon2$gender) #convert to factor
> felon2$felony<- impute(felon2$felony) #impute NAs (most frequent)
> felon2$gender<- impute(felon2$gender) #impute NAs
> felon2$natamer<- impute(felon2$natamer) #impute NAs
> na.pattern(felon2[,1:4]) #show no NAs left in these vars
pattern
0000
733
> fit32<- glm(hired ~ felony + gender + natamer, data=felon2, family=binomial)
> summary(fit32)

Call:
glm(formula = hired ~ felony + gender + natamer, family = binomial,
    data = felon2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.994  -0.994  -0.767   1.373   2.063

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.0731     0.1532  -7.005 2.47e-12 ***
felonyFelony       -0.8145     0.3098  -2.630 0.008549 **
gender1             0.6253     0.1791   3.491 0.000482 ***
natamerNative Amer -0.7398     0.3011  -2.457 0.014015 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 910.41  on 732  degrees of freedom
Residual deviance: 883.10  on 729  degrees of freedom
AIC: 891.1

Number of Fisher Scoring iterations: 4

> for (iSet in 1:2) { #show results from 1st two imputation datasets
+   fit<- glm(hired ~ felony + gender + natamer,
+     data=complete(imp, iSet), family=binomial) #fit to iSet-th realization
+   print(summary(fit))
+ }

Call:
glm(formula = hired ~ felony + gender + natamer, family = binomial,
    data = complete(imp, iSet))

```


Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9843	-0.9843	-0.7604	1.3837	2.0125

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.0929	0.1531	-7.137	9.57e-13	***
felonyFelony	-0.4168	0.2734	-1.525	0.127382	
gender	0.6200	0.1788	3.468	0.000525	***
natamerNative Amer	-0.7905	0.2997	-2.638	0.008349	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 910.41 on 732 degrees of freedom
Residual deviance: 887.60 on 729 degrees of freedom
AIC: 895.6

Number of Fisher Scoring iterations: 4

Call:

```
glm(formula = hired ~ felony + gender + natamer, family = binomial,  
     data = complete(imp, iSet))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9792	-0.9792	-0.7559	1.3894	2.1339

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.1065	0.1532	-7.221	5.14e-13	***
felonyFelony	-0.2661	0.2613	-1.018	0.30861	
gender	0.6207	0.1786	3.475	0.00051	***
natamerNative Amer	-0.7960	0.2995	-2.657	0.00787	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 910.41 on 732 degrees of freedom
Residual deviance: 888.78 on 729 degrees of freedom
AIC: 896.78

Number of Fisher Scoring iterations: 4

```
> felon3<- read.dta('imputedx.dta') #read Stata format file
```

```
> head(felon3) #show some rows
```

	hired	felony	gender	natamer	mo6	educ	educ1
1	0	No Felony	0	non Nat Amer	Worked >6Mo prev job	Not HS Grad	1
2	0	No Felony	1	non Nat Amer	Worked >6Mo prev job	Not HS Grad	1
3	0	No Felony	0	non Nat Amer	<NA>	Not HS Grad	1
4	0	No Felony	0	Native Amer	Worked >6Mo prev job	HS Grad	0
5	0	Felony	0	Native Amer	Worked <6 Mo prev job	HS Grad	0
6	0	No Felony	0	non Nat Amer	Worked >6Mo prev job	Not HS Grad	1

	educ2	educ3	educ4	_i	_j
1	0	0	0	1	1
2	0	0	0	2	1
3	0	0	0	3	1
4	1	0	0	4	1
5	1	0	0	5	1
6	0	0	0	6	1

```

> str(felon3) #show structure
'data.frame': 3665 obs. of 12 variables:
 $ hired : int 0 0 0 0 0 0 0 0 0 0 ...
 $ felony : Factor w/ 2 levels "No Felony","Felony": 1 1 1 1 2 1 1 1 1 1 ...
 $ gender : int 0 1 0 0 0 0 0 0 0 0 ...
 $ natamer: Factor w/ 2 levels "non Nat Amer",...: 1 1 1 2 2 1 1 1 1 2 ...
 $ mo6 : Factor w/ 2 levels "Worked <6 Mo prev job ",...: 2 2 NA 2 1 2 2 1 2
1 ...
 $ educ : Factor w/ 4 levels "Not HS Grad",...: 1 1 1 2 2 1 3 2 2 2 ...
 $ educ1 : int 1 1 1 0 0 1 0 0 0 0 ...
 $ educ2 : int 0 0 0 1 1 0 0 1 1 1 ...
 $ educ3 : int 0 0 0 0 0 0 1 0 0 0 ...
 $ educ4 : int 0 0 0 0 0 0 0 0 0 0 ...
 $ _i : int 1 2 3 4 5 6 7 8 9 10 ...
 $ _j : int 1 1 1 1 1 1 1 1 1 1 ...
 - attr(*, "datalabel")= chr ""
 - attr(*, "time.stamp")= chr "29 Jun 2008 11:21"
 - attr(*, "formats")= chr "%8.0g" "%9.0g" "%9.0g" "%12.0g" ...
 - attr(*, "types")= int 251 251 251 251 251 251 251 251 251 251 ...
 - attr(*, "val.labels")= chr "" "felony" "" "natamer" ...
 - attr(*, "var.labels")= chr "1=hired; 0=not hired" "felony imput. (51
values)" "gender imput. (1 values)" "natamer imput. (12 values)" ...
 - attr(*, "version")= int 8
 - attr(*, "label.table")=List of 4
 ..$ mo6 : Named num 0 1
 .. ..- attr(*, "names")= chr "Worked <6 Mo prev job " "Worked >6Mo prev job"
 ..$ natamer: Named num 0 1
 .. ..- attr(*, "names")= chr "non Nat Amer" "Native Amer"
 ..$ felony : Named num 0 1
 .. ..- attr(*, "names")= chr "No Felony" "Felony"
 ..$ educ : Named num 1 2 3 4
 .. ..- attr(*, "names")= chr "Not HS Grad" "HS Grad" "JC" "2+ yrs College"
> nrow(felon3) #show number of records
[1] 3665
> summary(felon3) #descriptive stats on variables
 hired felony gender natamer
Min. :0.0000 No Felony:3245 Min. :0.0000 non Nat Amer:3252
1st Qu.:0.0000 Felony : 420 1st Qu.:0.0000 Native Amer : 413
Median :0.0000 Median :1.0000
Mean :0.3124 Mean :0.6584
3rd Qu.:1.0000 3rd Qu.:1.0000
Max. :1.0000 Max. :1.0000

 mo6 educ educ1
Worked <6 Mo prev job :1110 Not HS Grad :1100 Min. : 0.0000
Worked >6Mo prev job :1970 HS Grad :1305 1st Qu.: 0.0000
NA's : 585 JC : 585 Median : 0.0000
2+ yrs College: 205 Mean : 0.3443
NA's : 470 3rd Qu.: 1.0000
Max. : 1.0000
NA's :470.0000

 educ2 educ3 educ4 _i
Min. : 0.0000 Min. : 0.0000 Min. : 0.00000 Min. : 1
1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.00000 1st Qu.:184
Median : 0.0000 Median : 0.0000 Median : 0.00000 Median :367
Mean : 0.4085 Mean : 0.1831 Mean : 0.06416 Mean :367
3rd Qu.: 1.0000 3rd Qu.: 0.0000 3rd Qu.: 0.00000 3rd Qu.:550
Max. : 1.0000 Max. : 1.0000 Max. : 1.00000 Max. :733
NA's :470.0000 NA's :470.0000 NA's :470.00000

 _j
Min. :1
1st Qu.:2
Median :3

```

```

Mean      :3
3rd Qu.:4
Max.      :5

> fit33<- glm(hired ~ felony + gender + natamer, data=felon3[1:733,],
family=binomial)
> summary(fit33)

Call:
glm(formula = hired ~ felony + gender + natamer, family = binomial,
    data = felon3[1:733, ])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9896  -0.9896  -0.7634   1.3776   1.9970

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.0837     0.1531  -7.080 1.45e-12 ***
felonyFelony    -0.6045     0.2889  -2.092 0.036423 *
gender           0.6244     0.1790   3.489 0.000484 ***
natamerNative Amer -0.7640     0.3003  -2.544 0.010960 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 910.41  on 732  degrees of freedom
Residual deviance: 885.75  on 729  degrees of freedom
AIC: 893.75

Number of Fisher Scoring iterations: 4

> c(table(felon$felony), sum(is.na(felon$felony))) #frequency in original
No Felony   Felony
    604      78      51

> c(table(felon3$felony[1:733]), sum(is.na(felon3$felony[1:733]))) #frequency
in imputed
No Felony   Felony
    652      81      0

> c(table(felon$gender), sum(is.na(felon$gender))) #frequency in original
 0  1
250 482  1

> c(table(felon3$gender[1:733]), sum(is.na(felon3$gender[1:733]))) #frequency
in imputed
 0  1
250 483  0

> c(table(felon$natamer), sum(is.na(felon$natamer))) #frequency in original
non Nat Amer  Native Amer
    640      81      12

> c(table(felon3$natamer[1:733]), sum(is.na(felon3$natamer[1:733]))) #frequency
in imputed
non Nat Amer  Native Amer
    651      82      0

```

SECTION 5.11:

```
#SECTION 5.11: MODELING UNCERTAIN RESPONSE
library('foreign')
medpar<- read.dta('medparx.dta')
head(medpar) #show some rows
str(medpar) #show structure
nrow(medpar) #show number of records
summary(medpar) #descriptive stats on variables

fit34<- glm(died ~ los + type2 + type3, data=medpar, family=binomial)
summary(fit34)
exp(coef(fit34)) #show odds ratios
table(medpar$died, fitted(fit34, type='response')>=0.5) #show confusion matrix
(28/513) #sensitivity
(958/982) #specificity

library('PresenceAbsence') #package for sensitivity/specificity
d<- cbind(1:nrow(medpar), medpar$died, fitted(fit34, type='response'))
#construct data
confmat<- cmx(d, threshold = 0.5, which.model = 1) #get confusion matrix
confmat #show confusion matrix
sensitivity(confmat, st.dev = TRUE) #calc. sensitivity
specificity(confmat, st.dev = TRUE) #calc. specificity

#Magder-Hughes methodology not yet implemented in R
```

OUTPUT: 5.11

```
> #SECTION 5.11: MODELING UNCERTAIN RESPONSE
> library('foreign')
> medpar<- read.dta('medparx.dta')
> head(medpar) #show some rows
  provnum died white hmo los age80   age      type type1 type2 type3
1  030001    0     1   0   4     0 65-69 Elective     1     0     0
2  030001    0     1   1   9     0 65-69 Elective     1     0     0
3  030001    1     1   1   3     1 80-84 Elective     1     0     0
4  030001    0     1   0   9     0 75-79 Elective     1     0     0
5  030001    1     1   0   1     1 80-84 Elective     1     0     0
6  030001    1     1   0   4     0 70-74 Elective     1     0     0

> str(medpar) #show structure
'data.frame':  1495 obs. of  11 variables:
 $ provnum: chr  "030001" "030001" "030001" "030001" ...
 $ died   : int   0 0 1 0 1 1 1 1 0 0 ...
 $ white  : int   1 1 1 1 1 1 1 1 1 1 ...
 $ hmo    : int   0 1 1 0 0 0 0 0 0 0 ...
 $ los    : int   4 9 3 9 1 4 10 3 5 6 ...
 $ age80  : int   0 0 1 0 1 0 1 1 0 0 ...
 $ age    : Factor w/ 9 levels "<55","55-59",...: 4 4 7 6 7 5 8 7 4 4 ...
 $ type   : Factor w/ 3 levels "Elective","Urgent",...: 1 1 1 1 1 1 1 2 1 1 ...
 $ type1  : int   1 1 1 1 1 1 1 0 1 1 ...
 $ type2  : int   0 0 0 0 0 0 0 0 1 0 0 ...
 $ type3  : int   0 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "datalabel")= chr ""
 - attr(*, "time.stamp")= chr "29 Jun 2008 11:22"
 - attr(*, "formats")= chr  "%9s" "%9.0g" "%9.0g" "%9.0g" ...
 - attr(*, "types")= int   6 251 251 251 252 251 251 251 251 251 ...
```

```

- attr(*, "val.labels")= chr  "" "" "" "" ...
- attr(*, "var.labels")= chr  "Provider number" "1=died in hospital; 0=not
die" "1=white; 0=non-white" "HMO/readmit" ...
- attr(*, "version")= int 8
- attr(*, "label.table")=List of 2
..$ type: Named num  1 2 3
.. ..- attr(*, "names")= chr  "Elective" "Urgent" "Emergency"
..$ age : Named num  1 2 3 4 5 6 7 8 9
.. ..- attr(*, "names")= chr  "<55" "55-59" "60-64" "65-69" ...

```

```

> nrow(medpar) #show number of records
[1] 1495

```

```

> summary(medpar) #descriptive stats on variables
      provnum      died      white      hmo
Length:1495    Min.   :0.0000    Min.   :0.000    Min.   :0.0000
Class :character 1st Qu.:0.0000    1st Qu.:1.000    1st Qu.:0.0000
Mode  :character Median :0.0000    Median :1.000    Median :0.0000
      Mean   :0.3431    Mean   :0.915    Mean   :0.1599
      3rd Qu.:1.0000    3rd Qu.:1.000    3rd Qu.:0.0000
      Max.   :1.0000    Max.   :1.000    Max.   :1.0000

      los      age80      age      type
Min.   : 1.000    Min.   :0.0000    75-79 :328    Elective :1134
1st Qu.: 4.000    1st Qu.:0.0000    70-74 :317    Urgent  : 265
Median : 8.000    Median :0.0000    65-69 :291    Emergency: 96
Mean   : 9.854    Mean   :0.2207    80-84 :191
3rd Qu.:13.000    3rd Qu.:0.0000    60-64 :163
Max.   :116.000    Max.   :1.0000    85-89 : 93
                        (Other):112

      type1      type2      type3
Min.   :0.0000    Min.   :0.0000    Min.   :0.00000
1st Qu.:1.0000    1st Qu.:0.0000    1st Qu.:0.00000
Median :1.0000    Median :0.0000    Median :0.00000
Mean   :0.7585    Mean   :0.1773    Mean   :0.06421
3rd Qu.:1.0000    3rd Qu.:0.0000    3rd Qu.:0.00000
Max.   :1.0000    Max.   :1.0000    Max.   :1.00000

```

```

> fit34<- glm(died ~ los + type2 + type3, data=medpar, family=binomial)
> summary(fit34)

```

```

Call:
glm(formula = died ~ los + type2 + type3, family = binomial,
    data = medpar)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3779  -0.9294  -0.8330   1.3801   2.4450

```

```

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.427167   0.090525  -4.719 2.37e-06 ***
los          -0.037747   0.007784  -4.849 1.24e-06 ***
type2         0.397794   0.143305   2.776 0.00551 **
type3         0.924904   0.228301   4.051 5.09e-05 ***
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

(Dispersion parameter for binomial family taken to be 1)

```

```

    Null deviance: 1922.9  on 1494  degrees of freedom
Residual deviance: 1883.4  on 1491  degrees of freedom
AIC: 1891.4

```

Number of Fisher Scoring iterations: 4

```
> exp(coef(fit34)) #show odds ratios
(Intercept)      los      type2      type3
  0.6523544    0.9629564    1.4885380    2.5216264

> table(medpar$died, fitted(fit34, type='response')>=0.5) #show confusion
matrix

      FALSE TRUE
0     958    24
1     485    28

> (28/513) #sensitivity
[1] 0.0545809

> (958/982) #specificity
[1] 0.97556

> library('PresenceAbsence') #package for sensitivity/specificity
> d<- cbind(1:nrow(medpar), medpar$died, fitted(fit34, type='response'))
#construct data
> confmat<- cmx(d, threshold = 0.5, which.model = 1) #get confusion matrix
> confmat #show confusion matrix
      observed
predicted  1    0
          1  28  24
          0 485 958

> sensitivity(confmat, st.dev = TRUE) #calc. sensitivity
      sensitivity sensitivity.sd
1    0.0545809    0.01003916

> specificity(confmat, st.dev = TRUE) #calc. specificity
      specificity specificity.sd
1    0.97556    0.004929949
```

SECTION 5.12:

```
#SECTION 5.12: CONSTRAINING COEFFICIENTS
fit35<- glm(died ~ white + hmo + los + age80, data=medpar, family=binomial)
summary(fit35)

coefhmo<- coef(fit35)['hmo']*medpar$hmo #effect of hmo in fit
fit36<- glm(died ~ white + los + age80 + offset(coefhmo),
  data=medpar, family=binomial) #fix effect of hmo via offset()
summary(fit36)

coefage80<- coef(fit36)['age80']*medpar$age80 #effect of age80 in fit
coeflos<- coef(fit36)['los']*medpar$los #effect of age80 in fit
constr<- coefhmo + coefage80 + coeflos #combined effects
fit37<- glm(died ~ white + offset(constr), data=medpar,
  family=binomial) #fix effects of hmo, age80, los via offset()
summary(fit37)

fit38<- glm(died ~ white + type2 + type3 + offset(constr), data=medpar,
  family=binomial) #fix effects of hmo, age80, los via offset()
```

```
summary(fit38)
```

```
#There is no equivalent to constraint() in R other than offset().
```

OUTPUT: 5.12

```
> #SECTION 5.12: CONSTRAINING COEFFICIENTS
```

```
> fit35<- glm(died ~ white + hmo + los + age80, data=medpar, family=binomial)
```

```
> summary(fit35)
```

```
Call:
```

```
glm(formula = died ~ white + hmo + los + age80, family = binomial,  
     data = medpar)
```

```
Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-1.2189	-0.9126	-0.8319	1.3932	2.3796

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.705414	0.216920	-3.252	0.001146	**
white	0.206459	0.208044	0.992	0.321011	
hmo	-0.013947	0.151015	-0.092	0.926414	
los	-0.029567	0.007815	-3.783	0.000155	***
age80	0.625582	0.128294	4.876	1.08e-06	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 1922.9 on 1494 degrees of freedom  
Residual deviance: 1879.4 on 1490 degrees of freedom  
AIC: 1889.4
```

```
Number of Fisher Scoring iterations: 4
```

```
> coefhmo<- coef(fit35)['hmo']*medpar$hmo #effect of hmo in fit
```

```
> fit36<- glm(died ~ white + los + age80 + offset(coefhmo),  
+ data=medpar, family=binomial) #fix effect of hmo via offset()  
> summary(fit36)
```

```
Call:
```

```
glm(formula = died ~ white + los + age80 + offset(coefhmo), family = binomial,  
     data = medpar)
```

```
Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-1.2189	-0.9126	-0.8319	1.3932	2.3796

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.705414	0.216081	-3.265	0.001096	**
white	0.206459	0.207772	0.994	0.320378	
los	-0.029567	0.007807	-3.787	0.000152	***
age80	0.625582	0.128173	4.881	1.06e-06	***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 1922.9 on 1494 degrees of freedom
```

Residual deviance: 1879.4 on 1491 degrees of freedom
AIC: 1887.4

Number of Fisher Scoring iterations: 4

```
> coefage80<- coef(fit36)['age80']*medpar$age80 #effect of age80 in fit
> coeflos<- coef(fit36)['los']*medpar$los #effect of age80 in fit
> constr<- coefhmo + coefage80 + coeflos #combined effects
> fit37<- glm(died ~ white + offset(constr), data=medpar,
+ family=binomial) #fix effects of hmo, age80, los via offset()
> summary(fit37)
```

Call:

```
glm(formula = died ~ white + offset(constr), family = binomial,
    data = medpar)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.2189	-0.9126	-0.8319	1.3932	2.3796

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.7054	0.1991	-3.543	0.000396 ***
white	0.2065	0.2073	0.996	0.319203

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1880.4 on 1494 degrees of freedom
Residual deviance: 1879.4 on 1493 degrees of freedom
AIC: 1883.4

Number of Fisher Scoring iterations: 4

```
> fit38<- glm(died ~ white + type2 + type3 + offset(constr), data=medpar,
+ family=binomial) #fix effects of hmo, age80, los via offset()
> summary(fit38)
```

Call:

```
glm(formula = died ~ white + type2 + type3 + offset(constr),
    family = binomial, data = medpar)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5005	-0.8838	-0.7973	1.2939	2.1078

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.8992	0.2064	-4.356	1.32e-05 ***
white	0.2710	0.2092	1.296	0.19503
type2	0.4176	0.1438	2.904	0.00369 **
type3	0.9136	0.2228	4.101	4.12e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1880.4 on 1494 degrees of freedom
Residual deviance: 1857.3 on 1491 degrees of freedom
AIC: 1865.3

Number of Fisher Scoring iterations: 4


```
> #There is no equivalent to constraint() in R other than offset().
```

Chapter 6

SECTION 6.0: <INTRODUCTION>

```
# SECTION 6.0: <INTRODUCTION>
library('foreign')
medpar<- read.dta('medparx.dta')
summary(medpar)

with(medpar, table(age, type)) #show classification
library('gmodels') #for CrossTable
CrossTable(medpar$age, medpar$type, dnn=c('Age','Type'), prop.r=FALSE,
  prop.c=FALSE, prop.t=FALSE, prop.chisq=FALSE) #better table format
```

OUTPUT: 6.0

```
> library('foreign')
> medpar<- read.dta('medparx.dta')
> summary(medpar)
```

provnum	died	white	hmo
Length:1495	Min. :0.0000	Min. :0.000	Min. :0.0000
Class :character	1st Qu.:0.0000	1st Qu.:1.000	1st Qu.:0.0000
Mode :character	Median :0.0000	Median :1.000	Median :0.0000
	Mean :0.3431	Mean :0.915	Mean :0.1599
	3rd Qu.:1.0000	3rd Qu.:1.000	3rd Qu.:0.0000
	Max. :1.0000	Max. :1.000	Max. :1.0000

los	age80	age	type
Min. : 1.000	Min. :0.0000	75-79 :328	Elective :1134
1st Qu.: 4.000	1st Qu.:0.0000	70-74 :317	Urgent : 265
Median : 8.000	Median :0.0000	65-69 :291	Emergency: 96
Mean : 9.854	Mean :0.2207	80-84 :191	
3rd Qu.:13.000	3rd Qu.:0.0000	60-64 :163	
Max. :116.000	Max. :1.0000	85-89 : 93	
		(Other):112	

type1	type2	type3
Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.:1.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median :1.0000	Median :0.0000	Median :0.00000
Mean :0.7585	Mean :0.1773	Mean :0.06421
3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.0000	Max. :1.00000


```
> with(medpar, table(age, type)) #show classification
```

age	Elective	Urgent	Emergency
<55	4	2	0
55-59	46	9	5
60-64	133	22	8
65-69	206	63	22
70-74	240	55	22
75-79	247	60	21

```

80-84      148      31      12
85-89       73      17       3
90+        37       6       3

> library('gtools') #for CrossTable
> CrossTable(medpar$age, medpar$type, dnn=c('Age','Type'), prop.r=FALSE,
+   prop.c=FALSE, prop.t=FALSE, prop.chisq=FALSE) #better table format

```

```

      Cell Contents
|-----|
|              N |
|-----|

```

Total Observations in Table: 1495

Age	Type Elective	Urgent	Emergency	Row Total
<55	4	2	0	6
55-59	46	9	5	60
60-64	133	22	8	163
65-69	206	63	22	291
70-74	240	55	22	317
75-79	247	60	21	328
80-84	148	31	12	191
85-89	73	17	3	93
90+	37	6	3	46
Column Total	1134	265	96	1495

SECTION 6.1:

```

#SECTION 6.1: BINARY x BINARY INTERACTIONS
titanic<- read.dta('titanicx.dta') #get data from Stata file
summary(titanic)
tpass<- tpass<- titanic[-which(titanic$class=='crew'),] #drop crew from
dataset
with(tpass, table(survived))
with(tpass, table(sex))
with(tpass, table(age))
with(tpass, table(class))

fit6a<- glm(survived ~ sex + age, data=tpass, family=binomial(link=logit))
summary(fit6a)
exp(coef(fit6a)) #odds ratios

df<- data.frame(sex=c('women', 'women', 'man', 'man'),

```

```

    age=c('child', 'adults', 'child', 'adults'))
df

predict(fit6a, newdata=df, type='link') #show predictions
coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*0 #prediction for
sex='women', age='child'
coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*1 #prediction for
sex='women', age='adults'
coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*0 #prediction for
sex='man', age='child'
coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*1 #prediction for
sex='man', age='adults'

exp(predict(fit6a, newdata=df, type='link')) #predicted odds ratios
exp(coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*0) #OR for sex='women',
age='child'
exp(coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*1) #OR for sex='women',
age='adults'
exp(coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*0) #OR for sex='man',
age='child'
exp(coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*1) #OR for sex='man',
age='adults'

predict(fit6a, newdata=df, type='response') #show predictions in probabilities
1/(1+exp(1.4326))
0.1926889/(1 - 0.1926889)

fit6b<- glm(survived ~ sex + age + sex:age, data=tpass, family=binomial) #with
interaction
summary(fit6b)
exp(coef(fit6b)) #odds ratios
exp(predict(fit6b, newdata=df, type='link')) #predicted odds ratios
2.7924528/1.6470588 #OR for adults:child given sex=women
0.2215478/0.8285714 #OR for adults:child given sex=man
0.8285714/1.6470588 #OR for man:women given age=child
0.2215478/2.7924528 #OR for man:women given age=adults
0.07933806/0.5030612 #man:women|age=child / man:women|age=adults

#no plotgr3 in R

class2<- as.numeric(tpass$class=='2nd class') #indicator variable
class3<- as.numeric(tpass$class=='3rd class') #indicator variable
fit6c<- glm(survived ~ sex + age + sex:age + class2 + class3,
    data=tpass, family=binomial)
summary(fit6c)

fit6d<- glm(survived ~ sex*age + class, data=tpass, family=binomial)
summary(fit6d)
exp(coef(fit6d)) #ORs
exp(coef(fit6c)) #ORs

```

OUTPUT: 6.1

```

> #SECTION 6.1: BINARY x BINARY INTERACTIONS
> titanic<- read.dta('titanicx.dta') #get data from Stata file
> summary(titanic)
      class      age      sex      survived
1st class:325  child : 109  women: 470   no :1490
2nd class:285  adults:2092  man  :1731   yes: 711
3rd class:706
crew          :885

```

```

> tpass<- tpass<- titanic[-which(titanic$class=='crew'),] #drop crew from
dataset
> with(tpass, table(survived))
survived
  no yes
817 499

> with(tpass, table(sex))
sex
women  man
 447   869

> with(tpass, table(age))
age
child adults
 109   1207

> with(tpass, table(class))
class
1st class 2nd class 3rd class      crew
    325      285      706         0

> fit6a<- glm(survived ~ sex + age, data=tpass, family=binomial(link=logit))
> summary(fit6a)

Call:
glm(formula = survived ~ sex + age, family = binomial(link = logit),
    data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8664  -0.6543  -0.6543   0.8214   1.8148

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.5490     0.2408   6.432 1.26e-10 ***
sexman       -2.3458     0.1361 -17.241 < 2e-16 ***
ageadults    -0.6358     0.2332  -2.726  0.00641 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1391.6  on 1313  degrees of freedom
AIC: 1397.6

Number of Fisher Scoring iterations: 4

> exp(coef(fit6a)) #odds ratios
(Intercept)      sexman  ageadults
 4.70670241  0.09577172  0.52949473

> df<- data.frame(sex=c('women', 'women', 'man', 'man'),
+   age=c('child', 'adults', 'child', 'adults'))
> df
   sex   age
1 women child
2 women adults
3  man  child
4  man  adults
> predict(fit6a, newdata=df, type='link') #show predictions

```

```

      1      2      3      4
1.5489875 0.9131555 -0.7968003 -1.4326324

> coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*0 #prediction for
sex='women', age='child'
(Intercept)
  1.548988

> coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*1 #prediction for
sex='women', age='adults'
(Intercept)
  0.9131555

> coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*0 #prediction for
sex='man', age='child'
(Intercept)
 -0.7968003

> coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*1 #prediction for
sex='man', age='adults'
(Intercept)
 -1.432632

> exp(predict(fit6a, newdata=df, type='link')) #predicted odds ratios
      1      2      3      4
4.7067024 2.4921741 0.4507690 0.2386798

> exp(coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*0) #OR for
sex='women', age='child'
(Intercept)
  4.706702

> exp(coef(fit6a)[1] + coef(fit6a)[2]*0 + coef(fit6a)[3]*1) #OR for
sex='women', age='adults'
(Intercept)
  2.492174

> exp(coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*0) #OR for sex='man',
age='child'
(Intercept)
  0.450769

> exp(coef(fit6a)[1] + coef(fit6a)[2]*1 + coef(fit6a)[3]*1) #OR for sex='man',
age='adults'
(Intercept)
  0.2386798

> predict(fit6a, newdata=df, type='response') #show predictions in
probabilities
      1      2      3      4
0.8247675 0.7136454 0.3107104 0.1926889

> 1/(1+exp(1.4326))
[1] 0.1926939

> 0.1926889/(1 - 0.1926889)
[1] 0.2386799

> fit6b<- glm(survived ~ sex + age + sex:age, data=tpass, family=binomial)
#with interaction
> summary(fit6b)

```

Call:

```

glm(formula = survived ~ sex + age + sex:age, family = binomial,
    data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6328  -0.6326  -0.6326   0.7824   1.8478

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.4990     0.3075   1.623   0.1046
sexman        -0.6870     0.3970  -1.731   0.0835 .
ageadults      0.5279     0.3276   1.611   0.1071
sexman:ageadults -1.8470     0.4228  -4.368 1.25e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1373.9  on 1312  degrees of freedom
AIC: 1381.9

Number of Fisher Scoring iterations: 4

> exp(coef(fit6b)) #odds ratios
              (Intercept)              sexman              ageadults sexman:ageadults
              1.6470588              0.5030612              1.6954178              0.1577106

> exp(predict(fit6b, newdata=df, type='link')) #predicted odds ratios
              1              2              3              4
1.6470588 2.7924528 0.8285714 0.2215478

> 2.7924528/1.6470588 #OR for adults:child given sex=women
[1] 1.695418

> 0.2215478/0.8285714 #OR for adults:child given sex=man
[1] 0.2673853

> 0.8285714/1.6470588 #OR for man:women given age=child
[1] 0.5030612

> 0.2215478/2.7924528 #OR for man:women given age=adults
[1] 0.07933806

> 0.07933806/0.5030612 #man:women|age=child / man:women|age=adults
[1] 0.1577106

> #no plotgr3 in R
>
> class2<- as.numeric(tpass$class=='2nd class') #indicator variable
> class3<- as.numeric(tpass$class=='3rd class') #indicator variable
> fit6c<- glm(survived ~ sex + age + sex:age + class2 + class3,
+   data=tpass, family=binomial)
> summary(fit6c)

Call:
glm(formula = survived ~ sex + age + sex:age + class2 + class3,
    family = binomial, data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1266  -0.6483  -0.4528   0.7502   2.1581

```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.0626     0.3526   5.850 4.91e-09 ***
sexman           -0.7149     0.4061  -1.760  0.0784 .
ageadults         0.0886     0.3360   0.264  0.7920
class2            -1.0273     0.1990  -5.163 2.44e-07 ***
class3            -1.8006     0.1753 -10.269 < 2e-16 ***
sexman:ageadults -1.8621     0.4350  -4.281 1.86e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1259.0  on 1310  degrees of freedom
AIC: 1271.0

Number of Fisher Scoring iterations: 4

> fit6d<- glm(survived ~ sex*age + class, data=tpass, family=binomial)
> summary(fit6d)

Call:
glm(formula = survived ~ sex * age + class, family = binomial,
    data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1266  -0.6483  -0.4528   0.7502   2.1581

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.0626     0.3526   5.850 4.91e-09 ***
sexman           -0.7149     0.4061  -1.760  0.0784 .
ageadults         0.0886     0.3360   0.264  0.7920
class2nd class    -1.0273     0.1990  -5.163 2.44e-07 ***
class3rd class    -1.8006     0.1753 -10.269 < 2e-16 ***
sexman:ageadults -1.8621     0.4350  -4.281 1.86e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1259.0  on 1310  degrees of freedom
AIC: 1271.0

Number of Fisher Scoring iterations: 4

> exp(coef(fit6d)) #ORs
              (Intercept)      sexman      ageadults      class2nd class
              7.8666046      0.4892594      1.0926414      0.3579866
class3rd class sexman:ageadults
              0.1652041      0.1553494

> exp(coef(fit6c)) #ORs
              (Intercept)      sexman      ageadults      class2
              7.8666046      0.4892594      1.0926414      0.3579866
class3 sexman:ageadults
              0.1652041      0.1553494

```

SECTION 6.2:

```
#SECTION 6.2: BINARY x CATEGORICAL INTERACTIONS
fit6e<- glm(survived ~ age + sex + class2 + class3, data=tpass,
family=binomial)
summary(fit6e)
exp(coef(fit6e)) #ORs

agecl2<- as.numeric(tpass$age=='adults')*class2 #indicator variable
agecl3<- as.numeric(tpass$age=='adults')*class3 #indicator variable
fit6f<- glm(survived ~ age + sex + class2 + class3 + agecl2 + agecl3,
data=tpass, family=binomial)
summary(fit6f)

sexcl2<- as.numeric(tpass$sex=='man')*class2 #indicator variable
sexcl3<- as.numeric(tpass$sex=='man')*class3 #indicator variable
fit6g<- glm(survived ~ age + sex + class2 + class3 + sexcl2 + sexcl3,
data=tpass, family=binomial)
summary(fit6g)

fit6h<- glm(survived ~ age + sex*class, data=tpass, family=binomial)
summary(fit6h)

exp(coef(fit6g)) #ORs

#No plostgr3 in R

fit6i<- glm(survived ~ sex + class2 + class3, data=tpass, family=binomial)
summary(fit6i)
exp(coef(fit6i)) #ORs
fitcases<- matrix(c(1,0,0,0, 1,0,1,0, 1,0,0,1, 1,1,0,0, 1,1,1,0,
1,1,0,1), ncol=4, byrow=TRUE) #cases for sex x class
adjust<- matrix(fitcases%*%coef(fit6i), ncol=3, byrow=TRUE) #link version
colnames(adjust)<- c('1st class', '2nd class', '3rd class')
rownames(adjust)<- c('women', 'man')
exp(adjust) #Odds version
0.136942/1.46764 #OR for sex | 3rd class
0.276349/0.71128 #OR for 2nd:1st | man
0.136942/0.71128 #OR for 3rd:1st | women

fit6j<- glm(survived ~ sex*class, data=tpass, family=binomial)
summary(fit6j)
exp(coef(fit6j)) #ORs
fitcases2<- matrix(c(1,0,0,0,0,0, 1,0,1,0,0,0, 1,0,0,1,0,0,
1,1,0,0,0,0, 1,1,1,0,1,0, 1,1,0,1,0,1), ncol=6, byrow=TRUE) #cases for sex x
class
adjust2<- matrix(fitcases2%*%coef(fit6j), ncol=3, byrow=TRUE) #link version
colnames(adjust2)<- c('1st class', '2nd class', '3rd class')
rownames(adjust2)<- c('women', 'man')
exp(adjust2) #Odds version
0.5254237/35.25 #OR for man:women | 1st class
7.1538462/35.25 #OR for 2nd:1st | women
0.8490566/35.25 #OR for 3rd:1st | women

0.162338/7.15385
0.525424/35.25
0.0226924/0.01490565
```



```
0.208531/0.849057
0.525424/35.25
0.24560306/0.01490565
```

OUTPUT: 6.2

```
> #SECTION 6.2: BINARY x CATEGORICAL INTERACTIONS
> fit6e<- glm(survived ~ age + sex + class2 + class3, data=tpass,
family=binomial)
> summary(fit6e)

Call:
glm(formula = survived ~ age + sex + class2 + class3, family = binomial,
    data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0652  -0.6718  -0.4740   0.7930   2.1175

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.0619     0.2980  10.275 < 2e-16 ***
ageadults    -1.0556     0.2427  -4.350 1.36e-05 ***
sexman       -2.3695     0.1453 -16.313 < 2e-16 ***
class2       -1.0106     0.1949  -5.184 2.17e-07 ***
class3       -1.7664     0.1707 -10.347 < 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: 1746.8  on 1315  degrees of freedom
Residual deviance: 1276.2  on 1311  degrees of freedom
AIC: 1286.2

Number of Fisher Scoring iterations: 4

> exp(coef(fit6e)) #ORs
(Intercept)  ageadults  sexman  class2  class3
21.36772532  0.34798085  0.09353076  0.36401585  0.17095216

> agecl2<- as.numeric(tpass$age=='adults')*class2 #indicator variable
> agecl3<- as.numeric(tpass$age=='adults')*class3 #indicator variable
> fit6f<- glm(survived ~ age + sex + class2 + class3 + agecl2 + agecl3,
+ data=tpass, family=binomial)
> summary(fit6f)

Call:
glm(formula = survived ~ age + sex + class2 + class3 + agecl2 +
    agecl3, family = binomial, data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0666  -0.6149  -0.4946   0.5875   2.0792

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  18.6470     900.1824   0.021   0.983
ageadults    -16.6373     900.1824  -0.018   0.985
sexman       -2.4146     0.1465  -16.483 <2e-16 ***
class2       -0.6871     990.4066  -0.001   0.999
```

```

class3      -18.0268    900.1825    -0.020    0.984
agecl2      -0.4775    990.4067   -0.000482    1.000
agecl3      16.3925    900.1825     0.018    0.985
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1240.7  on 1309  degrees of freedom
AIC: 1254.7

Number of Fisher Scoring iterations: 15

> sexcl2<- as.numeric(tpass$sex=='man')*class2 #indicator variable
> sexcl3<- as.numeric(tpass$sex=='man')*class3 #indicator variable
> fit6g<- glm(survived ~ age + sex + class2 + class3 + sexcl2 + sexcl3,
+   data=tpass, family=binomial)
> summary(fit6g)

Call:
glm(formula = survived ~ age + sex + class2 + class3 + sexcl2 +
    sexcl3, family = binomial, data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6781  -0.5798  -0.5798   0.5336   2.0228

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.6115     0.5567   8.283 < 2e-16 ***
ageadults     -1.0537     0.2304  -4.573 4.81e-06 ***
sexman         -4.2331     0.5310  -7.972 1.56e-15 ***
class2         -1.6806     0.5878  -2.859  0.00425 **
class3         -3.8854     0.5287  -7.350 1.99e-13 ***
sexcl2          0.4483     0.6460   0.694  0.48772
sexcl3          2.8625     0.5633   5.082 3.73e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1211.3  on 1309  degrees of freedom
AIC: 1225.3

Number of Fisher Scoring iterations: 6

> fit6h<- glm(survived ~ age + sex*class, data=tpass, family=binomial)
> summary(fit6h)

Call:
glm(formula = survived ~ age + sex * class, family = binomial,
    data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6781  -0.5798  -0.5798   0.5336   2.0228

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.6115     0.5567   8.283 < 2e-16 ***
ageadults     -1.0537     0.2304  -4.573 4.81e-06 ***

```

```

sexman          -4.2331      0.5310   -7.972 1.56e-15 ***
class2nd class  -1.6806      0.5878   -2.859 0.00425 **
class3rd class  -3.8854      0.5287   -7.350 1.99e-13 ***
sexman:class2nd class  0.4483      0.6460    0.694 0.48772
sexman:class3rd class  2.8625      0.5633    5.082 3.73e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1211.3  on 1309  degrees of freedom
AIC: 1225.3

Number of Fisher Scoring iterations: 6

> exp(coef(fit6g)) #ORs
(Intercept)    ageadults      sexman      class2      class3      sexcl2
100.63949838    0.34865800    0.01450731    0.18625597    0.02053872    1.56566572
      sexcl3
17.50603017

> #No plostgr3 in R

> fit6i<- glm(survived ~ sex + class2 + class3, data=tpass, family=binomial)
> summary(fit6i)

Call:
glm(formula = survived ~ sex + class2 + class3, family = binomial,
    data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0758  -0.6986  -0.5066   0.7628   2.0574

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.0312     0.1691  12.013 < 2e-16 ***
sexman        -2.3719     0.1438 -16.494 < 2e-16 ***
class2        -0.9454     0.1933  -4.892 9.99e-07 ***
class3        -1.6475     0.1670  -9.867 < 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: 1746.8  on 1315  degrees of freedom
Residual deviance: 1295.1  on 1312  degrees of freedom
AIC: 1303.1

Number of Fisher Scoring iterations: 4

> exp(coef(fit6i)) #ORs
(Intercept)      sexman      class2      class3
7.62295652  0.09330763  0.38852334  0.19252861

> fitcases<- matrix(c(1,0,0,0, 1,0,1,0, 1,0,0,1, 1,1,0,0, 1,1,1,0,
+ 1,1,0,1), ncol=4, byrow=TRUE) #cases for sex x class
> adjust<- matrix(fitcases%%coef(fit6i), ncol=3, byrow=TRUE) #link version
> colnames(adjust)<- c('1st class', '2nd class', '3rd class')
> rownames(adjust)<- c('women', 'man')
> exp(adjust) #Odds version
      1st class 2nd class 3rd class

```

```
women 7.622957 2.9616965 1.4676372
man    0.711280 0.2763489 0.1369418

> 0.136942/1.46764 #OR for sex | 3rd class
[1] 0.09330762

> 0.276349/0.71128 #OR for 2nd:1st | man
[1] 0.3885235

> 0.136942/0.71128 #OR for 3rd:1st | women
[1] 0.1925290

> fit6j<- glm(survived ~ sex*class, data=tpass, family=binomial)
> summary(fit6j)

Call:
glm(formula = survived ~ sex * class, family = binomial, data = tpass)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6797  -0.6155  -0.6155   0.5115   1.9842

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.5625     0.5070   7.026 2.13e-12 ***
sexman           -4.2060     0.5307  -7.925 2.29e-15 ***
class2nd class   -1.5948     0.5872  -2.716  0.0066 **
class3rd class   -3.7261     0.5269  -7.072 1.53e-12 ***
sexman:class2nd class  0.4203     0.6449   0.652  0.5146
sexman:class3rd class  2.8020     0.5621   4.985 6.21e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1746.8  on 1315  degrees of freedom
Residual deviance: 1231.6  on 1310  degrees of freedom
AIC: 1243.6

Number of Fisher Scoring iterations: 6

> exp(coef(fit6j)) #ORs
              (Intercept)              sexman              class2nd class
              35.25000000              0.01490564              0.20294599
              class3rd class sexman:class2nd class sexman:class3rd class
              0.02408671              1.52240125              16.47718494

> fitcases2<- matrix(c(1,0,0,0,0,0, 1,0,1,0,0,0, 1,0,0,1,0,0,
+ 1,1,0,0,0,0, 1,1,1,0,1,0, 1,1,0,1,0,1), ncol=6, byrow=TRUE) #cases for sex
x class
> adjust2<- matrix(fitcases2*%coef(fit6j), ncol=3, byrow=TRUE) #link version
> colnames(adjust2)<- c('1st class', '2nd class', '3rd class')
> rownames(adjust2)<- c('women', 'man')
> exp(adjust2) #Odds version
              1st class 2nd class 3rd class
women 35.2500000 7.1538462 0.8490566
man    0.5254237 0.1623377 0.2085308

> 0.5254237/35.25 #OR for man:women | 1st class
[1] 0.01490564

> 7.1538462/35.25 #OR for 2nd:1st | women
[1] 0.202946
```

```

> 0.8490566/35.25 #OR for 3rd:1st | women
[1] 0.02408671

> 0.162338/7.15385
[1] 0.02269240

> 0.525424/35.25
[1] 0.01490565

> 0.0226924/0.01490565
[1] 1.522403

> 0.208531/0.849057
[1] 0.2456031

> 0.525424/35.25
[1] 0.01490565

> 0.24560306/0.01490565
[1] 16.47718

```

SECTION 6.3:

R CODE:

```

#SECTION 6.3: BINARY x CONTINUOUS INTERACTIONS
heart<- read.dta('heart01.dta')
heart$anterior<- factor(heart$anterior) #fix up anterior, dropping unused
levels
summary(heart$age) #show age stats
cat('Non-missing:', sum(!is.na(heart$age)), 'Missing:',
    sum(is.na(heart$age)), '\n')

fit6k<- glm(death ~ anterior + age, data=heart, family=binomial)
summary(fit6k)
exp(coef(fit6k)) #ORs

axa<- as.numeric(heart$anterior=='Anterior')*heart$age #interaction
fit6l<- glm(death ~ anterior + age + axa, data=heart, family=binomial)
summary(fit6l)
exp(coef(fit6l)) #ORs

fit6m<- glm(death ~ anterior*age, data=heart, family=binomial)
summary(fit6m)

xage<- 40:100 #age abscissae
eta1<- coef(fit6m)[1] + coef(fit6m)[3]*xage #link values
p1<- 1/(1+exp(-eta1)) #response values
eta2<- coef(fit6m)[1] + coef(fit6m)[2] + coef(fit6m)[3]*xage +
    coef(fit6m)[4]*xage #link values
p2<- 1/(1+exp(-eta2)) #response values

plot(xage, p1, type='l', ylim=c(0,0.25), xlab='Age (yr)', ylab='Probability',
     main='Relationship of Age and Anterior') #for Inferior
lines(xage, p2, col=2, pch=2) #for Anterior
legend('topleft', legend=c('Inferior', 'Anterior'), pch=c(1,2),
     col=c(1,2), bty='n')

```

```

plot(xage, etal, type='l', ylim=c(-5,-1), xlab='Age (yr)', ylab='Logit',
     main='Logit Diff: Anterior vs Interior on Age') #for Inferior
lines(xage, eta2, col=2, pch=2) #for Anterior
legend('topleft', legend=c('Inferior', 'Anterior'), pch=c(1,2),
      col=c(1,2), bty='n')

fit6n<- glm(died ~ white + los, data=medpar, family=binomial)
summary(fit6n)
exp(coef(fit6n)) #ORs

whlo<- medpar$white*medpar$los #interaction (both are numeric)
fit6o<- glm(died ~ white + los + whlo, data=medpar, family=binomial)
summary(fit6o)
exp(coef(fit6o)) #ORs

fit6p<- glm(died ~ white*los, data=medpar, family=binomial)
summary(fit6p)
plot(medpar$los, predict(fit6p, type='response'), ylim=c(0,.5),
     col=(medpar$white+1), pch=(medpar$white+1), xlab='Length of Stay',
     ylab='Logit', main='Interaction of White and LOS')
legend(locator(1), legend=c('White=0', 'White=1'), pch=c(1,2),
      col=c(1,2), bty='n')

exp(0.7703 + (-0.0477*21))

summary(medpar$los) #show some stats
quantile(medpar$los, probs=c(.01,.05,.10,.25,.5,.75,.9,.95,.99))
table(medpar$white) #classification

wlos<- medpar$white*medpar$los #interaction (both numeric)
fit6q<- glm(died ~ white + los + wlos, data=medpar, family=binomial)
summary(fit6q)

plot(medpar$los, predict(fit6p, type='link'),
     col=(medpar$white+1), pch=(medpar$white+1), xlab='Length of Stay',
     ylab='Logit', main='Interaction of White and LOS')
legend(locator(1), legend=c('White=0', 'White=1'), pch=c(1,2),
      col=c(1,2), bty='n')

plot(medpar$los, predict(fit6p, type='response'), ylim=c(0,.5),
     col=(medpar$white+1), pch=(medpar$white+1), xlab='Length of Stay',
     ylab='Logit', main='Interaction of White and LOS')
legend(locator(1), legend=c('White=0', 'White=1'), pch=c(1,2),
      col=c(1,2), bty='n')

```

OUTPUT: 6.3

```

> #SECTION 6.3: BINARY x CONTINUOUS INTERACTIONS
> heart<- read.dta('heart01.dta')
> heart$anterior<- factor(heart$anterior) #fix up anterior, dropping unused
levels
> summary(heart$age) #show age stats
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  40.00   54.00   66.00   65.81   76.00   100.00

> cat('Non-missing:', sum(!is.na(heart$age)), 'Missing:',
+     sum(is.na(heart$age)), '\n')
Non-missing: 5388 Missing: 0

```

```

> fit6k<- glm(death ~ anterior + age, data=heart, family=binomial)
> summary(fit6k)

Call:
glm(formula = death ~ anterior + age, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7243  -0.3019  -0.2273  -0.1622   3.2678

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.60343    0.42930  -17.711 < 2e-16 ***
anteriorAnterior  0.72661    0.15780   4.605 4.13e-06 ***
age             0.05673    0.00545  10.409 < 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: 1572.0  on 4695  degrees of freedom
Residual deviance: 1427.4  on 4693  degrees of freedom
(692 observations deleted due to missingness)
AIC: 1433.4

Number of Fisher Scoring iterations: 6

> exp(coef(fit6k)) #ORs
      (Intercept) anteriorAnterior          age
      0.00049874      2.06806076      1.05836611

> axa<- as.numeric(heart$anterior=='Anterior')*heart$age #interaction
> fit6l<- glm(death ~ anterior + age + axa, data=heart, family=binomial)
> summary(fit6l)

Call:
glm(formula = death ~ anterior + age + axa, family = binomial,
    data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7359  -0.3000  -0.2240  -0.1625   3.2382

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -7.396859    0.688433  -10.744 < 2e-16 ***
anteriorAnterior  0.400441    0.871632   0.459  0.646
age             0.053977    0.009039   5.971 2.35e-09 ***
axa             0.004308    0.011337   0.380  0.704
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1572.0  on 4695  degrees of freedom
Residual deviance: 1427.2  on 4692  degrees of freedom
(692 observations deleted due to missingness)
AIC: 1435.2

Number of Fisher Scoring iterations: 7

> exp(coef(fit6l)) #ORs
      (Intercept) anteriorAnterior          age          axa
      0.00049874      2.06806076      1.05836611      1.004308

```

```

0.0006131755      1.4924834311      1.0554601900      1.0043175128

> fit6m<- glm(death ~ anterior*age, data=heart, family=binomial)
> summary(fit6m)

Call:
glm(formula = death ~ anterior * age, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7359  -0.3000  -0.2240  -0.1625   3.2382

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -7.396859    0.688433  -10.744 < 2e-16 ***
anteriorAnterior    0.400441    0.871632    0.459  0.646
age              0.053977    0.009039    5.971 2.35e-09 ***
anteriorAnterior:age 0.004308    0.011337    0.380  0.704
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

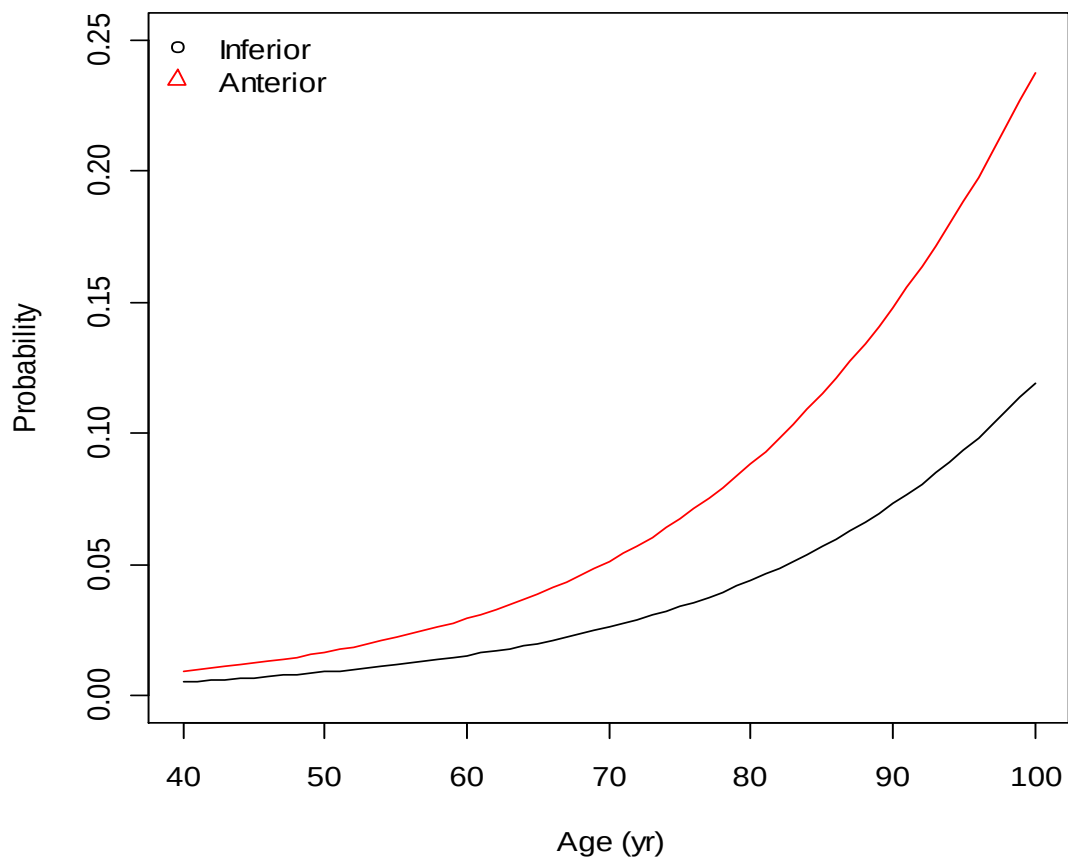
    Null deviance: 1572.0  on 4695  degrees of freedom
Residual deviance: 1427.2  on 4692  degrees of freedom
(692 observations deleted due to missingness)
AIC: 1435.2

Number of Fisher Scoring iterations: 7

> xage<- 40:100 #age abscissae
> eta1<- coef(fit6m)[1] + coef(fit6m)[3]*xage #link values
> p1<- 1/(1+exp(-eta1)) #response values
> eta2<- coef(fit6m)[1] + coef(fit6m)[2] + coef(fit6m)[3]*xage +
+   coef(fit6m)[4]*xage #link values
> p2<- 1/(1+exp(-eta2)) #response values
> plot(xage, p1, type='l', ylim=c(0,0.25), xlab='Age (yr)', ylab='Probability',
+   main='Relationship of Age and Anterior') #for Inferior
> lines(xage, p2, col=2, pch=2) #for Anterior
> legend('topleft', legend=c('Inferior', 'Anterior'), pch=c(1,2),
+   col=c(1,2), bty='n')

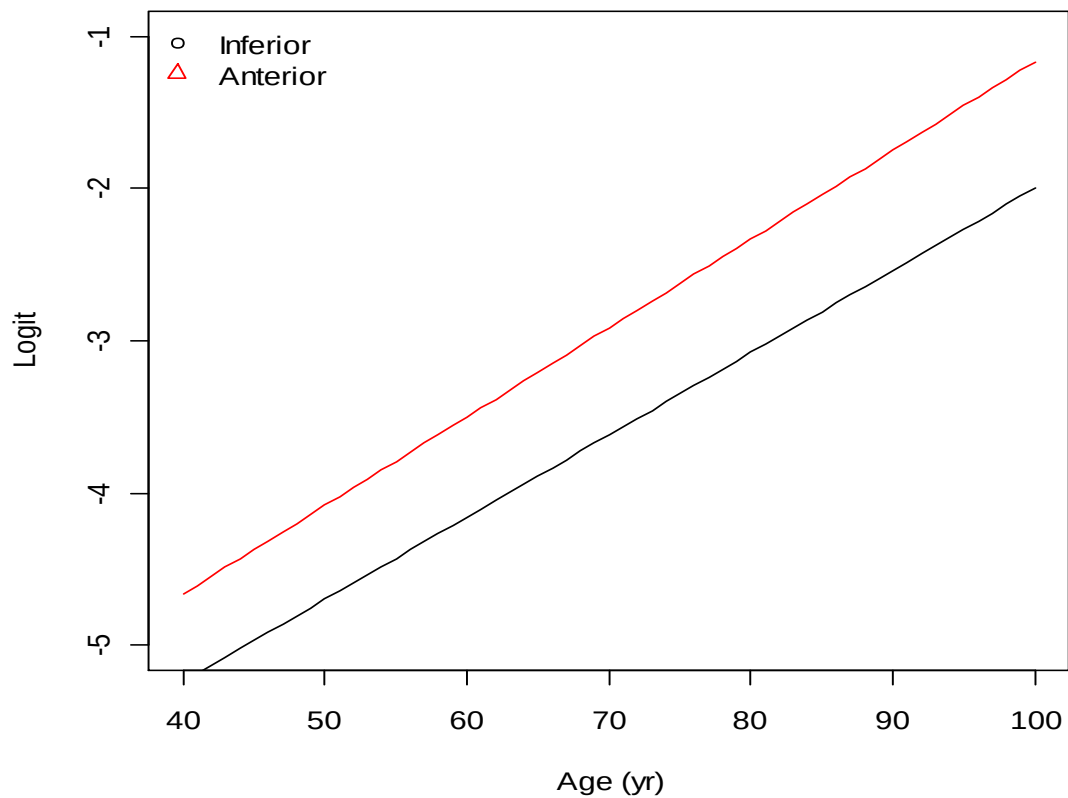
```


Relationship of Age and Anterior



```
> plot(xage, eta1, type='l', ylim=c(-5,-1), xlab='Age (yr)', ylab='Logit',  
+      main='Logit Diff: Anterior vs Interior on Age') #for Inferior  
> lines(xage, eta2, col=2, pch=2) #for Anterior  
> legend('topleft', legend=c('Inferior', 'Anterior'), pch=c(1,2),  
+      col=c(1,2), bty='n')
```

Logit Diff: Anterior vs Interior on Age



```
> fit6n<- glm(died ~ white + los, data=medpar, family=binomial)
> summary(fit6n)
```

```
Call:
glm(formula = died ~ white + los, family = binomial, data = medpar)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0225  -0.9522  -0.8630   1.3673   2.5934
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.598683   0.213268  -2.807   0.005 **
white        0.252681   0.206552   1.223   0.221
los         -0.029987   0.007704  -3.893 9.92e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 1922.9  on 1494  degrees of freedom
Residual deviance: 1903.0  on 1492  degrees of freedom
AIC: 1909
```

```
Number of Fisher Scoring iterations: 4
```

```
> exp(coef(fit6n)) #ORs
(Intercept)      white      los
```

```

0.5495351    1.2874723    0.9704583

> whlo<- medpar$white*medpar$los #interaction (both are numeric)
> fit6o<- glm(died ~ white + los + whlo, data=medpar, family=binomial)
> summary(fit6o)

Call:
glm(formula = died ~ white + los + whlo, family = binomial, data = medpar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0468  -0.9577  -0.8459   1.3388   2.4380

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.04834    0.28024  -3.741 0.000183 ***
white        0.77092    0.29560   2.608 0.009107 **
los          0.01002    0.01619   0.619 0.535986
whlo        -0.04776    0.01829  -2.611 0.009035 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1922.9  on 1494  degrees of freedom
Residual deviance: 1897.1  on 1491  degrees of freedom
AIC: 1905.1

Number of Fisher Scoring iterations: 4

> exp(coef(fit6o)) #ORs
(Intercept)      white      los      whlo
  0.3505181  2.1617619  1.0100692  0.9533621

> fit6p<- glm(died ~ white*los, data=medpar, family=binomial)
> summary(fit6p)

Call:
glm(formula = died ~ white * los, family = binomial, data = medpar)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0468  -0.9577  -0.8459   1.3388   2.4380

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.04834    0.28024  -3.741 0.000183 ***
white        0.77092    0.29560   2.608 0.009107 **
los          0.01002    0.01619   0.619 0.535986
white:los    -0.04776    0.01829  -2.611 0.009035 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

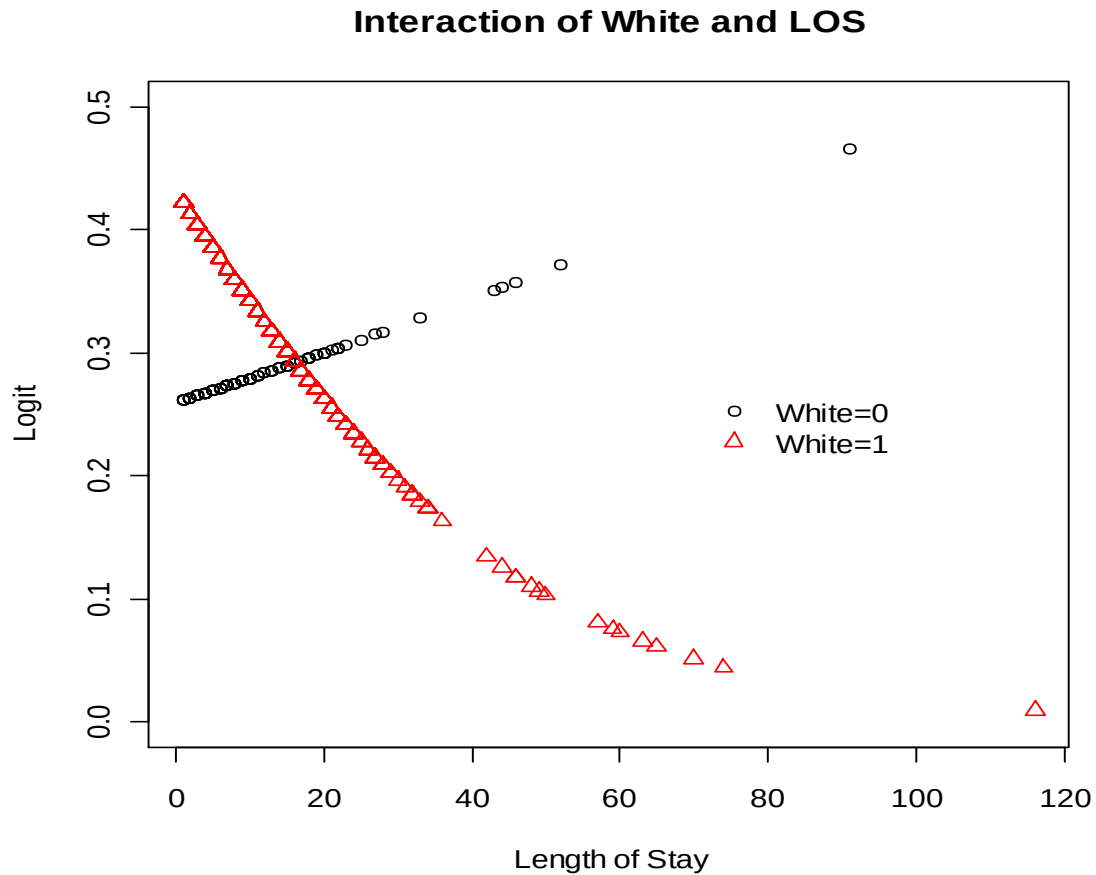
    Null deviance: 1922.9  on 1494  degrees of freedom
Residual deviance: 1897.1  on 1491  degrees of freedom
AIC: 1905.1

Number of Fisher Scoring iterations: 4

> plot(medpar$los,predict(fit6p, type='response'), ylim=c(0,.5),
+      col=(medpar$white+1), pch=(medpar$white+1), xlab='Length of Stay',

```

```
+ ylab='Logit', main='Interaction of White and LOS')
> legend(locator(1), legend=c('White=0', 'White=1'), pch=c(1,2),
+ col=c(1,2), bty='n')
```



```
> exp(0.7703 + (-0.0477*21))
[1] 0.793422
```

```
> summary(medpar$los) #show some stats
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   4.000   8.000  9.854  13.000 116.000
```

```
> quantile(medpar$los, probs=c(.01,.05,.10,.25,.5,.75,.9,.95,.99))
 1%  5% 10% 25% 50% 75% 90% 95% 99%
 1   1   2   4   8  13  19  23  46
```

```
> table(medpar$white) #classification
```

```
  0    1
127 1368
```

```
> wlos<- medpar$white*medpar$los #interaction (both numeric)
> fit6q<- glm(died ~ white + los + wlos, data=medpar, family=binomial)
> summary(fit6q)
```

```
Call:
glm(formula = died ~ white + los + wlos, family = binomial, data = medpar)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0468	-0.9577	-0.8459	1.3388	2.4380

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.04834	0.28024	-3.741	0.000183	***
white	0.77092	0.29560	2.608	0.009107	**
los	0.01002	0.01619	0.619	0.535986	
wlos	-0.04776	0.01829	-2.611	0.009035	**

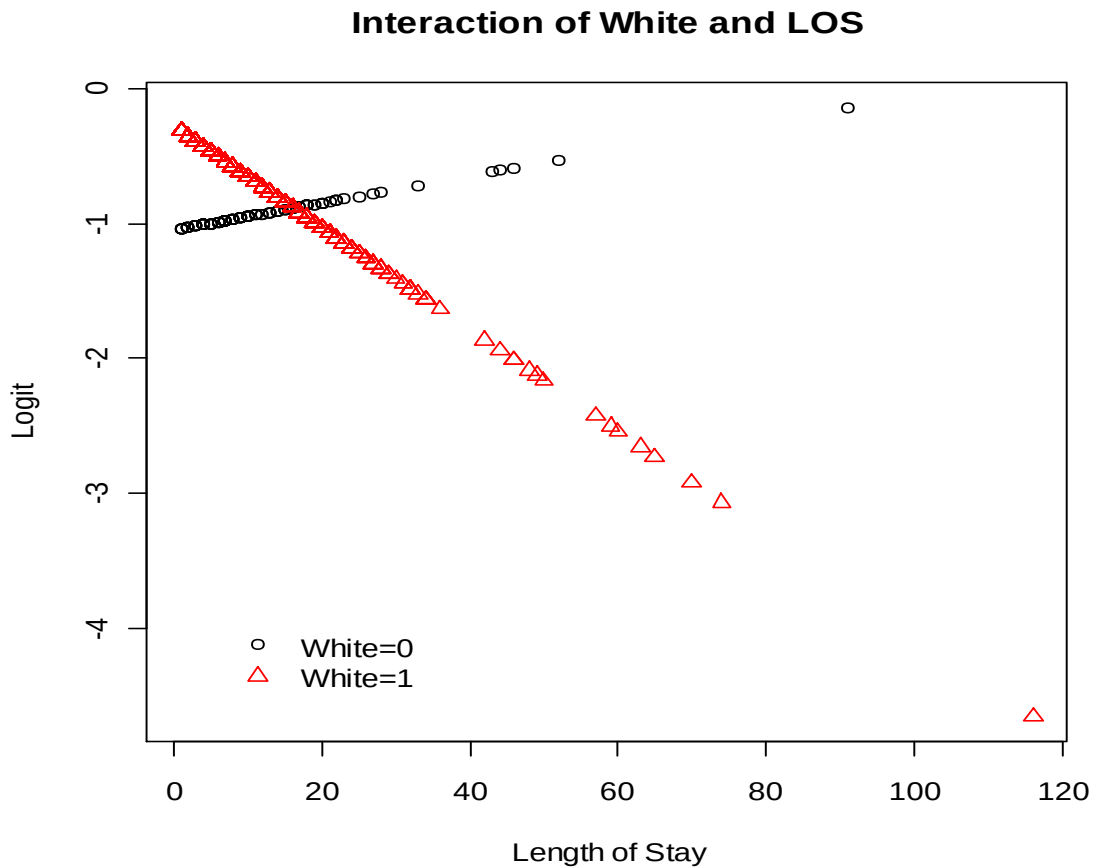
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1922.9 on 1494 degrees of freedom
Residual deviance: 1897.1 on 1491 degrees of freedom
AIC: 1905.1

Number of Fisher Scoring iterations: 4

```
> plot(medpar$los, predict(fit6p, type='link'),  
+ col=(medpar$white+1), pch=(medpar$white+1), xlab='Length of Stay',  
+ ylab='Logit', main='Interaction of White and LOS')  
> legend(locator(1), legend=c('White=0', 'White=1'), pch=c(1,2),  
+ col=c(1,2), bty='n')
```

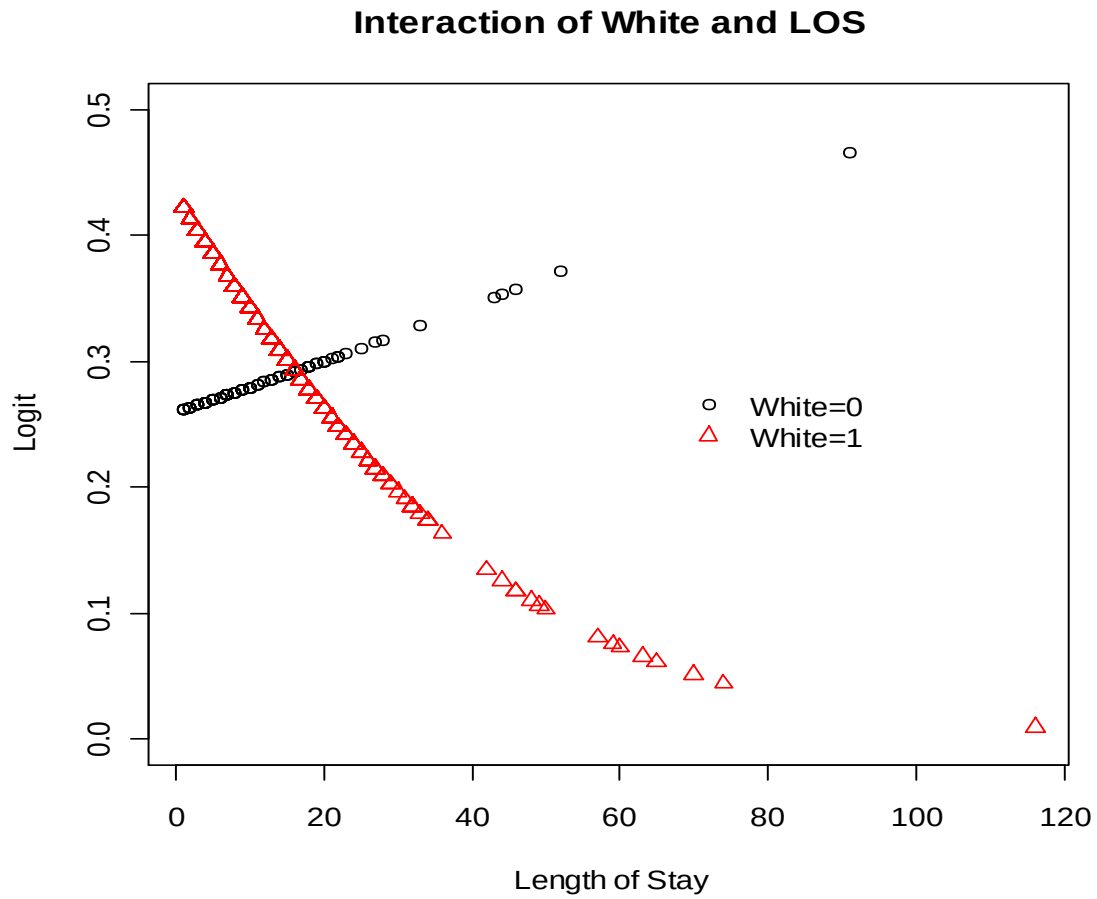


```
> plot(medpar$los, predict(fit6p, type='response'), ylim=c(0,.5),
```

```

+   col=(medpar$white+1), pch=(medpar$white+1), xlab='Length of Stay',
+   ylab='Logit', main='Interaction of White and LOS')
> legend(locator(1), legend=c('White=0', 'White=1'), pch=c(1,2),
+   col=c(1,2), bty='n')

```



SECTION 6.4:

```
#SECTION 6.4: CATEGORY x CONTINUOUS INTERACTIONS
table(heart$killip)
table(heart$killip)/sum(heart$killip>0, na.rm=TRUE)*100 #percents

fit6r<- glm(death ~ kk2 + kk3 + kk4, data=heart, family=binomial)
summary(fit6r)
exp(coef(fit6r)) #ORs

fit6s<- glm(death ~ age + factor(killip), data=heart, family=binomial)
summary(fit6s)
exp(coef(fit6s)) #ORs

xheart<- na.omit(heart) #drop NA's
plot(xheart$age, predict(fit6s, xheart, type='response'), ylim=c(0,1),
     col=xheart$killip, pch=xheart$killip,
     xlab='Age (yr)', ylab='Probability of Death',
     main='Prob. of Death: Killip x Age')
legend(locator(1), legend=c('Killip=1', 'Killip=2', 'Killip=3', 'Killip=4'),
     pch=1:4, col=1:4, bty='n')

fit6t<- glm(death ~ factor(killip)*age, data=heart, family=binomial)
summary(fit6t)
exp(coef(fit6t)) #ORs

plot(xheart$age, predict(fit6t, xheart, type='response'), ylim=c(0,1),
     col=xheart$killip, pch=xheart$killip,
     xlab='Age (yr)', ylab='Probability of Death',
     main='Prob. of Death: Killip x Age w/Interaction')
legend(locator(1), legend=c('Killip=1', 'Killip=2', 'Killip=3', 'Killip=4'),
     pch=1:4, col=1:4, bty='n')

ageXk1<- heart$age*heart$kk1 #interaction
ageXk3<- heart$age*heart$kk3 #interaction
ageXk4<- heart$age*heart$kk4 #interaction
fit6u<- glm(death ~ age + kk1 + kk3 + kk4 + ageXk1 + ageXk3 + ageXk4,
     data=heart, family=binomial) #fit relative to kk2
summary(fit6u)
exp(coef(fit6u)) #ORs
```

OUTPUT: 6.4

```
> #SECTION 6.4: CATEGORY x CONTINUOUS INTERACTIONS
> table(heart$killip)

  1    2    3    4
3768 1029 292  64

> table(heart$killip)/sum(heart$killip>0, na.rm=TRUE)*100 #percents

  1          2          3          4
73.122453 19.968950  5.666602  1.241995

> fit6r<- glm(death ~ kk2 + kk3 + kk4, data=heart, family=binomial)
> summary(fit6r)

Call:
glm(formula = death ~ kk2 + kk3 + kk4, family = binomial, data = heart)
```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0469  -0.2284  -0.2284  -0.2284   2.7054

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6335     0.1029 -35.323  < 2e-16 ***
kk2           1.0313     0.1603   6.432 1.26e-10 ***
kk3           1.4664     0.2185   6.712 1.92e-11 ***
kk4           3.3184     0.2732  12.146  < 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: 1849.1  on 5152  degrees of freedom
Residual deviance: 1698.6  on 5149  degrees of freedom
(235 observations deleted due to missingness)
AIC: 1706.6

Number of Fisher Scoring iterations: 6

> exp(coef(fit6r)) #ORs
      (Intercept)      kk2      kk3      kk4
0.02642332  2.80482319  4.33343826 27.61688492

> fit6s<- glm(death ~ age + factor(killip), data=heart, family=binomial)
> summary(fit6s)

Call:
glm(formula = death ~ age + factor(killip), family = binomial,
    data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5865  -0.2965  -0.2211  -0.1605   3.1930

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.080451   0.398005 -17.790  < 2e-16 ***
age           0.049725   0.005188   9.585  < 2e-16 ***
factor(killip)2 0.874362   0.163007   5.364 8.14e-08 ***
factor(killip)3 1.098789   0.224377   4.897 9.73e-07 ***
factor(killip)4 3.081978   0.287323  10.727  < 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: 1849.1  on 5152  degrees of freedom
Residual deviance: 1600.7  on 5148  degrees of freedom
(235 observations deleted due to missingness)
AIC: 1610.7

Number of Fisher Scoring iterations: 6

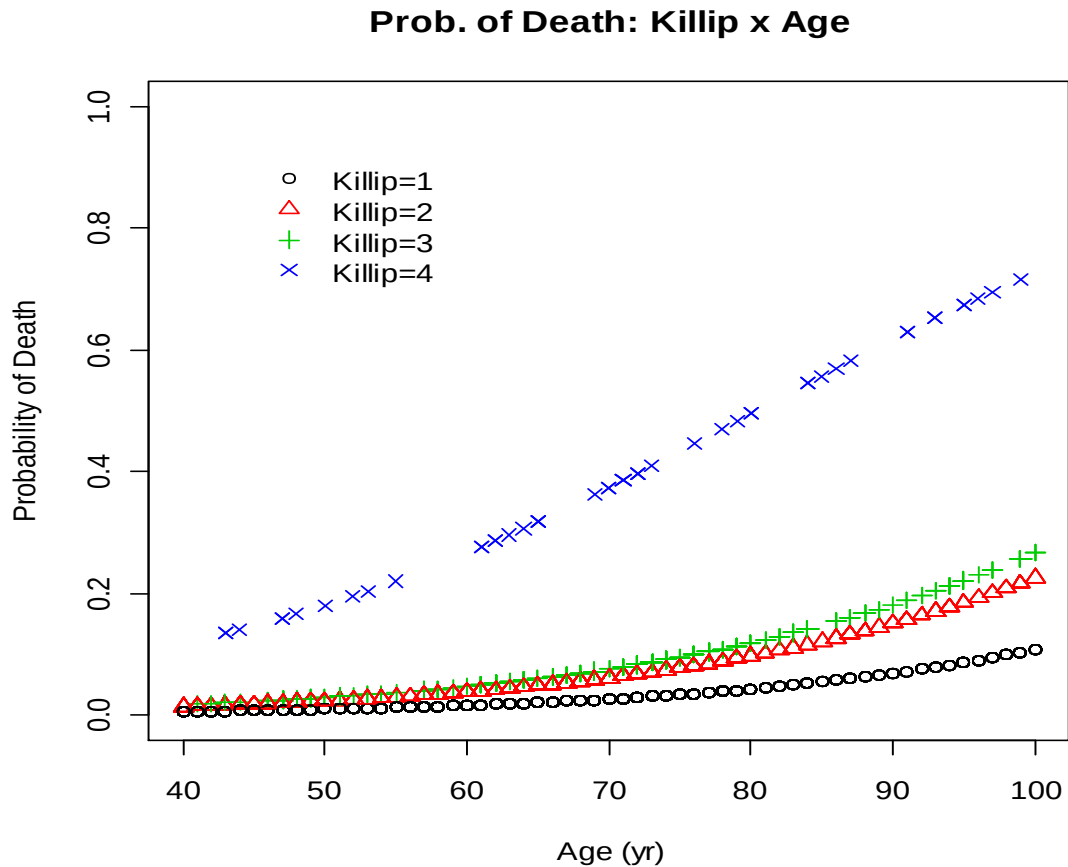
> exp(coef(fit6s)) #ORs
      (Intercept)      age factor(killip)2 factor(killip)3 factor(killip)4
8.413937e-04  1.050982e+00  2.397345e+00  3.000531e+00  2.180148e+01

> xheart<- na.omit(heart) #drop NA's
> plot(xheart$age, predict(fit6s, xheart, type='response'), ylim=c(0,1),
+      col=xheart$killip, pch=xheart$killip,

```



```
+ xlab='Age (yr)', ylab='Probability of Death',
+ main='Prob. of Death: Killip x Age')
> legend(locator(1), legend=c('Killip=1', 'Killip=2', 'Killip=3', 'Killip=4'),
+ pch=1:4, col=1:4, bty='n')
```



```
> fit6t<- glm(death ~ factor(killip)*age, data=heart, family=binomial)
> summary(fit6t)
```

Call:

```
glm(formula = death ~ factor(killip) * age, family = binomial,
    data = heart)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.1168	-0.3061	-0.2103	-0.1396	3.3416

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-8.120975	0.582859	-13.933	< 2e-16 ***
factor(killip)2	2.091800	0.917526	2.280	0.02262 *
factor(killip)3	3.880230	1.294041	2.999	0.00271 **
factor(killip)4	7.320069	1.411028	5.188	2.13e-07 ***
age	0.063543	0.007495	8.478	< 2e-16 ***
factor(killip)2:age	-0.016134	0.011830	-1.364	0.17265
factor(killip)3:age	-0.035975	0.016556	-2.173	0.02978 *
factor(killip)4:age	-0.056909	0.018737	-3.037	0.00239 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

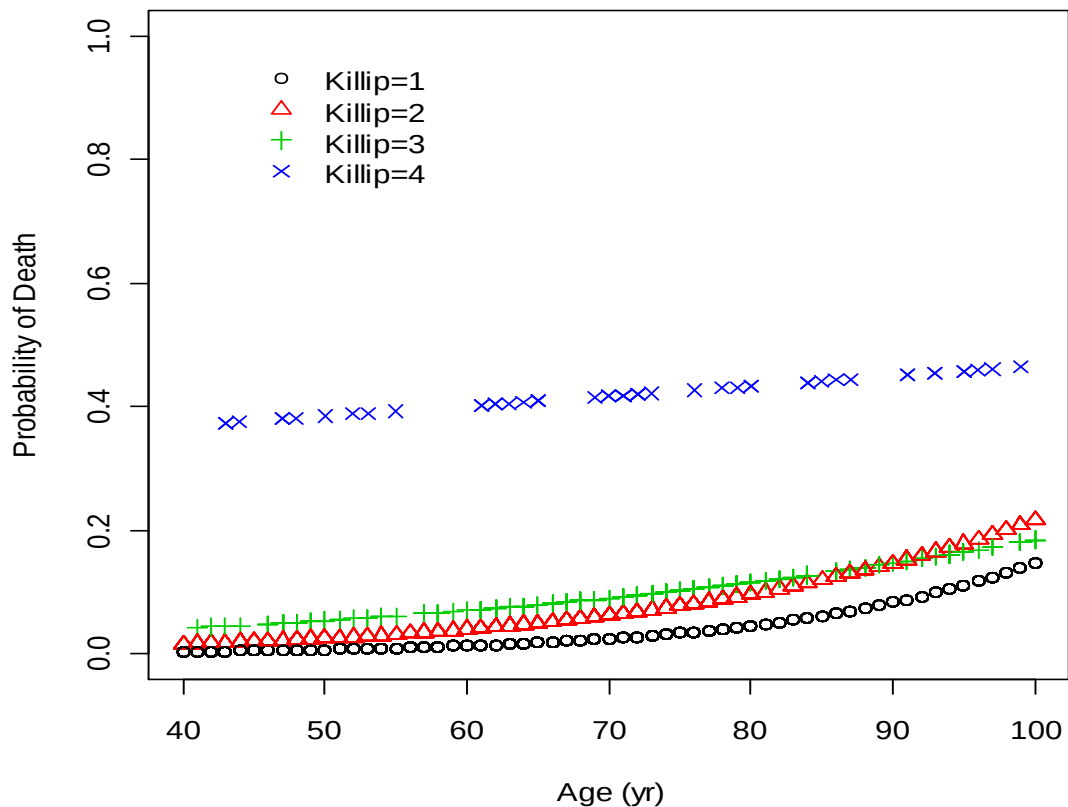
Null deviance: 1849.1 on 5152 degrees of freedom
Residual deviance: 1589.2 on 5145 degrees of freedom
(235 observations deleted due to missingness)
AIC: 1605.2

Number of Fisher Scoring iterations: 7

```
> exp(coef(fit6t)) #ORs
      (Intercept)      factor(killip)2      factor(killip)3      factor(killip)4
2.972388e-04      8.099483e+00      4.843537e+01      1.510308e+03
      age factor(killip)2:age factor(killip)3:age factor(killip)4:age
1.065605e+00      9.839959e-01      9.646640e-01      9.446797e-01

> plot(xheart$age, predict(fit6t, xheart, type='response'), ylim=c(0,1),
+      col=xheart$killip, pch=xheart$killip,
+      xlab='Age (yr)', ylab='Probability of Death',
+      main='Prob. of Death: Killip x Age w/Interaction')
> legend(locator(1), legend=c('Killip=1', 'Killip=2', 'Killip=3', 'Killip=4'),
+      pch=1:4, col=1:4, bty='n')
```

Prob. of Death: Killip x Age w/Interaction



```
> ageXk1<- heart$age*heart$kk1 #interaction
> ageXk3<- heart$age*heart$kk3 #interaction
> ageXk4<- heart$age*heart$kk4 #interaction
> fit6u<- glm(death ~ age + kk1 + kk3 + kk4 + ageXk1 + ageXk3 + ageXk4,
```

```

+ data=heart, family=binomial) #fit relative to kk2
> summary(fit6u)

Call:
glm(formula = death ~ age + kk1 + kk3 + kk4 + ageXk1 + ageXk3 +
    ageXk4, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1168  -0.3061  -0.2103  -0.1396   3.3416

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.029174   0.708610  -8.508 < 2e-16 ***
age           0.047409   0.009153   5.179 2.23e-07 ***
kk1          -2.091800   0.917526  -2.280 0.022618 *
kk3           1.788430   1.355340   1.320 0.186987
kk4           5.228269   1.467448   3.563 0.000367 ***
ageXk1        0.016134   0.011830   1.364 0.172650
ageXk3       -0.019842   0.017370  -1.142 0.253325
ageXk4       -0.040776   0.019460  -2.095 0.036135 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1849.1  on 5152  degrees of freedom
Residual deviance: 1589.2  on 5145  degrees of freedom
(235 observations deleted due to missingness)
AIC: 1605.2

Number of Fisher Scoring iterations: 7

> exp(coef(fit6u)) #ORs
(Intercept)          age          kk1          kk3          kk4          ageXk1
2.407481e-03 1.048551e+00 1.234647e-01 5.980057e+00 1.864697e+02 1.016264e+00
          ageXk3          ageXk4
9.803537e-01 9.600443e-01

```

Chapter 7

SECTION 7.1:

```

#SECTION 7.1: TRADITIONAL FIT TESTS
library('foreign')
heart<- read.dta('heart01.dta') #read Stata format file
heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that complicate
tabulations
heart$center<- factor(heart$center) #convert to factor from numeric levels
heart$killip<- factor(heart$killip) #convert to factor from numeric levels

fit7_1a<- glm(death ~ 1, data=heart, family=binomial(link=logit)) #intercept
only
summary(fit7_1a)

```

```
fit7_1b<- glm(death ~ age, data=heart, family=binomial)
summary(fit7_1b)
```

```
1- 1826.6/1974.8 #pseudo-R^2 from deviances
```

```
1974.8-1826.6 #Likelihood ratio from deviances
anova(fit7_1a, fit7_1b, test='Chisq') #LRT
```

OUTPUT: 7.1

```
> #SECTION 7.1: TRADITIONAL FIT TESTS
> library('foreign')
> heart<- read.dta('heart01.dta') #read Stata format file
> heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that
complicate tabulations
> heart$center<- factor(heart$center) #convert to factor from numeric levels
> heart$killip<- factor(heart$killip) #convert to factor from numeric levels
> fit7_1a<- glm(death ~ 1, data=heart, family=binomial(link=logit)) #intercept
only
> summary(fit7_1a)
```

Call:

```
glm(formula = death ~ 1, family = binomial(link = logit), data = heart)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.3032	-0.3032	-0.3032	-0.3032	2.4912

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.05704	0.06577	-46.48	<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: 1974.8 on 5387 degrees of freedom
Residual deviance: 1974.8 on 5387 degrees of freedom
AIC: 1976.8

Number of Fisher Scoring iterations: 5

```
> fit7_1b<- glm(death ~ age, data=heart, family=binomial)
> summary(fit7_1b)
```

Call:

```
glm(formula = death ~ age, family = binomial, data = heart)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.6569	-0.3375	-0.2559	-0.1780	3.1060

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.076784	0.372890	-18.98	<2e-16 ***
age	0.056529	0.004819	11.73	<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: 1974.8 on 5387 degrees of freedom
Residual deviance: 1826.6 on 5386 degrees of freedom
AIC: 1830.6

```

```

Number of Fisher Scoring iterations: 6

```

```

> 1- 1826.6/1974.8 #pseudo-R^2 from deviances
[1] 0.07504557

```

```

> 1974.8-1826.6 #Likelihood ratio from deviances
[1] 148.2

```

```

> anova(fit7_1a, fit7_1b, test='Chisq') #LRT
Analysis of Deviance Table

```

```

Model 1: death ~ 1
Model 2: death ~ age
  Resid. Df Resid. Dev    Df Deviance P(>|Chi|)
1     5387     1974.81      0    1974.81  1.000e+00
2     5386     1826.57      1    148.24  4.203e-34

```

SECTION 7.2:

```

#SECTION 7.2: HOSMER-LEMESHOW GOF TEST
fit7_2a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart, family=binomial)
summary(fit7_2a)
exp(coef(fit7_2a)) #ORs

source('ralhlGOFtest.r') #macro for H-L GOF test
heart2<- na.omit(heart) #drop rows with missing data
hlGOF.test(heart2$death, predict(fit7_2a, heart2, type='response'), breaks=10)
hlGOF.test(heart2$death, predict(fit7_2a, heart2, type='response'), breaks=6)

mean(heart$death)
sqrt(mean(heart$death)*(1-mean(heart$death))/length(heart$death))
library('binGroup') #binomial package
binBlaker(length(heart$death), sum(heart$death)) #Blaker C.I. for proportion

mu<- predict(fit7_2a, heart2, type='response') #get predicted values again
muSE<- sd(mu)/sqrt(length(mu)) #S.E. of mean
cat('Mean:', mean(mu), 'S.E.:', muSE, '95% C.I.:',
  mean(mu)-1.96*muSE, mean(mu)+1.96*muSE, '\n')
length(mu)

mean(heart2$death) #new mean for rows without missing data
sqrt(mean(heart2$death)*(1-mean(heart2$death))/length(heart2$death))
binBlaker(length(heart2$death), sum(heart2$death)) #Blaker C.I. for proportion

library('epicalc') #package for cmx() and lroc()
cmxdf<- data.frame(id=1:nrow(heart2), death=heart2$death, pred=mu)
cmx(cmxdf, threshold=0.05) #confusion matrix

lroc(fit7_2a) #ROC plot

```

OUTPUT: 7.2

```

> #SECTION 7.2: HOSMER-LEMESHOW GOF TEST
> fit7_2a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +

```

```

+ age3 + age4, data=heart, family=binomial)
> summary(fit7_2a)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
     age4, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8159     0.1934 -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg           0.7864     0.3527   2.229 0.025786 *
kk2             0.8249     0.1804   4.572 4.84e-06 ***
kk3            0.7967     0.2692   2.959 0.003083 **
kk4            2.6837     0.3565   7.529 5.11e-14 ***
age3           1.2668     0.2006   6.316 2.68e-10 ***
age4           1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3

Number of Fisher Scoring iterations: 7

> exp(coef(fit7_2a)) #ORs
      (Intercept) anteriorAnterior      hcabg      kk2
0.008099558      1.894095884      2.195519114      2.281692192
      kk3      kk4      age3      age4
2.218198595      14.639836908      3.549576517      6.964846706

> source('ralhlGOFtest.r') #macro for H-L GOF test
> heart2<- na.omit(heart) #drop rows with missing data
> hlGOF.test(heart2$death, predict(fit7_2a, heart2, type='response'),
breaks=10)

Hosmer-Lemeshow GOF test

For # Cuts = 8 # Data = 4503
Cut # Total #Patterns # Resp. # Pred. Mean Resp. Mean Pred.
1 563 1 7 4.52 0.01243 0.00803
2 563 1 2 4.52 0.00355 0.00803
3 563 2 6 7.77 0.01066 0.01380
4 563 4 6 8.87 0.01066 0.01575
5 562 2 8 13.79 0.01423 0.02455
6 563 7 25 21.07 0.04440 0.03742
7 563 8 38 32.68 0.06750 0.05805
8 563 24 84 82.77 0.14920 0.14702
Total # Data: 4503 Total over cuts: 4503
Chisq: 8.334831 d.f.: 6 P-value: 0.21458

For # Cuts = 10 # Data = 4503
Cut # Total #Patterns # Resp. # Pred. Mean Resp. Mean Pred.
1 450 1 6 3.62 0.01333 0.00803
2 451 1 3 3.62 0.00665 0.00803

```

3	450	2	3	4.47	0.00667	0.00994
4	450	1	3	6.80	0.00667	0.01511
5	451	4	6	7.17	0.01330	0.01591
6	450	2	7	10.66	0.01556	0.02370
7	450	4	12	13.95	0.02667	0.03099
8	450	5	24	21.94	0.05333	0.04875
9	451	7	43	31.90	0.09534	0.07073
10	450	24	69	71.87	0.15333	0.15970

Total # Data: 4503 Total over cuts: 4503
Chisq: 10.59860 d.f.: 8 P-value: 0.22550

For # Cuts = 12 # Data = 4503

Cut	# Total	#Patterns	# Resp.	# Pred.	Mean Resp.	Mean Pred.
1	375	1	6	3.01	0.01600	0.00803
2	375	1	1	3.01	0.00267	0.00803
3	376	1	2	3.02	0.00532	0.00803
4	375	2	5	4.93	0.01333	0.01315
5	375	1	2	5.67	0.00533	0.01511
6	376	4	5	6.04	0.01330	0.01606
7	375	2	5	8.57	0.01333	0.02285
8	375	4	10	10.97	0.02667	0.02925
9	375	4	18	15.32	0.04800	0.04086
10	375	3	20	19.80	0.05333	0.05280
11	376	9	41	31.75	0.10904	0.08445
12	375	21	61	63.90	0.16267	0.17040

Total # Data: 4503 Total over cuts: 4503
Chisq: 12.47930 d.f.: 10 P-value: 0.25426

Minimum P-value: 0.21458

> hlGOF.test(heart2\$death, predict(fit7_2a, heart2, type='response'), breaks=6)

Hosmer-Lemeshow GOF test

For # Cuts = 4 # Data = 4503

Cut	# Total	#Patterns	# Resp.	# Pred.	Mean Resp.	Mean Pred.
1	1126	1	9	9.05	0.00799	0.00803
2	1126	5	12	16.64	0.01066	0.01477
3	1125	8	33	34.86	0.02933	0.03099
4	1126	31	122	115.46	0.10835	0.10254

Total # Data: 4503 Total over cuts: 4503
Chisq: 1.827500 d.f.: 2 P-value: 0.40102

For # Cuts = 6 # Data = 4503

Cut	# Total	#Patterns	# Resp.	# Pred.	Mean Resp.	Mean Pred.
1	750	1	7	6.03	0.00933	0.00803
2	751	2	7	7.95	0.00932	0.01059
3	751	4	7	11.71	0.00932	0.01559
4	750	5	15	19.54	0.02000	0.02605
5	750	6	38	35.12	0.05067	0.04683
6	751	29	102	95.65	0.13582	0.12737

Total # Data: 4503 Total over cuts: 4503
Chisq: 4.007261 d.f.: 4 P-value: 0.40502

For # Cuts = 8 # Data = 4503

Cut	# Total	#Patterns	# Resp.	# Pred.	Mean Resp.	Mean Pred.
1	563	1	7	4.52	0.01243	0.00803
2	563	1	2	4.52	0.00355	0.00803
3	563	2	6	7.77	0.01066	0.01380
4	563	4	6	8.87	0.01066	0.01575
5	562	2	8	13.79	0.01423	0.02455
6	563	7	25	21.07	0.04440	0.03742
7	563	8	38	32.68	0.06750	0.05805

```

      8      563      24      84      82.77 0.14920 0.14702
Total # Data: 4503 Total over cuts: 4503
Chisq: 8.334831 d.f.: 6 P-value: 0.21458

Minimum P-value: 0.21458

> mean(heart$death)
[1] 0.04491463

> sqrt(mean(heart$death)*(1-mean(heart$death))/length(heart$death))
[1] 0.002821637

> library('binGroup') #binomial package
> binBlaker(length(heart$death), sum(heart$death)) #Blaker C.I. for proportion
[1] 0.03953938 0.05078944

> mu<- predict(fit7_2a, heart2, type='response') #get predicted values again
> muSE<- sd(mu)/sqrt(length(mu)) #S.E. of mean
> cat('Mean:', mean(mu), 'S.E.: ', muSE, '95% C.I.: ',
+     mean(mu)-1.96*muSE, mean(mu)+1.96*muSE, '\n')
Mean: 0.03908505 S.E.: 0.0008121251 95% C.I.: 0.03749329 0.04067682

> length(mu)
[1] 4503

> mean(heart2$death) #new mean for rows without missing data
[1] 0.03908505

> sqrt(mean(heart2$death)*(1-mean(heart2$death))/length(heart2$death))
[1] 0.002887998

> binBlaker(length(heart2$death), sum(heart2$death)) #Blaker C.I. for
proportion
[1] 0.03361441 0.04516326

> library('epicalc') #package for cmx() and lroc()
> cmxdf<- data.frame(id=1:nrow(heart2), death=heart2$death, pred=mu)
> cmx(cmxdf, threshold=0.05) #confusion matrix
      observed
predicted    1    0
      1  133 1138
      0   43 3189

> lroc(fit7_2a) #ROC plot
$model.description
[1] "logit (death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4)"

$auc
[1] 0.7968307

$predicted.table
predicted.prob Non-diseased Diseased
      0.0080      1220      10
      0.0151       878       9
      0.0174        47       0
      0.0176        37       2
      0.0181       241       3
      0.0279       453       8
      0.0325        16       1
      0.0329        43       3
      0.0338       243       7
      0.0379         1       0
      0.0389        10       0

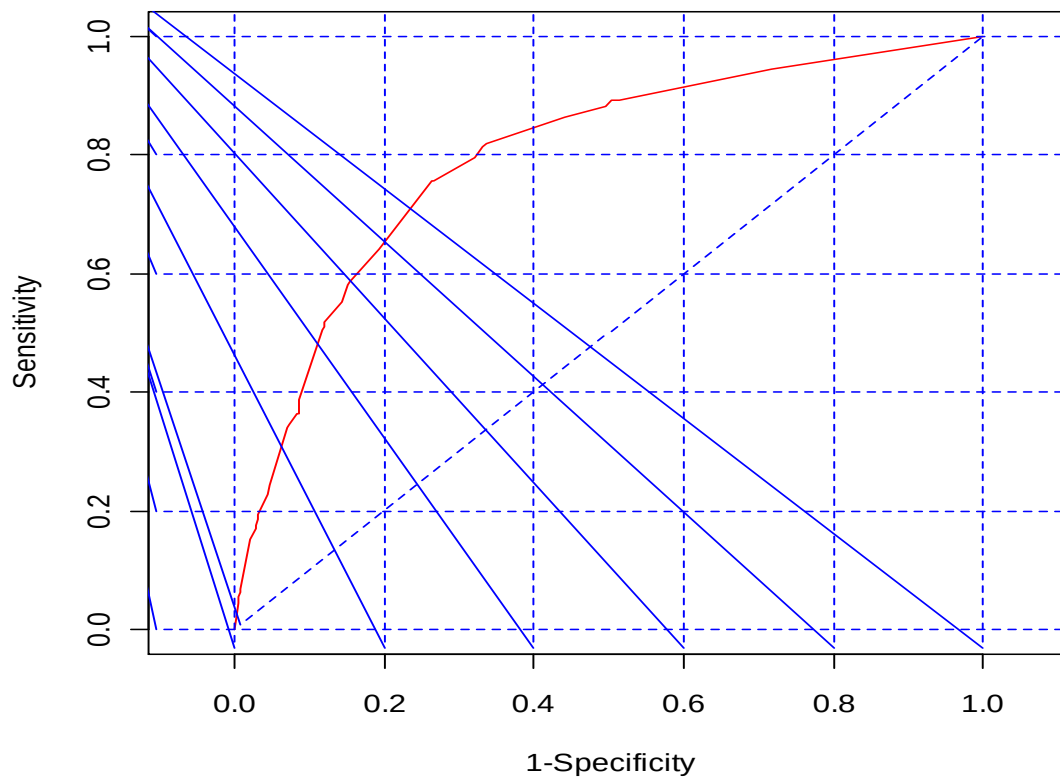
```


	0.0516	310	20
	0.0533	163	10
...			
	0.4435	5	4
	0.4523	3	0
	0.6100	7	4

```

$diagnostic.table
      1-Specificity Sensitivity
> 0.0080 1.000000000 1.00000000
> 0.0151 0.718049457 0.94318182
> 0.0174 0.504275480 0.89204545
> 0.0176 0.495724520 0.88068182
> 0.0181 0.440027733 0.86363636
...
> 0.4523 0.001617749 0.02272727
> 0.6100 0.000000000 0.00000000

```



SECTION 7.3:

```
#SECTION 7.3: INFORMATION CRITERIA TESTS
fit7_3a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart, family=binomial)
summary(fit7_3a)
exp(coef(fit7_3a)) #ORs
cat('AIC:',AIC(fit7_3a, k=2), 'BIC:',
  AIC(fit7_3a, k=log(nrow(!is.na(heart))))), '\n')
1276.3+8*2 #calc. AIC from deviance

fit7_3b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4,
  data=heart, family=binomial)
summary(fit7_3b)
exp(coef(fit7_3b)) #ORs
cat('AIC:',AIC(fit7_3b, k=2), 'BIC:',
  AIC(fit7_3b, k=log(nrow(!is.na(heart))))), '\n')
1372.6 + 6*2 #calc. AIC from deviance

1 - 1292.319/1384.575 #change in AIC between 2 models

-2*(as.numeric(logLik(fit7_3a)) - 8 - 8*(8+1)/(4503-8-1))/4503 #AICfs
-2*(as.numeric(logLik(fit7_3b)) - 6 - 6*(6+1)/(4503-6-1))/4503 #AICfs
0.2869978*4503 #compare to AIC=1292.3
0.3074825*4503 #compare to AIC=1384.6

(as.numeric(logLik(fit7_3a)) - 8)/(4503/2) - (1+log(2*pi)) #AIClimdep
(as.numeric(logLik(fit7_3b)) - 6)/(4503/2) - (1+log(2*pi)) #AIClimdep

(as.numeric(logLik(fit7_3a)) + 8*log(4503))/4503 #AICs
(as.numeric(logLik(fit7_3b)) + 6*log(4503))/4503 #AICs

cat('AIC:',AIC(fit7_3a, k=2), 'BIC:',
  AIC(fit7_3a, k=log(nrow(!is.na(heart))))), '\n')
deviance(fit7_3a) - 4495*log(4503)
-2*as.numeric(logLik(fit7_3a)) -4495*log(4503) #BIC

AIC(fit7_3a, k=log(nrow(!is.na(heart)))) - AIC(fit7_3b,
k=log(nrow(!is.na(heart))))

-2*(as.numeric(logLik(fit7_3a)) - 8*log(8))/4503 #HQIC
-2*(as.numeric(logLik(fit7_3b)) - 6*log(6))/4503 #HQIC

step(fit7_3a, direction='backward', k=2) #backward selection via AIC
```

OUTPUT: 7.3

```
> #SECTION 7.3: INFORMATION CRITERIA TESTS
> fit7_3a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart, family=binomial)
> summary(fit7_3a)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
  age4, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
```

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.8159     0.1934 -24.896 < 2e-16 ***
anteriorAnterior 0.6387     0.1675   3.812 0.000138 ***
hcabg          0.7864     0.3527   2.229 0.025786 *
kk2            0.8249     0.1804   4.572 4.84e-06 ***
kk3            0.7967     0.2692   2.959 0.003083 **
kk4            2.6837     0.3565   7.529 5.11e-14 ***
age3           1.2668     0.2006   6.316 2.68e-10 ***
age4           1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3

Number of Fisher Scoring iterations: 7

> exp(coef(fit7_3a)) #ORs
              (Intercept) anteriorAnterior          hcabg          kk2
0.008099558      1.894095884      2.195519114      2.281692192
              kk3          kk4          age3          age4
2.218198595      14.639836908      3.549576517      6.964846706

> cat('AIC:',AIC(fit7_3a, k=2), 'BIC:',
+   AIC(fit7_3a, k=log(nrow(!is.na(heart)))), '\n')
AIC: 1292.319 BIC: 1345.055

> 1276.3+8*2 #calc. AIC from deviance
[1] 1292.3

> fit7_3b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4,
+   data=heart, family=binomial)
> summary(fit7_3b)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4, family = binomial,
    data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0015  -0.2627  -0.2627  -0.1838   2.8598

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.0722     0.1495 -27.244 < 2e-16 ***
anteriorAnterior 0.7227     0.1647   4.389 1.14e-05 ***
hcabg          0.7185     0.3461   2.076 0.0379 *
kk2            0.9453     0.1771   5.337 9.42e-08 ***
kk3            1.1769     0.2621   4.490 7.14e-06 ***
kk4            2.9206     0.3300   8.849 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1372.6  on 4497  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1384.6

```

Number of Fisher Scoring iterations: 6

```
> exp(coef(fit7_3b)) #ORs
      (Intercept) anteriorAnterior      hcabg      kk2
      0.01703982      2.05989014      2.05136913      2.57369176
      kk3      kk4
      3.24425144      18.55298477
> cat('AIC:',AIC(fit7_3b, k=2), 'BIC:',
+     AIC(fit7_3b, k=log(nrow(!is.na(heart))))), '\n')
AIC: 1384.575 BIC: 1424.127

> 1372.6 + 6*2 #calc. AIC from deviance
[1] 1384.6

> 1 - 1292.319/1384.575 #change in AIC between 2 models
[1] 0.06663128

> -2*(as.numeric(logLik(fit7_3a)) - 8 - 8*(8+1)/(4503-8-1))/4503 #AICfs
[1] 0.2869978

> -2*(as.numeric(logLik(fit7_3b)) - 6 - 6*(6+1)/(4503-6-1))/4503 #AICfs
[1] 0.3074825

> 0.2869978*4503 #compare to AIC=1292.3
[1] 1292.351

> 0.3074825*4503 #compare to AIC=1384.6
[1] 1384.594

> (as.numeric(logLik(fit7_3a)) - 8)/(4503/2) - (1+log(2*pi)) #AIClimdep
[1] -3.124868

> (as.numeric(logLik(fit7_3b)) - 6)/(4503/2) - (1+log(2*pi)) #AIClimdep
[1] -3.145355

> (as.numeric(logLik(fit7_3a)) + 8*log(4503))/4503 #AICs
[1] -0.1267732

> (as.numeric(logLik(fit7_3b)) + 6*log(4503))/4503 #AICs
[1] -0.1411975

> cat('AIC:',AIC(fit7_3a, k=2), 'BIC:',
+     AIC(fit7_3a, k=log(nrow(!is.na(heart))))), '\n')
AIC: 1292.319 BIC: 1345.055

> deviance(fit7_3a) - 4495*log(4503)
[1] -36537.86

> -2*as.numeric(logLik(fit7_3a)) -4495*log(4503) #BIC
[1] -36537.86

> AIC(fit7_3a, k=log(nrow(!is.na(heart)))) - AIC(fit7_3b,
k=log(nrow(!is.na(heart))))
[1] -79.07202

> -2*(as.numeric(logLik(fit7_3a)) - 8*log(8))/4503 #HQIC
[1] 0.2908262

> -2*(as.numeric(logLik(fit7_3b)) - 6*log(6))/4503 #HQIC
[1] 0.3095883

> step(fit7_3a, direction='backward', k=2) #backward selection via AIC
```

```

Start:  AIC=1292.32
death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4

           Df Deviance    AIC
<none>             1276.3 1292.3
- hcabg           1   1280.5 1294.5
- kk3             1   1284.0 1298.0
- anterior        1   1291.3 1305.3
- kk2             1   1296.0 1310.0
- age3            1   1317.1 1331.1
- kk4             1   1320.1 1334.1
- age4            1   1361.7 1375.7

Call:  glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
age4, family = binomial, data = heart)

Coefficients:
      (Intercept)  anteriorAnterior          hcabg          kk2
          -4.8159           0.6387           0.7864           0.8249
           kk3           kk4           age3           age4
           0.7967           2.6837           1.2668           1.9409

Degrees of Freedom: 4502 Total (i.e. Null);  4495 Residual
(885 observations deleted due to missingness)
Null Deviance:      1486
Residual Deviance: 1276      AIC: 1292

```

SECTION 7.4:

```

#SECTION 7.4: RESIDUAL ANALYSIS
fit7_4a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart, family=binomial)
mu<- predict(fit7_4a, type='response')
raw<- heart2$death - mu #use heart2 because dropped missing data rows
head(cbind(mu, raw)) #predicted and raw residuals

pearson<- residuals(fit7_4a, type='pearson')
variance<- mu*(1-mu)
head(cbind(death=heart2$death, mu=mu, raw=raw,
  variance=variance, pearson=pearson))
-0.015109539/sqrt(0.01488124) #calc. pearson residual
pear<- sum(pearson*pearson) #chi-sq.
pear
length(pearson)

deviance<- residuals(fit7_4a, type='deviance')
head(cbind(death=heart2$death, mu=mu, deviance=deviance))
dev<- sum(deviance*deviance) #chi-sq.
dev

pred<- predict(fit7_4a, se.fit=TRUE, type='response')
mu<- pred$fit #predictions of probs
stdp<- pred$se.fit #S.E. of fit
variance<- mu*(1-mu)
h<- stdp*stdp*variance #hat statistics
hmd<- hatvalues(fit7_4a)
head(cbind(hmd=hmd, hat=h))

stdpear<- pearson/sqrt(1-hmd) #std. pearson resid.
head(cbind(pearson, stdpear))

```

```

summary(hmd)

sdeviance<- rstandard(fit7_4a) #standardized deviance residuals
head(cbind(deviance, sdeviance))

sdeviance2<- sdeviance*sdeviance
summary(sdeviance2[heart2$death==1])
summary(sdeviance2[heart2$death==0])

heart2[sdeviance2>9.63727 & heart2$anterior=='Inferior',c(1:7,13:16)]

mhat<- hatvalues(fit7_4a)
msdev<- rstandard(fit7_4a) #use std. deviance, std. pearson not available
plot(msdev, mhat, main='Std. Deviance by Hat', xlab='Std. Dev. Residual',
     ylab='Hat value', col='blue', xlim=c(-3,4))
abline(v=0, col='red')

rstud<- rstudent(fit7_4a)
dbeta<- dfbeta(fit7_4a)
dbetas<- dfbetas(fit7_4a)

layout(matrix(1:4, ncol=2))
plot(fit7_4a)

fit7_4b<- glm(death ~ anterior + age, data=heart, family=binomial)
summary(fit7_4b)
L1<- coef(fit7_4b)[1] + coef(fit7_4b)[3]*heart$age +
     coef(fit7_4b)[2]*(heart$anterior=='Anterior')
Y1<- 1/(1+exp(-L1))
L2<- coef(fit7_4b)[1] + coef(fit7_4b)[3]*heart$age +
     coef(fit7_4b)[2]*(heart$anterior!='Anterior')
Y2<- 1/(1+exp(-L2))
layout(1)
plot(heart$age, Y1, col=1, main='P[Death] within 48 hr admission',
     xlab='Age')
lines(heart$age, Y2, col=2, type='p')

knum<- as.numeric(heart$killip) #convert factor to level values
fit7_4c<- glm(death ~ age + knum, data=heart, family=binomial)
K1<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
     coef(fit7_4c)[3]*1
R1<- 1/(1+exp(-K1))
K2<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
     coef(fit7_4c)[3]*2
R2<- 1/(1+exp(-K2))
K3<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
     coef(fit7_4c)[3]*3
R3<- 1/(1+exp(-K3))
K4<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
     coef(fit7_4c)[3]*4
R4<- 1/(1+exp(-K4))

layout(1)
plot(heart$age, R1, col=1, main='P[Death] within 48 hr admission',
     sub='Killip Level', xlab='Age', ylim=c(0,0.6))
lines(heart$age, R2, col=2, type='p')
lines(heart$age, R3, col=3, type='p')
lines(heart$age, R4, col=4, type='p')

```

OUTPUT: 7.4

```
> #SECTION 7.4: RESIDUAL ANALYSIS
> fit7_4a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart, family=binomial)
> mu<- predict(fit7_4a, type='response')
> raw<- heart2$death - mu      #use heart2 because dropped missing data rows
> head(cbind(mu, raw))      #predicted and raw residuals
      mu      raw
1 0.015109539 -0.015109539
2 0.114036940  0.885963060
3 0.027946538 -0.027946538
4 0.110518233  0.889481767
5 0.033820360 -0.033820360
7 0.008034482 -0.008034482

> pearson<- residuals(fit7_4a, type='pearson')
> variance<- mu*(1-mu)
> head(cbind(death=heart2$death, mu=mu, raw=raw,
+   variance=variance, pearson=pearson))
      death      mu      raw  variance  pearson
1      0 0.015109539 -0.015109539 0.01488124 -0.12386016
2      1 0.114036940  0.885963060 0.10103252  2.78730843
3      0 0.027946538 -0.027946538 0.02716553 -0.16955825
4      1 0.110518233  0.889481767 0.09830395  2.83694920
5      0 0.033820360 -0.033820360 0.03267654 -0.18709413
7      0 0.008034482 -0.008034482 0.00796993 -0.08999754

> -0.015109539/sqrt(0.01488124) #calc. pearson residual
[1] -0.1238602

> pear<- sum(pearson*pearson) #chi-sq.
> pear
[1] 4212.632

> length(pearson)
[1] 4503

> deviance<- residuals(fit7_4a, type='deviance')
> head(cbind(death=heart2$death, mu=mu, deviance=deviance))
      death      mu  deviance
1      0 0.015109539 -0.1744984
2      1 0.114036940  2.0838584
3      0 0.027946538 -0.2380944
4      1 0.110518233  2.0988448
5      0 0.033820360 -0.2623185
7      0 0.008034482 -0.1270192

> dev<- sum(deviance*deviance) #chi-sq.
> dev
[1] 1276.319

> pred<- predict(fit7_4a, se.fit=TRUE, type='response')
> mu<- pred$fit      #predictions of probs
> stdp<- pred$se.fit #S.E. of fit
> variance<- mu*(1-mu)
> h<- stdp*stdp*variance #hat statistics
> hmd<- hatvalues(fit7_4a)
> head(cbind(hmd=hmd, hat=h))
      hmd      hat
1 0.0005094531 1.128190e-07
2 0.0046062711 4.701883e-05
```

```

3 0.0008865432 6.542386e-07
4 0.0032823295 3.171934e-05
5 0.0013680912 1.460787e-06
7 0.0002982295 1.894346e-08

> stdpear<- pearson/sqrt(1-hmd) #std. pearson resid.
> head(cbind(pearson, stdpear))
      pearson      stdpear
1 -0.12386016 -0.12389172
2  2.78730843  2.79375025
3 -0.16955825 -0.16963346
4  2.83694920  2.84161659
5 -0.18709413 -0.18722225
7 -0.08999754 -0.09001097

> summary(hmd)
      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
0.0002982 0.0002982 0.0008815 0.0017770 0.0013990 0.0419900

> sdeviance<- rstandard(fit7_4a) #standardized deviance residuals
> head(cbind(deviance, sdeviance))
      deviance  sdeviance
1 -0.1744984 -0.1745429
2  2.0838584  2.0886744
3 -0.2380944 -0.2382000
4  2.0988448  2.1022979
5 -0.2623185 -0.2624981
7 -0.1270192 -0.1270381

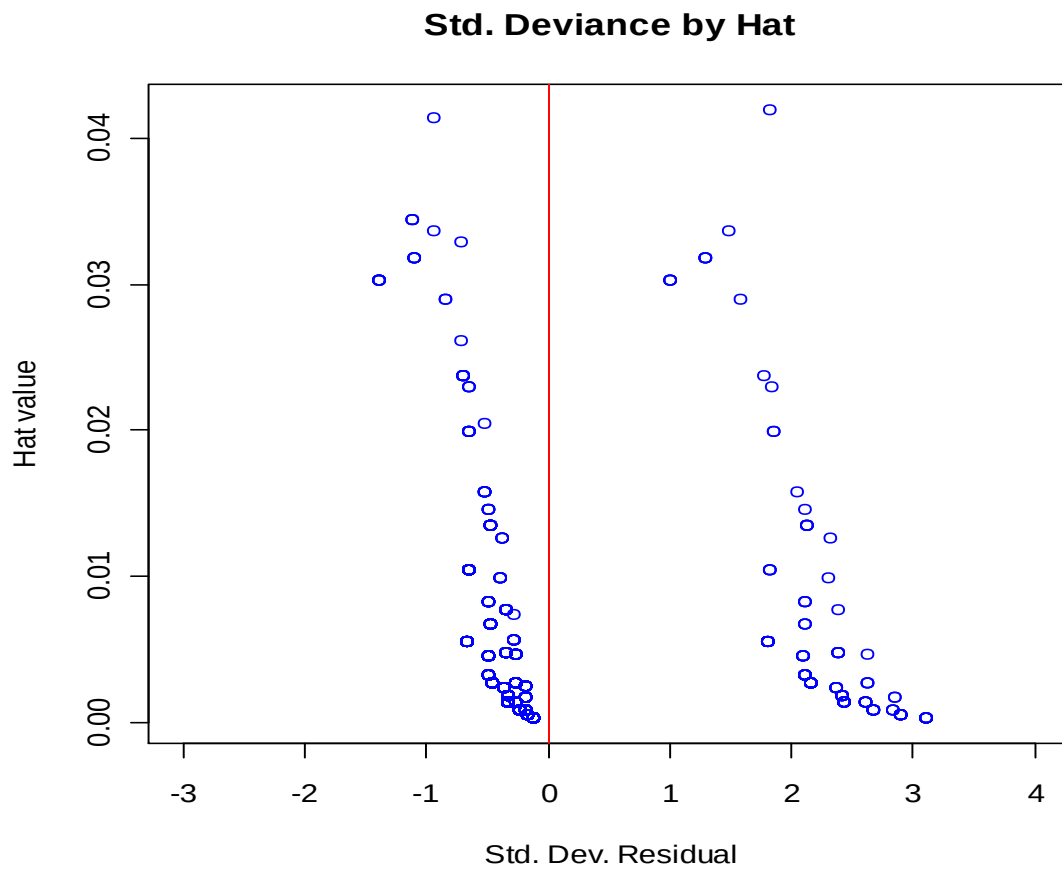
> sdeviance2<- sdeviance*sdeviance
> summary(sdeviance2[heart2$death==1])
      Min. 1st Qu.  Median     Mean 3rd Qu.    Max.
 1.019   4.420   4.689   5.369   5.935   9.651

> summary(sdeviance2[heart2$death==0])
      Min. 1st Qu.  Median     Mean 3rd Qu.    Max.
0.01614 0.01614 0.03567 0.07785 0.10620 1.94200

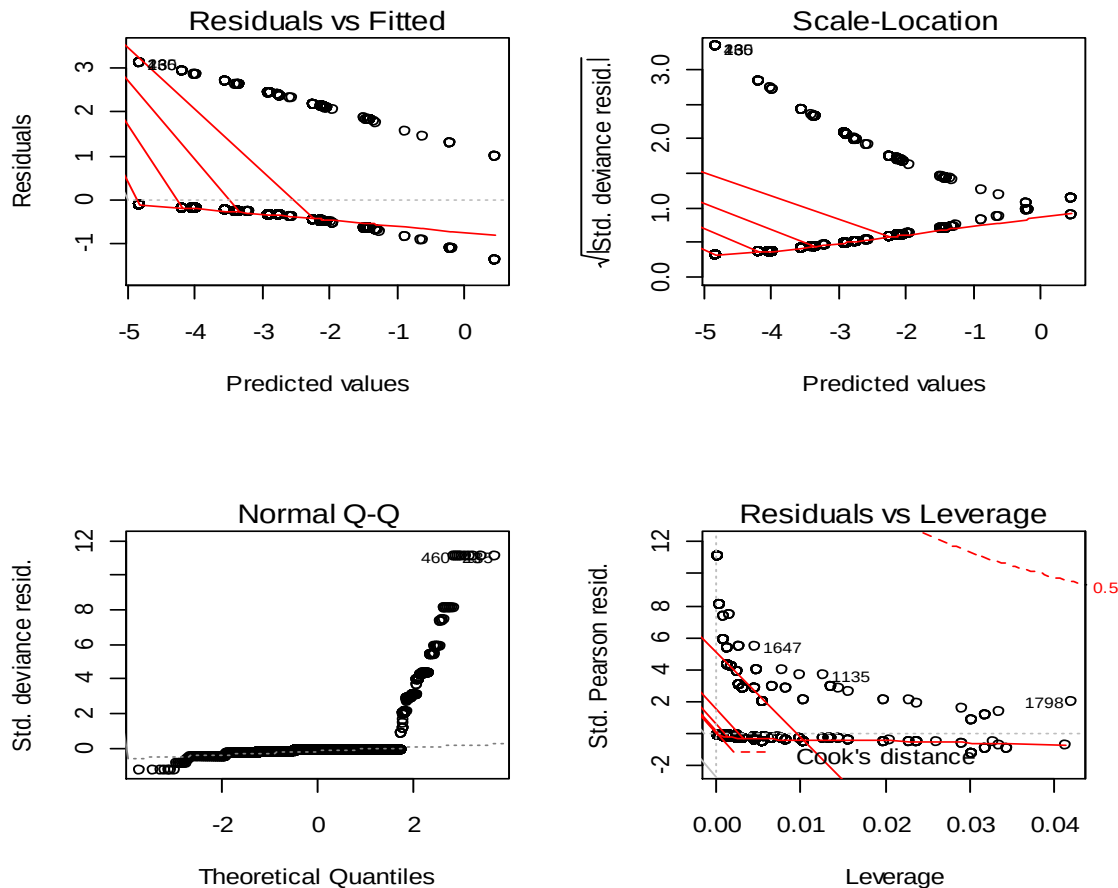
> heart2[sdeviance2>9.63727 & heart2$anterior=='Inferior',c(1:7,13:16)]
      death anterior hcabg kk1 kk2 kk3 kk4 age1 age2 age3 age4
23       1 Inferior     0   1   0   0   0   1   0   0   0
135      1 Inferior     0   1   0   0   0   1   0   0   0
460      1 Inferior     0   1   0   0   0   0   1   0   0
587      1 Inferior     0   1   0   0   0   0   1   0   0
895      1 Inferior     0   1   0   0   0   1   0   0   0
1003     1 Inferior     0   1   0   0   0   0   1   0   0
1961     1 Inferior     0   1   0   0   0   1   0   0   0
3399     1 Inferior     0   1   0   0   0   0   1   0   0
3642     1 Inferior     0   1   0   0   0   0   1   0   0
5350     1 Inferior     0   1   0   0   0   1   0   0   0

> mhat<- hatvalues(fit7_4a)
> msdev<- rstandard(fit7_4a) #use std. deviance, std. pearson not available
> plot(msdev, mhat, main='Std. Deviance by Hat', xlab='Std. Dev. Residual',
+       ylab='Hat value', col='blue', xlim=c(-3,4))
> abline(v=0, col='red')

```

```
> rstud<- rstudent(fit7_4a)
> dbeta<- dfbeta(fit7_4a)
> dbetas<- dfbetas(fit7_4a)
> layout(matrix(1:4, ncol=2))
> plot(fit7_4a)
```



```
> fit7_4b<- glm(death ~ anterior + age, data=heart, family=binomial)
> summary(fit7_4b)
```

Call:

```
glm(formula = death ~ anterior + age, family = binomial, data = heart)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.7243	-0.3019	-0.2273	-0.1622	3.2678

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.60343	0.42930	-17.711	< 2e-16 ***
anteriorAnterior	0.72661	0.15780	4.605	4.13e-06 ***
age	0.05673	0.00545	10.409	< 2e-16 ***

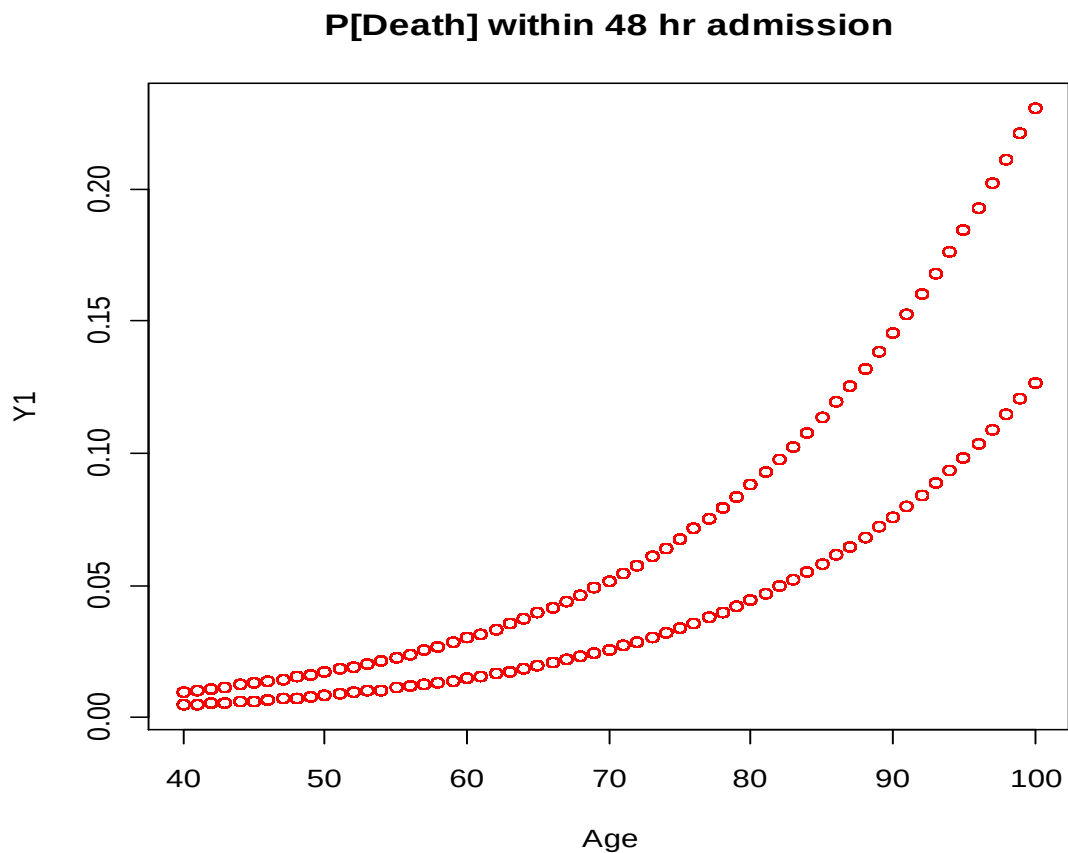
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1572.0 on 4695 degrees of freedom
 Residual deviance: 1427.4 on 4693 degrees of freedom
 (692 observations deleted due to missingness)
 AIC: 1433.4

Number of Fisher Scoring iterations: 6

```
> L1<- coef(fit7_4b)[1] + coef(fit7_4b)[3]*heart$age +
+   coef(fit7_4b)[2]*(heart$anterior=='Anterior')
> Y1<- 1/(1+exp(-L1))
> L2<- coef(fit7_4b)[1] + coef(fit7_4b)[3]*heart$age +
+   coef(fit7_4b)[2]*(heart$anterior!='Anterior')
> Y2<- 1/(1+exp(-L2))
> layout(1)
> plot(heart$age, Y1, col=1, main='P[Death] within 48 hr admission',
+   xlab='Age')
> lines(heart$age, Y2, col=2, type='p')
```

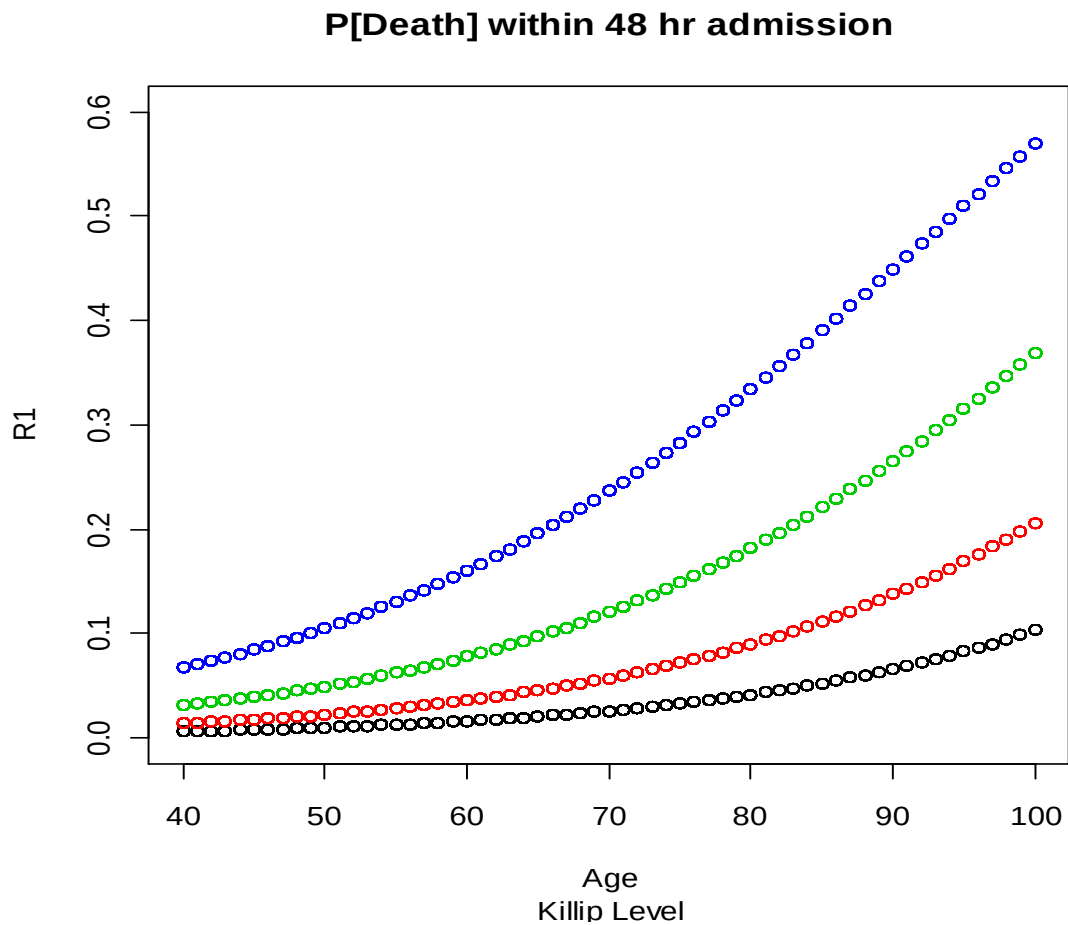


```
> knum<- as.numeric(heart$killip) #convert factor to level values
> fit7_4c<- glm(death ~ age + knum, data=heart, family=binomial)
> K1<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
+   coef(fit7_4c)[3]*1
> R1<- 1/(1+exp(-K1))
> K2<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
+   coef(fit7_4c)[3]*2
> R2<- 1/(1+exp(-K2))
> K3<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
+   coef(fit7_4c)[3]*3
> R3<- 1/(1+exp(-K3))
> K4<- coef(fit7_4c)[1] + coef(fit7_4c)[2]*heart$age +
+   coef(fit7_4c)[3]*4
> R4<- 1/(1+exp(-K4))
```

```

> layout(1)
> plot(heart$age, R1, col=1, main='P[Death] within 48 hr admission',
+      sub='Killip Level', xlab='Age', ylim=c(0,0.6))
> lines(heart$age, R2, col=2, type='p')
> lines(heart$age, R3, col=3, type='p')
> lines(heart$age, R4, col=4, type='p')

```



SECTION 7.5:

```
#SECTION 7.5: VALIDATION MODELS
fit7_5a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart, family=binomial)
summary(fit7_5a)
deviance(fit7_5a)/fit7_5a$df.residual #dispersion

i20<- sample(1:nrow(heart), round(0.2*nrow(heart)), replace=FALSE)
heart20<- heart[i20,]
fit7_5b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart20, family=binomial)
summary(fit7_5b)
deviance(fit7_5b)/fit7_5b$df.residual #dispersion
```

OUTPUT: 7.5

```
> #SECTION 7.5: VALIDATION MODELS
> fit7_5a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart, family=binomial)
> summary(fit7_5a)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial, data = heart)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.8159     0.1934 -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg          0.7864     0.3527   2.229 0.025786 *
kk2            0.8249     0.1804   4.572 4.84e-06 ***
kk3            0.7967     0.2692   2.959 0.003083 **
kk4            2.6837     0.3565   7.529 5.11e-14 ***
age3           1.2668     0.2006   6.316 2.68e-10 ***
age4           1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3

Number of Fisher Scoring iterations: 7

> deviance(fit7_5a)/fit7_5a$df.residual #dispersion
[1] 0.2839420

> i20<- sample(1:nrow(heart), round(0.2*nrow(heart)), replace=FALSE)
> heart20<- heart[i20,]
> fit7_5b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart20, family=binomial)
> summary(fit7_5b)
```

```

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial, data = heart20)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3341  -0.3197  -0.1754  -0.1358   3.0631

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.6820     0.4002  -11.700 < 2e-16 ***
anteriorAnterior  0.5151     0.3577   1.440  0.14990
hcabg           1.3285     0.6032   2.202  0.02764 *
kk2              0.8338     0.4078   2.045  0.04087 *
kk3              1.0663     0.6015   1.773  0.07627 .
kk4              2.7001     0.6138   4.399 1.09e-05 ***
age3             1.2186     0.4282   2.846  0.00443 **
age4             1.8280     0.4502   4.060 4.90e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 322.26  on 912  degrees of freedom
Residual deviance: 268.58  on 905  degrees of freedom
(165 observations deleted due to missingness)
AIC: 284.58

Number of Fisher Scoring iterations: 7

> deviance(fit7_5b)/fit7_5b$df.residual #dispersion
[1] 0.2967742

```

Chapter 8

```

#CHAPTER 8: BINOMIAL LOGISTIC REGRESSION
eg<- data.frame(y=c(1,1,2,0,2,0), cases=c(3,1,2,1,2,1),
    x1=c(1,1,0,0,1,0), x2=c(0,1,0,1,0,1), x3=c(1,1,1,1,0,0))
eg$noty<- eg$cases - eg$y #get failures
fit8a<- glm(cbind(y, noty) ~ x1 + x2 + x3, data=eg,
family=binomial(link=logit))
summary(fit8a)
exp(coef(fit8a)) #ORs

binomialX2<- function(y, m, e) { #Compute Pearson chi-square for binomial
    #y: number successes
    #m: number trials
    #e: expected probability
    return(sum(m*(y/m-e)^2/(e*(1-e))))
}

fit8aX2<- binomialX2(eg$y, eg$cases, fitted(fit8a, type='response'))
cat('Pearson X2:', fit8aX2, 'Dispersion:', fit8aX2/fit8a$df.residual, '\n')

library('foreign')
heart<- read.dta('heart01.dta') #read Stata format file

```

```

heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that complicate
tabulations
heart$center<- factor(heart$center) #convert to factor from numeric levels
heart$killip<- factor(heart$killip) #convert to factor from numeric levels

fit8b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart, family=binomial)
summary(fit8b)
exp(coef(fit8b)) #ORs

library(reshape) #for cast(), melt()
heart2 <- na.omit(data.frame(cast(melt(heart, measure="death"),
  anterior + hcabg + kk1 + kk2 + kk3 + kk4 + age1 + age2 + age3 + age4 + center
~ .,
  function(x) { c(alive=sum(x == 0), dead=sum(x == 1)) })))

fit8c<- glm(cbind(dead, alive) ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart2, family=binomial)
summary(fit8c)
exp(coef(fit8c)) #ORs
exp(confint(fit8c)) #CI for ORs
fit8cX2<- binomialX2(heart2$dead, (heart2$dead+heart2$alive),
  fitted(fit8c, type='response'))
cat('Pearson X2:', fit8cX2, 'Dispersion:', fit8cX2/fit8c$df.residual, '\n')

fit8d<- glm(cbind(dead, alive) ~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4, data=heart2, family=quasibinomial)
summary(fit8d)
exp(coef(fit8d)) #ORs
fit8dX2<- binomialX2(heart2$dead, (heart2$dead+heart2$alive),
  fitted(fit8d, type='response'))
cat('Pearson X2:', fit8dX2, 'Dispersion:', fit8dX2/fit8d$df.residual, '\n')

```

OUTPUT: Ch 8

```

> #CHAPTER 8: BINOMIAL LOGISTIC REGRESSION
> eg<- data.frame(y=c(1,1,2,0,2,0), cases=c(3,1,2,1,2,1),
+   x1=c(1,1,0,0,1,0), x2=c(0,1,0,1,0,1), x3=c(1,1,1,1,0,0))
> eg$noty<- eg$cases - eg$y #get failures
> fit8a<- glm(cbind(y, noty) ~ x1 + x2 + x3, data=eg,
family=binomial(link=logit))
> summary(fit8a)

Call:
glm(formula = cbind(y, noty) ~ x1 + x2 + x3, family = binomial(link = logit),
    data = eg)

Deviance Residuals:
    1      2      3      4      5      6
-1.2902  1.5172  1.2935 -0.8114  0.9490 -1.0160

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.2050     1.8348   0.657   0.511
x1             0.1714     1.4909   0.115   0.908
x2            -1.5972     1.6011  -0.998   0.318
x3            -0.5499     1.5817  -0.348   0.728

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 9.6411  on 5  degrees of freedom

```

Residual deviance: 8.2310 on 2 degrees of freedom
AIC: 17.853

Number of Fisher Scoring iterations: 4

```
> exp(coef(fit8a)) #ORs
(Intercept)      x1      x2      x3
  3.3367121  1.1869473  0.2024631  0.5770337

> binomialX2<- function(y, m, e) { #Compute Pearson chi-square for binomial
+   #y: number successes
+   #m: number trials
+   #e: expected probability
+   return(sum(m*(y/m-e)^2/(e*(1-e))))
+ }

> fit8aX2<- binomialX2(eg$y, eg$cases, fitted(fit8a, type='response'))
> cat('Pearson X2:', fit8aX2, 'Dispersion:', fit8aX2/fit8a$df.residual, '\n')
Pearson X2: 6.630003 Dispersion: 3.315001

> library('foreign')
> heart<- read.dta('heart01.dta') #read Stata format file
> heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that
complicate tabulations
> heart$center<- factor(heart$center) #convert to factor from numeric levels
> heart$killip<- factor(heart$killip) #convert to factor from numeric levels
> fit8b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart, family=binomial)
> summary(fit8b)
```

Call:

```
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial, data = heart)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.3724	-0.3257	-0.1745	-0.1270	3.1061

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.8159	0.1934	-24.896	< 2e-16 ***
anteriorAnterior	0.6387	0.1675	3.812	0.000138 ***
hcabg	0.7864	0.3527	2.229	0.025786 *
kk2	0.8249	0.1804	4.572	4.84e-06 ***
kk3	0.7967	0.2692	2.959	0.003083 **
kk4	2.6837	0.3565	7.529	5.11e-14 ***
age3	1.2668	0.2006	6.316	2.68e-10 ***
age4	1.9409	0.2080	9.329	< 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: 1486.2 on 4502 degrees of freedom
Residual deviance: 1276.3 on 4495 degrees of freedom
(885 observations deleted due to missingness)
AIC: 1292.3

Number of Fisher Scoring iterations: 7

```
> exp(coef(fit8b)) #ORs
(Intercept) anteriorAnterior      hcabg      kk2
  0.008099558      1.894095884  2.195519114  2.281692192
```



```

                kk3                kk4                age3                age4
2.218198595      14.639836908      3.549576517      6.964846706

> library(reshape) #for cast(), melt()
> heart2 <- na.omit(data.frame(cast(melt(heart, measure="death"),
+   anterior + hcabg + kk1 + kk2 + kk3 + kk4 + age1 + age2 + age3 + age4 +
center ~ .,
+   function(x) { c(alive=sum(x == 0), dead=sum(x == 1)) })))
> fit8c<- glm(cbind(dead, alive) ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart2, family=binomial)
> summary(fit8c)

Call:
glm(formula = cbind(dead, alive) ~ anterior + hcabg + kk2 + kk3 +
    kk4 + age3 + age4, family = binomial, data = heart2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4945  -0.6369  -0.3848  -0.1914   2.9400

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8159     0.1934 -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg           0.7864     0.3527   2.229 0.025786 *
kk2             0.8249     0.1804   4.572 4.84e-06 ***
kk3             0.7967     0.2692   2.959 0.003083 **
kk4             2.6837     0.3565   7.529 5.11e-14 ***
age3            1.2668     0.2006   6.316 2.68e-10 ***
age4            1.9409     0.2080   9.329 < 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: 826.1  on 894  degrees of freedom
Residual deviance: 616.2  on 887  degrees of freedom
AIC: 836.42

Number of Fisher Scoring iterations: 6

> exp(coef(fit8c)) #ORs
                (Intercept) anteriorAnterior                hcabg                kk2
0.008099558      1.894095884      2.195519114      2.281692192
                kk3                kk4                age3                age4
2.218198595      14.639836908      3.549576517      6.964846706
> exp(confint(fit8c)) #CI for ORs
Waiting for profiling to be done...
                2.5 %                97.5 %
(Intercept)    0.005461512    0.01166807
anteriorAnterior 1.368418117    2.64220438
hcabg           1.037061789    4.19021301
kk2             1.595839214    3.24176424
kk3             1.278043511    3.68924231
kk4             7.154252752    29.13630155
age3            2.404385923    5.28930691
age4            4.645436584    10.52267115

> fit8cX2<- binomialX2(heart2$dead, (heart2$dead+heart2$alive),
+   fitted(fit8c, type='response'))
> cat('Pearson X2:', fit8cX2, 'Dispersion:', fit8cX2/fit8c$df.residual, '\n')
Pearson X2: 936.504 Dispersion: 1.055811

```

```

> fit8d<- glm(cbind(dead, alive) ~ anterior + hcabg + kk2 + kk3 + kk4 +
+   age3 + age4, data=heart2, family=quasibinomial)
> summary(fit8d)

Call:
glm(formula = cbind(dead, alive) ~ anterior + hcabg + kk2 + kk3 +
    kk4 + age3 + age4, family = quasibinomial, data = heart2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4945  -0.6369  -0.3848  -0.1914   2.9400

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -4.8159     0.1988  -24.229  < 2e-16 ***
anteriorAnterior  0.6387     0.1722   3.710  0.00022 ***
hcabg           0.7864     0.3625   2.170  0.03030 *
kk2             0.8249     0.1854   4.449  9.71e-06 ***
kk3             0.7967     0.2766   2.880  0.00407 **
kk4             2.6837     0.3663   7.327  5.28e-13 ***
age3            1.2668     0.2061   6.147  1.19e-09 ***
age4            1.9409     0.2138   9.079  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 1.055811)

Null deviance: 826.1  on 894  degrees of freedom
Residual deviance: 616.2  on 887  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6

> exp(coef(fit8d))  #ORs
      (Intercept) anteriorAnterior      hcabg      kk2
0.008099558      1.894095884      2.195519114      2.281692192
      kk3      kk4      age3      age4
2.218198595      14.639836908      3.549576517      6.964846706

> fit8dX2<- binomialX2(heart2$dead, (heart2$dead+heart2$alive),
+   fitted(fit8d, type='response'))
> cat('Pearson X2:', fit8dX2, 'Dispersion:', fit8dX2/fit8d$df.residual, '\n')
Pearson X2: 936.504 Dispersion: 1.055811

```

Chapter 9

SECTION 9.1: no examples

SECTION 9.2:

```

#SECTION 9.2: BINOMIAL OVERDISPERSION
library('foreign')
simul<- read.dta('overexx.dta') #read Stata format file
simul$noty<- simul$d - simul$y #numbers of failures
simul$notyi<- simul$d - simul$yi #numbers of failures
simul$notysq<- simul$d - simul$ysq #numbers of failures
simul$notyp<- simul$d - simul$yp #numbers of failures

```

```

summary(simul)
head(simul)

source('pearsonX2.r') #reinstall binomialX2()

fit9_2a<- glm(cbind(y, noty) ~ x1 + x2 + x3, data=simul, family=binomial)
summary(fit9_2a)
fit9_2aX2<- binomialX2(simul$y, simul$d, fitted(fit9_2a, type='response'))
cat('Pearson X2:', fit9_2aX2, 'Dispersion:', fit9_2aX2/fit9_2a$df.residual,
    'AIC/d.f.:', fit9_2a$aic/fit9_2a$df.residual, '\n')

fit9_2b<- glm(cbind(y, noty) ~ x2 + x3, data=simul, family=binomial)
summary(fit9_2b)
fit9_2bX2<- binomialX2(simul$y, simul$d, fitted(fit9_2b, type='response'))
cat('Pearson X2:', fit9_2bX2, 'Dispersion:', fit9_2bX2/fit9_2b$df.residual,
    'AIC/d.f.:', fit9_2b$aic/fit9_2b$df.residual, '\n')

fit9_2c<- glm(cbind(yi, notyi) ~ x1 + x2 + x3 + x23, data=simul,
family=binomial)
summary(fit9_2c)
fit9_2cX2<- binomialX2(simul$yi, simul$d, fitted(fit9_2c, type='response'))
cat('Pearson X2:', fit9_2cX2, 'Dispersion:', fit9_2cX2/fit9_2c$df.residual,
    'AIC/d.f.:', fit9_2c$aic/fit9_2c$df.residual, '\n')

fit9_2d<- glm(cbind(yi, notyi) ~ x1 + x2 + x3, data=simul, family=binomial)
summary(fit9_2d)
fit9_2dX2<- binomialX2(simul$yi, simul$d, fitted(fit9_2d, type='response'))
cat('Pearson X2:', fit9_2dX2, 'Dispersion:', fit9_2dX2/fit9_2d$df.residual,
    'AIC/d.f.:', fit9_2d$aic/fit9_2d$df.residual, '\n')

fit9_2e<- glm(cbind(ysq, notysq) ~ x1sq + x2 + x3, data=simul, family=binomial)
summary(fit9_2e)
fit9_2eX2<- binomialX2(simul$ysq, simul$d, fitted(fit9_2e, type='response'))
cat('Pearson X2:', fit9_2eX2, 'Dispersion:', fit9_2eX2/fit9_2e$df.residual,
    'AIC/d.f.:', fit9_2e$aic/fit9_2e$df.residual, '\n')

fit9_2f<- glm(cbind(ysq, notysq) ~ x1 + x2 + x3, data=simul, family=binomial)
summary(fit9_2f)
fit9_2fX2<- binomialX2(simul$ysq, simul$d, fitted(fit9_2f, type='response'))
cat('Pearson X2:', fit9_2fX2, 'Dispersion:', fit9_2fX2/fit9_2f$df.residual,
    'AIC/d.f.:', fit9_2f$aic/fit9_2f$df.residual, '\n')

fit9_2g<- glm(cbind(yp, notyp) ~ x1 + x2 + x3, data=simul,
    family=binomial(link=probit))
summary(fit9_2g)
fit9_2gX2<- binomialX2(simul$yp, simul$d, fitted(fit9_2g, type='response'))
cat('Pearson X2:', fit9_2gX2, 'Dispersion:', fit9_2gX2/fit9_2g$df.residual,
    'AIC/d.f.:', fit9_2g$aic/fit9_2g$df.residual, '\n')

fit9_2h<- glm(cbind(yp, notyp) ~ x1 + x2 + x3, data=simul,
    family=binomial(link=logit))
summary(fit9_2h)
fit9_2hX2<- binomialX2(simul$yp, simul$d, fitted(fit9_2h, type='response'))
cat('Pearson X2:', fit9_2hX2, 'Dispersion:', fit9_2hX2/fit9_2h$df.residual,
    'AIC/d.f.:', fit9_2h$aic/fit9_2h$df.residual, '\n')

cancer<- read.dta('cancerx.dta') #read Stata format file
cancer$drug<- factor(cancer$drug) #convert to factor
summary(cancer)
head(cancer)
nrow(cancer)

fit9_2i<- glm(studytim ~ drug, data=cancer, family=poisson)

```

```

poissonX2 = function(y, e) { #Compute Pearson chi-square for poisson
  #y: number successes
  #e: expected probability
  return(sum((y-e)^2/e))
}
summary(fit9_2i)
fit9_2iX2<- poissonX2(cancer$studytim, fitted(fit9_2i, type='response'))
cat('Pearson X2:', fit9_2iX2, 'Dispersion:', fit9_2iX2/fit9_2i$df.residual,
    'AIC/d.f.:', fit9_2i$aic/fit9_2i$df.residual, '\n')

cancer$time<- cancer$studytim #make copy
cancer$time[cancer$studytim==39]<- 390 #create outliers
fit9_2j<- glm(time ~ drug, data=cancer, family=poisson)
summary(fit9_2j)
fit9_2jX2<- poissonX2(cancer$time, fitted(fit9_2j, type='response'))
cat('Pearson X2:', fit9_2jX2, 'Dispersion:', fit9_2jX2/fit9_2j$df.residual,
    'AIC/d.f.:', fit9_2j$aic/fit9_2j$df.residual, '\n')
cancer[cancer$time==390,] #show rows for time=390

summary(cancer$age) #show stats
cat('S.D.:', sd(cancer$age), 'S.E. mean:',
    sd(cancer$age)/sqrt(length(cancer$age)), '\n')

fit9_2k<- glm(died ~ age, data=cancer, family=binomial)
summary(fit9_2k)
fit9_2kX2<- binomialX2(cancer$died, rep(1, nrow(cancer)),
  fitted(fit9_2k, type='response'))
cat('Pearson X2:', fit9_2kX2, 'Dispersion:', fit9_2kX2/fit9_2k$df.residual,
    'AIC/d.f.:', fit9_2k$aic/fit9_2k$df.residual, '\n')

quantile(cancer$age, probs=0.75)
sum(cancer$age==60)
which(cancer$age==60)
cancer$age[37]<- 600 #make outlier

fit9_2l<- glm(died ~ age, data=cancer, family=binomial)
summary(fit9_2l)
fit9_2lX2<- binomialX2(cancer$died, rep(1, nrow(cancer)),
  fitted(fit9_2l, type='response'))
cat('Pearson X2:', fit9_2lX2, 'Dispersion:', fit9_2lX2/fit9_2l$df.residual,
    'AIC/d.f.:', fit9_2l$aic/fit9_2l$df.residual, '\n')
exp(coef(fit9_2l)) #ORs
exp(coef(fit9_2k)) #ORs

yd<- simul$y/simul$d #proportions
summary(yd) #stats
cat('S.D.:', sd(yd), 'S.E. mean:', sd(yd)/sqrt(length(yd)), '\n')

summary(fit9_2a) #show previous fit again
cat('Pearson X2:', fit9_2aX2, 'Dispersion:', fit9_2aX2/fit9_2a$df.residual,
    'AIC/d.f.:', fit9_2a$aic/fit9_2a$df.residual, '\n')

fit9_2m<- glm(y ~ x1 + x2 + x3, data=simul, offset=d, family=poisson)
summary(fit9_2m)
fit9_2mX2<- poissonX2(simul$y, fitted(fit9_2m, type='response'))
cat('Pearson X2:', fit9_2mX2, 'Dispersion:', fit9_2mX2/fit9_2m$df.residual,
    'AIC/d.f.:', fit9_2m$aic/fit9_2m$df.residual, '\n')

require('MASS')
fit9_2n<- glm.nb(y ~ x1 + x2 + x3 + offset(d), data=simul)
summary(fit9_2n)
fit9_2nX2<- binomialX2(simul$y, simul$d, fitted(fit9_2n, type='response'))
cat('Pearson X2:', fit9_2nX2, 'Dispersion:', fit9_2nX2/fit9_2n$df.residual,

```

```
'AIC/d.f.:', fit9_2n$aic/fit9_2n$df.residual, '\n')
```

OUTPUT: CH 9.2

```
> #SECTION 9.2: BINOMIAL OVERDISPERSION
> library('foreign')
> simul<- read.dta('overexx.dta') #read Stata format file
> simul$noty<- simul$d - simul$y #numbers of failures
> simul$notyi<- simul$d - simul$yi #numbers of failures
> simul$notysq<- simul$d - simul$ysq #numbers of failures
> simul$notyp<- simul$d - simul$yp #numbers of failures
> summary(simul)
```

x1		x2		x3		d	
Min.	:-4.116852	Min.	:-4.057574	Min.	:-3.951092	Min.	:50.0
1st Qu.	:-0.678305	1st Qu.	:-0.681405	1st Qu.	:-0.681466	1st Qu.	:60.0
Median	: 0.018527	Median	: 0.007926	Median	: 0.008751	Median	:72.5
Mean	: 0.004121	Mean	: 0.014198	Mean	:-0.008423	Mean	:72.5
3rd Qu.	: 0.678992	3rd Qu.	: 0.702251	3rd Qu.	: 0.669799	3rd Qu.	:85.0
Max.	: 3.617816	Max.	: 4.089732	Max.	: 3.997190	Max.	:95.0

xb		y		x23		xbi	
Min.	:-7.3710	Min.	: 0.0	Min.	:-7.914881	Min.	:-7.8080
1st Qu.	:-0.6167	1st Qu.	:24.0	1st Qu.	:-0.352768	1st Qu.	:-0.6038
Median	: 0.5020	Median	:43.0	Median	: 0.002682	Median	: 0.5231
Mean	: 0.4843	Mean	:42.2	Mean	: 0.018788	Mean	: 0.4880
3rd Qu.	: 1.5988	3rd Qu.	:59.0	3rd Qu.	: 0.374929	3rd Qu.	: 1.6164
Max.	: 6.1114	Max.	:95.0	Max.	: 7.857721	Max.	: 6.4930

yi		xlsq		xbsq		ysq	
Min.	: 0.00	Min.	:2.822e-10	Min.	:-4.379016	Min.	: 0.00
1st Qu.	:24.00	1st Qu.	:1.024e-01	1st Qu.	:-0.009142	1st Qu.	:34.00
Median	:43.00	Median	:4.603e-01	Median	: 0.938693	Median	:49.00
Mean	:42.41	Mean	:1.008e+00	Mean	: 0.984310	Mean	:48.34
3rd Qu.	:59.00	3rd Qu.	:1.340e+00	3rd Qu.	: 1.923928	3rd Qu.	:63.00
Max.	:95.00	Max.	:1.695e+01	Max.	: 8.110692	Max.	:95.00

yp		yc		noty		notyi		notysq	
Min.	: 0.00	Min.	: 0.00	Min.	: 0.0	Min.	: 0.00	Min.	: 0.00
1st Qu.	:18.00	1st Qu.	:29.00	1st Qu.	:12.0	1st Qu.	:11.00	1st Qu.	: 9.00
Median	:47.00	Median	:52.00	Median	:26.0	Median	:26.00	Median	:20.00
Mean	:43.56	Mean	:50.06	Mean	:30.3	Mean	:30.09	Mean	:24.16
3rd Qu.	:65.00	3rd Qu.	:70.00	3rd Qu.	:46.0	3rd Qu.	:46.00	3rd Qu.	:36.00
Max.	:95.00	Max.	:95.00	Max.	:94.0	Max.	:94.00	Max.	:94.00


```
notyp
Min. : 0.00
1st Qu.: 4.00
Median :22.00
Mean :28.94
3rd Qu.:50.00
Max. :95.00
> head(simul)
```

	x1	x2	x3	d	xb	y	x23	xbi	yi
1	0.50633353	0.19542192	0.1617462	50	0.8024927	30	0.03160875	0.8088144	36
2	-0.17592467	-1.27175009	-0.1025323	50	1.8881298	44	0.13039549	1.9142090	41
3	-2.49378181	-0.60230917	-0.6137104	50	-1.3943230	10	0.36964339	-1.3203943	8
4	-0.50829357	-0.04621129	1.2532927	50	0.3627937	32	-0.05791628	0.3512105	27
5	0.52022028	-0.86786199	-1.0278013	50	1.8480974	44	0.89198965	2.0264955	41
6	0.05115141	0.45371607	-0.7813335	50	-0.2113270	17	-0.35450357	-0.2822278	26

	xlsq	xbsq	ysq	yp	yc	noty	notyi	notysq	notyp
1	0.256373644	0.4243460	25	38	48	20	14	25	12
2	0.030949490	2.0795293	39	48	50	6	9	11	2
3	6.218947887	4.2089329	48	5	10	40	42	2	45
4	0.258362353	1.0002685	37	36	37	18	23	13	14

```

5 0.270629138 1.4631917 42 50 50 6 9 8 0
6 0.002616467 -0.2611702 23 22 27 33 24 27 28

> source('pearsonX2.r') #reinstall binomialX2()
> fit9_2a<- glm(cbind(y, noty) ~ x1 + x2 + x3, data=simul, family=binomial)
> summary(fit9_2a)

Call:
glm(formula = cbind(y, noty) ~ x1 + x2 + x3, family = binomial,
    data = simul)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.81009  -0.66364   0.01385   0.70967   3.44838

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.498530   0.002955  168.7  <2e-16 ***
x1           0.999543   0.003373  296.3  <2e-16 ***
x2          -1.245868   0.003597 -346.3  <2e-16 ***
x3           0.249151   0.002952   84.4  <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: 270037  on 9999  degrees of freedom
Residual deviance: 10097  on 9996  degrees of freedom
AIC: 51172

Number of Fisher Scoring iterations: 4

> fit9_2aX2<- binomialX2(simul$y, simul$d, fitted(fit9_2a, type='response'))
> cat('Pearson X2:', fit9_2aX2, 'Dispersion:', fit9_2aX2/fit9_2a$df.residual,
+     'AIC/d.f.:', fit9_2a$aic/fit9_2a$df.residual, '\n')
Pearson X2: 9906.11 Dispersion: 0.9910073 AIC/d.f.: 5.119241

> fit9_2b<- glm(cbind(y, noty) ~ x2 + x3, data=simul, family=binomial)
> summary(fit9_2b)

Call:
glm(formula = cbind(y, noty) ~ x2 + x3, family = binomial, data = simul)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-13.4433  -2.3589   0.3346   2.5796  12.7210

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.423406   0.002681  157.90  <2e-16 ***
x2          -1.046665   0.003158 -331.48  <2e-16 ***
x3           0.198955   0.002697   73.77  <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: 270037  on 9999  degrees of freedom
Residual deviance: 121319  on 9997  degrees of freedom
AIC: 162392

Number of Fisher Scoring iterations: 4

```

```
> fit9_2bX2<- binomialX2(simul$y, simul$d, fitted(fit9_2b, type='response'))
> cat('Pearson X2:', fit9_2bX2, 'Dispersion:', fit9_2bX2/fit9_2b$df.residual,
+ 'AIC/d.f.:', fit9_2b$aic/fit9_2b$df.residual, '\n')
Pearson X2: 117527.2 Dispersion: 11.75625 AIC/d.f.: 16.24404
```

```
> fit9_2c<- glm(cbind(yi, notyi) ~ x1 + x2 + x3 + x23, data=simul,
family=binomial)
> summary(fit9_2c)
```

Call:

```
glm(formula = cbind(yi, notyi) ~ x1 + x2 + x3 + x23, family = binomial,
data = simul)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-3.65749	-0.68259	0.01596	0.69428	3.77910

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.504371	0.002975	169.51	<2e-16 ***
x1	1.002341	0.003389	295.75	<2e-16 ***
x2	-1.255015	0.003638	-345.00	<2e-16 ***
x3	0.256492	0.003014	85.09	<2e-16 ***
x23	0.199889	0.003512	56.92	<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: 274679 on 9999 degrees of freedom
Residual deviance: 10380 on 9995 degrees of freedom
AIC: 51254

Number of Fisher Scoring iterations: 4

```
> fit9_2cX2<- binomialX2(simul$yi, simul$d, fitted(fit9_2c, type='response'))
> cat('Pearson X2:', fit9_2cX2, 'Dispersion:', fit9_2cX2/fit9_2c$df.residual,
+ 'AIC/d.f.:', fit9_2c$aic/fit9_2c$df.residual, '\n')
Pearson X2: 10105.69 Dispersion: 1.011074 AIC/d.f.: 5.127943
```

```
> fit9_2d<- glm(cbind(yi, notyi) ~ x1 + x2 + x3, data=simul, family=binomial)
> summary(fit9_2d)
```

Call:

```
glm(formula = cbind(yi, notyi) ~ x1 + x2 + x3, family = binomial,
data = simul)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-4.914196	-0.784267	-0.003704	0.756701	4.843536

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.517251	0.002966	174.40	<2e-16 ***
x1	1.000164	0.003381	295.86	<2e-16 ***
x2	-1.249446	0.003607	-346.38	<2e-16 ***
x3	0.281058	0.002966	94.74	<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: 274679 on 9999 degrees of freedom

```
Residual deviance: 13679 on 9996 degrees of freedom
AIC: 54550
```

```
Number of Fisher Scoring iterations: 4
```

```
> fit9_2dX2<- binomialX2(simul$yi, simul$d, fitted(fit9_2d, type='response'))
> cat('Pearson X2:', fit9_2dX2, 'Dispersion:', fit9_2dX2/fit9_2d$df.residual,
+ 'AIC/d.f.:', fit9_2d$aic/fit9_2d$df.residual, '\n')
Pearson X2: 13408.44 Dispersion: 1.341380 AIC/d.f.: 5.457203
```

```
> fit9_2e<- glm(cbind(ysq, notysq) ~ x1sq + x2 + x3, data=simul,
family=binomial)
> summary(fit9_2e)
```

```
Call:
```

```
glm(formula = cbind(ysq, notysq) ~ x1sq + x2 + x3, family = binomial,
data = simul)
```

```
Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-3.97013	-0.64430	0.02335	0.71769	3.96765

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.500260	0.003700	135.20	<2e-16 ***
x1sq	0.495654	0.003041	163.01	<2e-16 ***
x2	-1.246583	0.003618	-344.53	<2e-16 ***
x3	0.247263	0.002958	83.59	<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: 214597 on 9999 degrees of freedom
Residual deviance: 10340 on 9996 degrees of freedom
AIC: 51033
```

```
Number of Fisher Scoring iterations: 4
```

```
> fit9_2eX2<- binomialX2(simul$ysq, simul$d, fitted(fit9_2e, type='response'))
> cat('Pearson X2:', fit9_2eX2, 'Dispersion:', fit9_2eX2/fit9_2e$df.residual,
+ 'AIC/d.f.:', fit9_2e$aic/fit9_2e$df.residual, '\n')
Pearson X2: 10177.22 Dispersion: 1.018129 AIC/d.f.: 5.10532
```

```
> fit9_2f<- glm(cbind(ysq, notysq) ~ x1 + x2 + x3, data=simul, family=binomial)
> summary(fit9_2f)
```

```
Call:
```

```
glm(formula = cbind(ysq, notysq) ~ x1 + x2 + x3, family = binomial,
data = simul)
```

```
Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-5.0639	-1.3708	-0.3061	1.1091	21.8245

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.910365	0.002996	303.844	<2e-16 ***
x1	0.002378	0.002827	0.841	0.4
x2	-1.190336	0.003465	-343.515	<2e-16 ***
x3	0.232204	0.002875	80.763	<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: 214597 on 9999 degrees of freedom
Residual deviance: 46880 on 9996 degrees of freedom
AIC: 87572

Number of Fisher Scoring iterations: 4

```
> fit9_2fX2<- binomialX2(simul$ysq, simul$d, fitted(fit9_2f, type='response'))
> cat('Pearson X2:', fit9_2fX2, 'Dispersion:', fit9_2fX2/fit9_2f$df.residual,
+ 'AIC/d.f.:', fit9_2f$aic/fit9_2f$df.residual, '\n')
Pearson X2: 44215.15 Dispersion: 4.423284 AIC/d.f.: 8.760743
```

```
> fit9_2g<- glm(cbind(yp, notyp) ~ x1 + x2 + x3, data=simul,
+ family=binomial(link=probit))
> summary(fit9_2g)
```

Call:

```
glm(formula = cbind(yp, notyp) ~ x1 + x2 + x3, family = binomial(link =
probit),
data = simul)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.96946	-0.62141	0.06853	0.68208	4.16980

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.498844	0.002103	237.2	<2e-16 ***
x1	0.999848	0.002628	380.5	<2e-16 ***
x2	-1.244400	0.002912	-427.3	<2e-16 ***
x3	0.247808	0.002045	121.2	<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: 461561.7 on 9999 degrees of freedom
Residual deviance: 9700.7 on 9996 degrees of freedom
AIC: 41595

Number of Fisher Scoring iterations: 4

```
> fit9_2gX2<- binomialX2(simul$yp, simul$d, fitted(fit9_2g, type='response'))
> cat('Pearson X2:', fit9_2gX2, 'Dispersion:', fit9_2gX2/fit9_2g$df.residual,
+ 'AIC/d.f.:', fit9_2g$aic/fit9_2g$df.residual, '\n')
Pearson X2: 9802.15 Dispersion: 0.9806073 AIC/d.f.: 4.161138
```

```
> fit9_2h<- glm(cbind(yp, notyp) ~ x1 + x2 + x3, data=simul,
+ family=binomial(link=logit))
> summary(fit9_2h)
```

Call:

```
glm(formula = cbind(yp, notyp) ~ x1 + x2 + x3, family = binomial(link = logit),
data = simul)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8928	-0.6502	0.1187	0.7929	4.4498

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
--	----------	------------	---------	----------

```

(Intercept)  0.873293    0.003787    230.6    <2e-16 ***
x1           1.751515    0.004935    355.0    <2e-16 ***
x2          -2.180025    0.005562   -392.0    <2e-16 ***
x3           0.433979    0.003612    120.1    <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: 461562  on 9999  degrees of freedom
Residual deviance: 10566  on 9996  degrees of freedom
AIC: 42460

Number of Fisher Scoring iterations: 4

> fit9_2hX2<- binomialX2(simul$yp, simul$d, fitted(fit9_2h, type='response'))
> cat('Pearson X2:', fit9_2hX2, 'Dispersion:', fit9_2hX2/fit9_2h$df.residual,
+     'AIC/d.f.: ', fit9_2h$aic/fit9_2h$df.residual, '\n')
Pearson X2: 9489.328 Dispersion: 0.9493125 AIC/d.f.: 4.247658

> cancer<- read.dta('cancerx.dta') #read Stata format file
> cancer$drug<- factor(cancer$drug) #convert to factor
> summary(cancer)
      studytim      died      drug      age
Min.   : 1.00   Min.   :0.0000   1:20   Min.   :47.00
1st Qu.: 7.75   1st Qu.:0.0000   2:14   1st Qu.:50.75
Median :12.50   Median :1.0000   3:14   Median :56.00
Mean   :15.50   Mean   :0.6458             Mean   :55.88
3rd Qu.:23.00   3rd Qu.:1.0000             3rd Qu.:60.00
Max.   :39.00   Max.   :1.0000             Max.   :67.00
> head(cancer)
  studytim died drug age
1         1   1    1  61
2         1   1    1  65
3         2   1    1  59
4         3   1    1  52
5         4   1    1  56
6         4   1    1  67

> nrow(cancer)
[1] 48

> fit9_2i<- glm(studytim ~ drug, data=cancer, family=poisson)
> summary(fit9_2i)

Call:
glm(formula = studytim ~ drug, family = poisson, data = cancer)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.6280 -1.6849 -0.1715  1.2521  3.8936

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.19722    0.07454  29.479  < 2e-16 ***
drug2        0.50605    0.10169   4.977 6.47e-07 ***
drug3        1.03584    0.09150  11.320  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 330.47  on 47  degrees of freedom

```

```
Residual deviance: 192.40 on 45 degrees of freedom
AIC: 405.32
```

```
Number of Fisher Scoring iterations: 5
```

```
> fit9_2iX2<- poissonX2(cancer$studytim, fitted(fit9_2i, type='response'))
> cat('Pearson X2:', fit9_2iX2, 'Dispersion:', fit9_2iX2/fit9_2i$df.residual,
+ 'AIC/d.f.:', fit9_2i$aic/fit9_2i$df.residual, '\n')
Pearson X2: 185.1356 Dispersion: 4.114123 AIC/d.f.: 9.007198
```

```
> cancer$time<- cancer$studytim #make copy
> cancer$time[cancer$studytim==39]<- 390 #create outliers
> fit9_2j<- glm(time ~ drug, data=cancer, family=poisson)
> summary(fit9_2j)
```

```
Call:
```

```
glm(formula = time ~ drug, family = poisson, data = cancer)
```

```
Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-7.9569	-2.7941	-1.4568	0.7208	30.2724

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.19722	0.07454	29.479	< 2e-16 ***
drug2	0.50605	0.10169	4.977	6.47e-07 ***
drug3	1.72333	0.08350	20.639	< 2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for poisson family taken to be 1)
```

```
Null deviance: 1927.9 on 47 degrees of freedom
Residual deviance: 1319.9 on 45 degrees of freedom
AIC: 1535.1
```

```
Number of Fisher Scoring iterations: 6
```

```
> fit9_2jX2<- poissonX2(cancer$time, fitted(fit9_2j, type='response'))
> cat('Pearson X2:', fit9_2jX2, 'Dispersion:', fit9_2jX2/fit9_2j$df.residual,
+ 'AIC/d.f.:', fit9_2j$aic/fit9_2j$df.residual, '\n')
Pearson X2: 2620.232 Dispersion: 58.22738 AIC/d.f.: 34.11343
```

```
> cancer[cancer$time==390,] #show rows for time=390
  studytim died drug age time
48      39    0    3  52  390
```

```
> summary(cancer$age) #show stats
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	47.00	50.75	56.00	55.88	60.00	67.00

```
> cat('S.D.: ', sd(cancer$age), 'S.E. mean:',
+ sd(cancer$age)/sqrt(length(cancer$age)), '\n')
S.D.: 5.659205 S.E. mean: 0.8168358
```

```
> fit9_2k<- glm(died ~ age, data=cancer, family=binomial)
> summary(fit9_2k)
```

```
Call:
```

```
glm(formula = died ~ age, family = binomial, data = cancer)
```

```
Deviance Residuals:
```

	Min	1Q	Median	3Q	Max
	-1.8243	-1.2261	0.7447	0.9504	1.2436

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.35393     3.23876  -1.344   0.179
age          0.08935     0.05859   1.525   0.127

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 62.399  on 47  degrees of freedom
Residual deviance: 59.891  on 46  degrees of freedom
AIC: 63.891

Number of Fisher Scoring iterations: 4

> fit9_2kX2<- binomialX2(cancer$died, rep(1, nrow(cancer)),
+   fitted(fit9_2k, type='response'))
> cat('Pearson X2:', fit9_2kX2, 'Dispersion:', fit9_2kX2/fit9_2k$df.residual,
+   'AIC/d.f.:', fit9_2k$aic/fit9_2k$df.residual, '\n')
Pearson X2: 48.05722 Dispersion: 1.044722 AIC/d.f.: 1.388930

> quantile(cancer$age, probs=0.75)
75%
60
> sum(cancer$age==60)
[1] 2
> which(cancer$age==60)
[1] 37 45
> cancer$age[37]<- 600 #make outlier
> fit9_2l<- glm(died ~ age, data=cancer, family=binomial)
> summary(fit9_2l)

Call:
glm(formula = died ~ age, family = binomial, data = cancer)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4872  -1.4607   0.9057   0.9196   0.9406

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.001647    0.540915   1.852   0.064 .
age          -0.006201    0.007574  -0.819   0.413
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 62.399  on 47  degrees of freedom
Residual deviance: 60.817  on 46  degrees of freedom
AIC: 64.817

Number of Fisher Scoring iterations: 5

> fit9_2lX2<- binomialX2(cancer$died, rep(1, nrow(cancer)),
+   fitted(fit9_2l, type='response'))
> cat('Pearson X2:', fit9_2lX2, 'Dispersion:', fit9_2lX2/fit9_2l$df.residual,
+   'AIC/d.f.:', fit9_2l$aic/fit9_2l$df.residual, '\n')
Pearson X2: 47.48936 Dispersion: 1.032377 AIC/d.f.: 1.409057

> exp(coef(fit9_2l)) #ORs
(Intercept)      age
  2.7227622   0.9938186

```

```

> exp(coef(fit9_2k)) #ORs
(Intercept)      age
0.01285621  1.09346718

> yd<- simul$y/simul$sd #proportions
> summary(yd) #stats
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000 0.3474  0.6211  0.5823  0.8333  1.0000
> cat('S.D.: ', sd(yd), 'S.E. mean:', sd(yd)/sqrt(length(yd)), '\n')
S.D.: 0.2824357 S.E. mean: 0.002824357

> summary(fit9_2a) #show previous fit again

Call:
glm(formula = cbind(y, noty) ~ x1 + x2 + x3, family = binomial,
    data = simul)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.81009  -0.66364   0.01385   0.70967   3.44838

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.498530    0.002955  168.7  <2e-16 ***
x1           0.999543    0.003373   296.3  <2e-16 ***
x2          -1.245868    0.003597  -346.3  <2e-16 ***
x3           0.249151    0.002952   84.4  <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: 270037  on 9999  degrees of freedom
Residual deviance: 10097  on 9996  degrees of freedom
AIC: 51172

Number of Fisher Scoring iterations: 4

> cat('Pearson X2:', fit9_2aX2, 'Dispersion:', fit9_2aX2/fit9_2a$df.residual,
+     'AIC/d.f.:', fit9_2a$aic/fit9_2a$df.residual, '\n')
Pearson X2: 9906.11 Dispersion: 0.9910073 AIC/d.f.: 5.119241

> fit9_2m<- glm(y ~ x1 + x2 + x3, data=simul, offset=d, family=poisson)
Warning messages:
1: In glm.fit(x = X, y = Y, weights = weights, start = start, etastart =
etastart, :
   fitted rates numerically 0 occurred
2: In glm.fit(x = X[, "(Intercept)", drop = FALSE], y = Y, weights = weights,
:
   fitted rates numerically 0 occurred
> summary(fit9_2m)

Call:
glm(formula = y ~ x1 + x2 + x3, family = poisson, data = simul,
    offset = d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-48.44   17.82   34.15   48.03   63.15

Coefficients:
              Estimate Std. Error  z value Pr(>|z|)
(Intercept) -89.072411   0.001702 -52331.61  <2e-16 ***

```

```

x1          0.305112    0.001595    191.29    <2e-16 ***
x2          -0.360476    0.001524   -236.53    <2e-16 ***
x3           0.075894    0.001559    48.67     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 14643129  on 9999  degrees of freedom
Residual deviance: 14559774  on 9996  degrees of freedom
AIC: 14613357

Number of Fisher Scoring iterations: 23

> fit9_2mX2<- poissonX2(simul$y, fitted(fit9_2m, type='response'))
> cat('Pearson X2:', fit9_2mX2, 'Dispersion:', fit9_2mX2/fit9_2m$df.residual,
+     'AIC/d.f.:', fit9_2m$aic/fit9_2m$df.residual, '\n')
Pearson X2: 5.372253e+21 Dispersion: 5.374403e+17 AIC/d.f.: 1461.921

> require('MASS')
> fit9_2n<- glm.nb(y ~ x1 + x2 + x3 + offset(d), data=simul)
Error: no valid set of coefficients has been found: please supply starting
values
In addition: Warning message:
In glm.fitter(x = X, y = Y, w = w, start = start, etastart = etastart, :
  fitted rates numerically 0 occurred

```

SECTION 9.3:

```

#SECTION 9.3: BINARY OVERDISPERSION
heart<- read.dta('heart01.dta') #read Stata format file
heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that complicate
tabulations
heart$center<- factor(heart$center) #convert to factor from numeric levels
heart$killip<- factor(heart$killip) #convert to factor from numeric levels

heart2<- na.omit(heart) #drop NAs

fit9_3a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
  data=heart2, family=binomial)
summary(fit9_3a)
fit9_3aX2<- binomialX2(heart2$death, rep(1, length(heart2$death)),
  fitted(fit9_3a, type='response'))
cat('Pearson X2:', fit9_3aX2, 'Dispersion:', fit9_3aX2/fit9_3a$df.residual,
  'AIC/d.f.:', fit9_3a$aic/fit9_3a$df.residual, '\n')

fit9_3b<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4 +
center,
  data=heart2, family=binomial)
summary(fit9_3b)
fit9_3bX2<- binomialX2(heart2$death, rep(1, length(heart2$death)),
  fitted(fit9_3b, type='response'))
cat('Pearson X2:', fit9_3bX2, 'Dispersion:', fit9_3bX2/fit9_3b$df.residual,
  'AIC/d.f.:', fit9_3b$aic/fit9_3b$df.residual, '\n')

```

OUTPUT: CH 9.3

```

> #SECTION 9.3: BINARY OVERDISPERSION
> heart<- read.dta('heart01.dta') #read Stata format file
> heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that
  complicate tabulations

```

```

> heart$center<- factor(heart$center) #convert to factor from numeric levels
> heart$killip<- factor(heart$killip) #convert to factor from numeric levels
> heart2<- na.omit(heart) #drop NAs
>
> fit9_3a<- glm(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
+   data=heart2, family=binomial)
> summary(fit9_3a)

Call:
glm(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
    age4, family = binomial, data = heart2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3724  -0.3257  -0.1745  -0.1270   3.1061

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.8159     0.1934  -24.896 < 2e-16 ***
anteriorAnterior  0.6387     0.1675   3.812 0.000138 ***
hcabg           0.7864     0.3527   2.229 0.025786 *
kk2             0.8249     0.1804   4.572 4.84e-06 ***
kk3            0.7967     0.2692   2.959 0.003083 **
kk4            2.6837     0.3565   7.529 5.11e-14 ***
age3           1.2668     0.2006   6.316 2.68e-10 ***
age4           1.9409     0.2080   9.329 < 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: 1486.2  on 4502  degrees of freedom
Residual deviance: 1276.3  on 4495  degrees of freedom
AIC: 1292.3

Number of Fisher Scoring iterations: 7

> fit9_3aX2<- binomialX2(heart2$death, rep(1, length(heart2$death)),
+   fitted(fit9_3a, type='response'))
> cat('Pearson X2:', fit9_3aX2, 'Dispersion:', fit9_3aX2/fit9_3a$df.residual,
+   'AIC/d.f.:', fit9_3a$aic/fit9_3a$df.residual, '\n')
Pearson X2: 4212.632 Dispersion: 0.9371817 AIC/d.f.: 0.2875015

```

Chapter 10

SECTION 10.1:

R CODE:

```

#CHAPTER 10: ORDERED LOGISTIC REGRESSION
#SECTION 10.1: PROPORTIONAL ODDS
library('foreign')
gss2002<- read.dta('gss2002_10x.dta') #read Stata format file
gss2002$sex<- factor(gss2002$sex) #convert to factor
gss2002$polviews<- factor(gss2002$polviews, ordered=TRUE) #convert to ordered
factor

```

```

summary(gss2002)
head(gss2002)

library('MASS') #for polr()
fit10_1a<- polr(polviews ~ sex, data=gss2002, method='logistic')
summary(fit10_1a)

medpar<- read.dta('medparx.dta') #read Stata format file
summary(medpar)
head(medpar)
table(medpar$type)
fit10_1b<- polr(type ~ age80, data=medpar, method='logistic')
summary(fit10_1b)

pfit10_1b<- predict(fit10_1b, type='p')
summary(pfit10_1b)

fit10_1b$zeta[1] #first intercept
1/(1+exp(-fit10_1b$zeta[1]))
fit10_1b$zeta[2] #2nd intercept
1/(1+exp(-fit10_1b$zeta[2])) - 0.7517818
1-1/(1+exp(-fit10_1b$zeta[2]))

exp(-fit10_1b$zeta[1])
0.3301732/1.3301732
1-0.2482182

exp(-fit10_1b$zeta[2])
0.07111143/1.07111143
0.2482182-0.06639032

1-(0.7517818 + 0.06639032)

exp(-fit10_1b$zeta[2])
0.07111143/1.07111143

-0.1708724 - fit10_1b$zeta[1]
1/(1+exp(-1.2790103))
-0.1708724 - fit10_1b$zeta[2]
1/(1+exp(-2.8143796)) - 1/(1+exp(-1.2790103))
1 - 1/(1+exp(-2.8143796))

-0.1708724 - fit10_1b$zeta[1]
exp(-1.2790103)
0.27831261/1.27831261
1-0.21771874

exp(-0.1708724 - fit10_1b$zeta[1])
0.27831261/1.27831261
exp(-0.1708724 - fit10_1b$zeta[2])
0.0599419/1.0599419
1-(0.78228126 + 0.05655206)

-0.1780724 - fit10_1b$zeta[2]
exp(-2.8143796)
0.0599419/1.0599419

str(medpar)
summary(medpar)
unique(medpar$provnum)
sum(is.na(medpar))

```



```

fit10_1c<- polr(type ~ died + white + hmo + los, data=medpar,
method='logistic')
summary(fit10_1c)
pfit10_1c<- predict(fit10_1c, type='p')
head(cbind(medpar$type, pfit10_1c))

#No Brant's test in R available
confint(fit10_1c)
exp(coef(fit10_1c))
exp(fit10_1c$zeta)

fit10_1d<- polr(type ~ los, data=medpar, method='logistic')
pfit10_1d<- predict(fit10_1d, type='p')
plot(medpar$los, pfit10_1d[,1], pch=1, col=1, xlab='LOS', ylab='Probability',
main='Hospital Admission Types on LOS')
points(medpar$los, pfit10_1d[,2], pch=2, col=2)
points(medpar$los, pfit10_1d[,3], pch=3, col=3)

fit10_1e<- polr(type ~ died + white, data=medpar, method='logistic')
summary(fit10_1e)

medpar2<- medpar[,c('died', 'white', 'type')]
ftable(medpar2)

```

OUTPUT:

```

> #CHAPTER 10: ORDERED LOGISTIC REGRESSION
> #SECTION 10.1: PROPORTIONAL ODDS
> library('foreign')
> gss2002<- read.dta('gss2002_10x.dta') #read Stata format file
> gss2002$sex<- factor(gss2002$sex) #convert to factor
> gss2002$polviews<- factor(gss2002$polviews, ordered=TRUE) #convert to ordered
factor
> summary(gss2002)

```

id		hrs1		wrkslf		age	
Min.	: 1	Min.	: 1.00	nap	: 0	Min.	:18.00
1st Qu.:	692	1st Qu.:	36.00	self-employed:	307	1st Qu.:	32.00
Median	:1383	Median	: 40.00	someone else	:2362	Median	:44.00
Mean	:1383	Mean	: 41.78	dk	: 0	Mean	:46.28
3rd Qu.:	2074	3rd Qu.:	50.00	na	: 0	3rd Qu.:	58.00
Max.	:2765	Max.	: 89.00	NA's	: 96	Max.	:89.00
		NA's	:1036.00			NA's	:14.00

educ		paeduc		maeduc		sex	
Min.	: 0.00	Min.	: 0.00	Min.	: 0.00	male	:1228
1st Qu.:	12.00	1st Qu.:	8.00	1st Qu.:	10.00	female:	1537
Median	:13.00	Median	: 12.00	Median	: 12.00		
Mean	:13.36	Mean	: 11.34	Mean	: 11.45		
3rd Qu.:	16.00	3rd Qu.:	14.00	3rd Qu.:	13.00		
Max.	:20.00	Max.	: 20.00	Max.	: 20.00		
NA's	:12.00	NA's	:730.00	NA's	:367.00		

polviews		sei		pasei	
moderate	: 522	Min.	: 17.10	Min.	: 17.10
conservative	: 210	1st Qu.:	33.10	1st Qu.:	32.30
slightly conservative:	209	Median	: 41.60	Median	: 38.20
slightly liberal	: 159	Mean	: 49.21	Mean	: 47.16
liberal	: 143	3rd Qu.:	63.50	3rd Qu.:	63.50
(Other)	: 88	Max.	: 97.20	Max.	: 97.20
NA's	:1434	NA's	:128.00	NA's	:653.00

masei	
Min.	: 17.10
1st Qu.:	29.00

```

Median : 38.00
Mean   : 42.97
3rd Qu.: 57.10
Max.    : 97.20
NA's    :1149.00
> head(gss2002)
  id hrs1      wrkslf age educ paeduc maeduc  sex  polviews  sei pasei
1  1   40 someone else  25  14    12    12 female conservative 57.1  36.5
2  2   72 someone else  43  16    12    12  male conservative 48.5  26.5
3  3   40 someone else  30  13    12    NA female          <NA> 68.1  33.3
4  4   60 someone else  55   2     8     8 female          <NA> 73.2  37.3
5  5   40 someone else  37   7    16    10  male          <NA> 44.7  38.2
6  6   42 someone else  47  12     8     8  male conservative 47.6  35.1
  masei
1     NA
2     NA
3     NA
4     NA
5  26.7
6  33.1

```

```

> library('MASS') #for polr()
> fit10_la<- polr(polviews ~ sex, data=gss2002, method='logistic')
> summary(fit10_la)

```

Re-fitting to get Hessian

```

Call:
polr(formula = polviews ~ sex, data = gss2002, method = "logistic")

```

```

Coefficients:
                Value Std. Error  t value
sexfemale 0.02279898  0.0986791  0.2310417

```

```

Intercepts:
                Value Std. Error t value
extremely liberal|liberal      -3.2951  0.1573  -20.9467
liberal|slightly liberal      -1.7804  0.0941  -18.9140
slightly liberal|moderate     -1.0223  0.0813  -12.5677
moderate|slghtly conservative  0.6504  0.0777   8.3741
slghtly conservative|conservative  1.4710  0.0871  16.8867
conservative|extrmly conservative  3.4610  0.1669  20.7336

```

```

Residual Deviance: 4439.939
AIC: 4453.939
(1434 observations deleted due to missingness)

```

```

> medpar<- read.dta('medparx.dta') #read Stata format file
> summary(medpar)
  provnum      died      white      hmo
Length:1495   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
Class :character 1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.0000
Mode  :character Median :0.0000   Median :1.000   Median :0.0000
                Mean  :0.3431   Mean  :0.915   Mean  :0.1599
                3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.0000
                Max.   :1.0000   Max.   :1.000   Max.   :1.0000

  los      age80      age      type
Min.   : 1.000   Min.   :0.0000   75-79 :328   Elective :1134
1st Qu.: 4.000   1st Qu.:0.0000   70-74 :317   Urgent   : 265
Median : 8.000   Median :0.0000   65-69 :291   Emergency: 96
Mean   : 9.854   Mean   :0.2207   80-84 :191
3rd Qu.:13.000   3rd Qu.:0.0000   60-64 :163

```

```

Max.      :116.000    Max.      :1.0000    85-89   : 93
                                   (Other):112
      type1          type2          type3
Min.      :0.0000    Min.      :0.0000    Min.      :0.00000
1st Qu.   :1.0000    1st Qu.   :0.0000    1st Qu.   :0.00000
Median    :1.0000    Median    :0.0000    Median    :0.00000
Mean      :0.7585    Mean      :0.1773    Mean      :0.06421
3rd Qu.   :1.0000    3rd Qu.   :0.0000    3rd Qu.   :0.00000
Max.      :1.0000    Max.      :1.0000    Max.      :1.00000

> head(medpar)
  provnum died white hmo los age80   age      type type1 type2 type3
1  030001    0     1   0   4     0 65-69 Elective     1     0     0
2  030001    0     1   1   9     0 65-69 Elective     1     0     0
3  030001    1     1   1   3     1 80-84 Elective     1     0     0
4  030001    0     1   0   9     0 75-79 Elective     1     0     0
5  030001    1     1   0   1     1 80-84 Elective     1     0     0
6  030001    1     1   0   4     0 70-74 Elective     1     0     0

> table(medpar$type)

 Elective      Urgent Emergency
    1134         265         96

> fit10_1b<- polr(type ~ age80, data=medpar, method='logistic')
> summary(fit10_1b)

Re-fitting to get Hessian

Call:
polr(formula = type ~ age80, data = medpar, method = "logistic")

Coefficients:
              Value Std. Error   t value
age80 -0.1708723   0.1485985 -1.149892

Intercepts:
              Value Std. Error t value
Elective|Urgent   1.1081   0.0677  16.3594
Urgent|Emergency  2.6435   0.1096  24.1135

Residual Deviance: 2069.593
AIC: 2075.593

> pfit10_1b<- predict(fit10_1b, type='p')
> summary(pfit10_1b)
      Elective      Urgent      Emergency
Min.      :0.7518    Min.      :0.1612    Min.      :0.05655
1st Qu.   :0.7518    1st Qu.   :0.1818    1st Qu.   :0.06639
Median    :0.7518    Median    :0.1818    Median    :0.06639
Mean      :0.7585    Mean      :0.1773    Mean      :0.06422
3rd Qu.   :0.7518    3rd Qu.   :0.1818    3rd Qu.   :0.06639
Max.      :0.7823    Max.      :0.1818    Max.      :0.06639

> fit10_1b$zeta[1] #first intercept
Elective|Urgent
 1.108138
> 1/(1+exp(-fit10_1b$zeta[1]))
Elective|Urgent
 0.7517818
> fit10_1b$zeta[2] #2nd intercept
Urgent|Emergency
 2.643507
> 1/(1+exp(-fit10_1b$zeta[2])) - 0.7517818

```

```

Urgent|Emergency
0.1818279
> 1-1/(1+exp(-fit10_lb$zeta[2]))
Urgent|Emergency
0.06639032

> exp(-fit10_lb$zeta[1])
Elective|Urgent
0.3301732
> 0.3301732/1.3301732
[1] 0.2482182
> 1-0.2482182
[1] 0.7517818

> exp(-fit10_lb$zeta[2])
Urgent|Emergency
0.07111143
> 0.07111143/1.07111143
[1] 0.06639032
> 0.2482182-0.06639032
[1] 0.1818279

> 1-(0.7517818 + 0.06639032)
[1] 0.1818279

> exp(-fit10_lb$zeta[2])
Urgent|Emergency
0.07111143
> 0.07111143/1.07111143
[1] 0.06639032

> -0.1708724 - fit10_lb$zeta[1]
Elective|Urgent
-1.279010
> 1/(1+exp(-1.2790103))
[1] 0.7822813
> -0.1708724 - fit10_lb$zeta[2]
Urgent|Emergency
-2.814380
> 1/(1+exp(-2.8143796)) - 1/(1+exp(-1.2790103))
[1] 0.1611667
> 1 - 1/(1+exp(-2.8143796))
[1] 0.05655206

> -0.1708724 - fit10_lb$zeta[1]
Elective|Urgent
-1.279010
> exp(-1.2790103)
[1] 0.2783126
> 0.27831261/1.27831261
[1] 0.2177187
> 1-0.21771874
[1] 0.7822813

> exp(-0.1708724 - fit10_lb$zeta[1])
Elective|Urgent
0.2783126
> 0.27831261/1.27831261
[1] 0.2177187
> exp(-0.1708724 - fit10_lb$zeta[2])
Urgent|Emergency
0.0599419
> 0.0599419/1.0599419

```

```

[1] 0.05655206
> 1-(0.78228126 + 0.05655206)
[1] 0.1611667

> -0.1780724 - fit10_lb$zeta[2]
Urgent|Emergency
-2.821580
> exp(-2.8143796)
[1] 0.0599419
> 0.0599419/1.0599419
[1] 0.05655206

> str(medpar)
'data.frame': 1495 obs. of 11 variables:
 $ provnum: chr "030001" "030001" "030001" "030001" ...
 $ died : int 0 0 1 0 1 1 1 1 0 0 ...
 $ white : int 1 1 1 1 1 1 1 1 1 1 ...
 $ hmo : int 0 1 1 0 0 0 0 0 0 0 ...
 $ los : int 4 9 3 9 1 4 10 3 5 6 ...
 $ age80 : int 0 0 1 0 1 0 1 1 0 0 ...
 $ age : Factor w/ 9 levels "<55","55-59",...: 4 4 7 6 7 5 8 7 4 4 ...
 $ type : Factor w/ 3 levels "Elective","Urgent",...: 1 1 1 1 1 1 2 1 1 ...
 $ type1 : int 1 1 1 1 1 1 1 0 1 1 ...
 $ type2 : int 0 0 0 0 0 0 0 1 0 0 ...
 $ type3 : int 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "datalabel")= chr ""
 - attr(*, "time.stamp")= chr "29 Jun 2008 11:22"
 - attr(*, "formats")= chr "%9s" "%9.0g" "%9.0g" "%9.0g" ...
 - attr(*, "types")= int 6 251 251 251 252 251 251 251 251 251 ...
 - attr(*, "val.labels")= chr "" "" "" "" ...
 - attr(*, "var.labels")= chr "Provider number" "1=died in hospital; 0=not
die" "1=white; 0=non-white" "HMO/readmit" ...
 - attr(*, "version")= int 8
 - attr(*, "label.table")=List of 2
 ..$ type: Named num 1 2 3
 .. ..- attr(*, "names")= chr "Elective" "Urgent" "Emergency"
 ..$ age : Named num 1 2 3 4 5 6 7 8 9
 .. ..- attr(*, "names")= chr "<55" "55-59" "60-64" "65-69" ...
> summary(medpar)
 provnum      died      white      hmo
Length:1495   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
Class :character 1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
Mode :character  Median :0.0000   Median :1.0000   Median :0.0000
                Mean   :0.3431   Mean   :0.915    Mean   :0.1599
                3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
                Max.   :1.0000   Max.   :1.0000   Max.   :1.0000

      los      age80      age      type
Min.   : 1.000   Min.   :0.0000   75-79 :328   Elective :1134
1st Qu.: 4.000   1st Qu.:0.0000   70-74 :317   Urgent  : 265
Median : 8.000   Median :0.0000   65-69 :291   Emergency: 96
Mean   : 9.854   Mean   :0.2207   80-84 :191
3rd Qu.:13.000   3rd Qu.:0.0000   60-64 :163
Max.   :116.000   Max.   :1.0000   85-89 : 93
                        (Other):112

      type1      type2      type3
Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.00000
Median :1.0000   Median :0.0000   Median :0.00000
Mean   :0.7585   Mean   :0.1773   Mean   :0.06421
3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
Max.   :1.0000   Max.   :1.0000   Max.   :1.00000

```

```
> unique(medpar$provnum)
[1] "030001" "030002" "030003" "030006" "030007" "030008" "030009" "030010"
[9] "030011" "030012" "030013" "030014" "030016" "030017" "030018" "030019"
[17] "030022" "030023" "030024" "030025" "030030" "030033" "030035" "030036"
[25] "030037" "030038" "030043" "030044" "030055" "030059" "030060" "030061"
[33] "030062" "030064" "030065" "030067" "030068" "030069" "030073" "030078"
[41] "030080" "030083" "030084" "030085" "030086" "030087" "030088" "030089"
[49] "030092" "030093" "030094" "032000" "032002" "032003"
```

```
> sum(is.na(medpar))
[1] 0
```

```
> fit10_1c<- polr(type ~ died + white + hmo + los, data=medpar,
method='logistic')
> summary(fit10_1c)
```

Re-fitting to get Hessian

Call:

```
polr(formula = type ~ died + white + hmo + los, data = medpar,
      method = "logistic")
```

Coefficients:

	Value	Std. Error	t value
died	0.54863197	0.127843541	4.291433
white	-0.50516779	0.196339867	-2.572925
hmo	-0.71078455	0.198593146	-3.579099
los	0.05779401	0.006835093	8.455483

Intercepts:

	Value	Std. Error	t value
Elective Urgent	1.3938	0.2104	6.6234
Urgent Emergency	3.0469	0.2321	13.1253

Residual Deviance: 1955.333

AIC: 1967.333

```
> pfit10_1c<- predict(fit10_1c, type='p')
> head(cbind(medpar$type, pfit10_1c))
      Elective      Urgent      Emergency
1 1 0.8412872 0.12385160 0.03486121
2 1 0.8898933 0.08696518 0.02314156
3 1 0.8685026 0.10332524 0.02817218
4 1 0.7988089 0.15518690 0.04600418
5 1 0.7845847 0.16547265 0.04994264
6 1 0.7538412 0.18731738 0.05884145
```

```
> confint(fit10_1c)
Waiting for profiling to be done...
```

Re-fitting to get Hessian

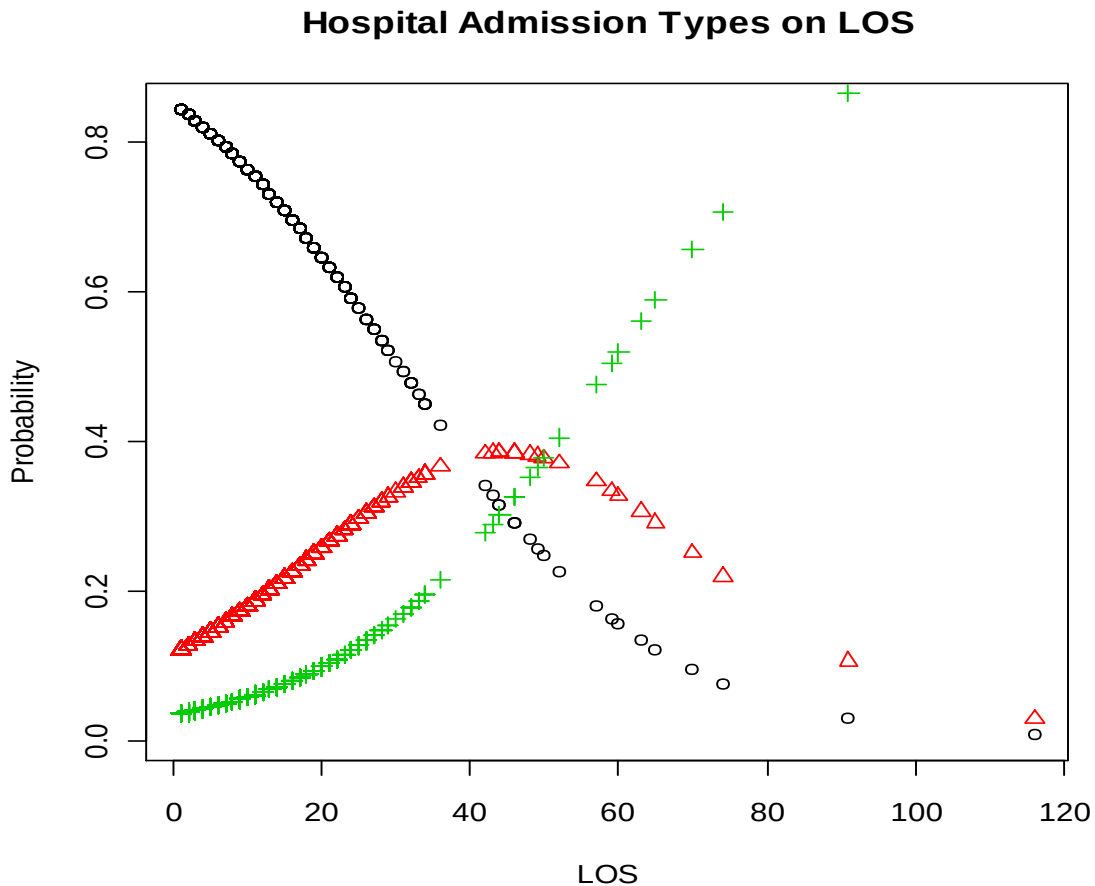
	2.5 %	97.5 %
died	0.29751888	0.7989575
white	-0.88409862	-0.1129585
hmo	-1.11488662	-0.3341545
los	0.04457789	0.0713888

```
> exp(coef(fit10_1c))
      died      white      hmo      los
1.7308835 0.6034043 0.4912586 1.0594967
> exp(fit10_1c$zeta)
      Elective|Urgent Urgent|Emergency
      4.030317      21.050101
```

```

> fit10_1d<- polr(type ~ los, data=medpar, method='logistic')
> pfit10_1d<- predict(fit10_1d, type='p')
> plot(medpar$los, pfit10_1d[,1], pch=1, col=1, xlab='LOS', ylab='Probability',
+      main='Hospital Admission Types on LOS')
> points(medpar$los, pfit10_1d[,2], pch=2, col=2)
> points(medpar$los, pfit10_1d[,3], pch=3, col=3)

```



```

> fit10_1e<- polr(type ~ died + white, data=medpar, method='logistic')
> summary(fit10_1e)

```

Re-fitting to get Hessian

Call:

```
polr(formula = type ~ died + white, data = medpar, method = "logistic")
```

Coefficients:

	Value	Std. Error	t value
died	0.4293199	0.1238931	3.465245
white	-0.6564387	0.1904411	-3.446938

Intercepts:

	Value	Std. Error	t value
Elective Urgent	0.7103	0.1837	3.8672
Urgent Emergency	2.2609	0.2008	11.2623

Residual Deviance: 2048.87

AIC: 2056.87

```
> medpar2<- medpar[,c('died', 'white', 'type')]
> ftable(medpar2)
      type Elective Urgent Emergency
died white
0      0           62      23        6
      1          708     138       45
1      0           18      14        4
      1          346      90       41
```

CHAPTER 11 and 12

No code for Chapters 11 and 12 (and 14/15)

CHAPTER 13

#SECTION 13.1: GENERALIZED ESTIMATING EQUATIONS

```
library('foreign')
german<- read.dta('rwm_1980x.dta') #read Stata format file
summary(german)
head(german)
str(german)

fit13_1a<- glm(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, family=binomial(link='logit'))
summary(fit13_1a)
exp(coef(fit13_1a)) #ORs

library('gee') #one package for gee()
fit13_1b<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, id=id, corstr='independence', family=binomial)
summary(fit13_1b)
exp(coef(fit13_1b)) #ORs

fit13_1c<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, id=id, corstr='exchangeable', family=binomial)
summary(fit13_1c)
exp(coef(fit13_1c)) #ORs

library('geepack') #another package for gee() and geese()
fit13_1d<- geeglm(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, id=id, corstr='ar1', family=binomial(link='logit'))
summary(fit13_1d)
exp(coef(fit13_1d)) #ORs

fit13_1e<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, id=id, corstr='unstructured', family=binomial)
summary(fit13_1e)
exp(coef(fit13_1e)) #ORs

fit13_1f<- geeglm(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
```



```

data=german, id=id, corstr='unstructured', family=binomial(link='logit'))
summary(fit13_1f)
exp(coef(fit13_1f)) #ORs

fit13_1g<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, id=id, corstr='stat_M_dep', Mv=1, family=binomial)
summary(fit13_1g)
exp(coef(fit13_1g)) #ORs

fit13_1h<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
  data=german, id=id, corstr='non_stat_M_dep', Mv=1, family=binomial)
summary(fit13_1h)
exp(coef(fit13_1h)) #ORs

heart<- read.dta('heart01.dta') #read Stata format file
heart$anterior <- heart$anterior[,drop=TRUE] #drop empty levels that complicate
tabulations
heart$center<- factor(heart$center) #convert to factor from numeric levels
heart$killip<- factor(heart$killip) #convert to factor from numeric levels

fit13_1i<- gee(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
  data=heart, id=center, corstr='exchangeable', family=binomial)
fit13_1i

heartgrp<- read.dta('heart02grp.dta')
head(heartgrp)
ix<- order(heartgrp$center) #get indices for center order
heartgrp2<- heartgrp[ix,] #sort by center

fit13_1j<- gee(cbind(dead, cases-dead)~ anterior + hcabg + kk2 + kk3 + kk4 +
  age3 + age4,
  data=heartgrp2, id=center, corstr='exchangeable', family=binomial)
fit13_1j

```

OUTPUT: 13.1

```

> #CHAPTER 13: PANEL MODELS
> #SECTION 13.1: GEE
> library('foreign')
> german<- read.dta('rwm_1980x.dta') #read Stata format file
> summary(german)

```

id		female	year	age
Min. :	1	Min. :0.0000	Min. :1984	Min. :25.00
1st Qu.:	1746	1st Qu.:0.0000	1st Qu.:1985	1st Qu.:34.00
Median :	3471	Median :0.0000	Median :1986	Median :44.00
Mean :	3448	Mean :0.4805	Mean :1986	Mean :43.79
3rd Qu.:	5134	3rd Qu.:1.0000	3rd Qu.:1987	3rd Qu.:53.00
Max. :	7028	Max. :1.0000	Max. :1988	Max. :64.00

hsat		handdum	handper	hhninc
Min. :	0.000	Min. :0.0000	Min. : 0.000	Min. : 0.000
1st Qu.:	5.000	1st Qu.:0.0000	1st Qu.: 0.000	1st Qu.: 2.200
Median :	7.000	Median :0.0000	Median : 0.000	Median : 3.000
Mean :	6.819	Mean :0.2555	Mean : 7.451	Mean : 3.239
3rd Qu.:	8.000	3rd Qu.:1.0000	3rd Qu.: 3.397	3rd Qu.: 4.000
Max. :	10.000	Max. :1.0000	Max. :100.000	Max. :30.671

hhkids		educ	married	haupts
Min. :	0.0000	Min. : 7.00	Min. :0.0000	Min. :0.0000
1st Qu.:	0.0000	1st Qu.:10.50	1st Qu.:1.0000	1st Qu.:0.0000
Median :	0.0000	Median :10.50	Median :1.0000	Median :1.0000
Mean :	0.4132	Mean :11.25	Mean :0.7736	Mean :0.6478
3rd Qu.:	1.0000	3rd Qu.:11.50	3rd Qu.:1.0000	3rd Qu.:1.0000

```

Max.      :1.0000   Max.      :18.00   Max.      :1.0000   Max.      :1.0000
  reals          fachhs          abitur          univ
Min.      :0.0000   Min.      :0.0000   Min.      :0.0000   Min.      :0.00000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
Median :0.0000   Median :0.0000   Median :0.0000   Median :0.00000
Mean     :0.1879   Mean     :0.0385   Mean     :0.1099   Mean     :0.06874
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
Max.      :1.0000   Max.      :1.0000   Max.      :1.0000   Max.      :1.00000
  working        bluec          whitec          self
Min.      :0.0000   Min.      :0.0000   Min.      :0.0000   Min.      :0.00000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
Median :1.0000   Median :0.0000   Median :0.0000   Median :0.00000
Mean     :0.6561   Mean     :0.2462   Mean     :0.2860   Mean     :0.06313
3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
Max.      :1.0000   Max.      :1.0000   Max.      :1.0000   Max.      :1.00000
  beamt          docvis          hospvis          public
Min.      :0.00000   Min.      : 0.000   Min.      : 0.0000   Min.      :0.0000
1st Qu.:0.00000   1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.:0.0000
Median :0.00000   Median : 1.000   Median : 0.0000   Median :1.0000
Mean     :0.07527   Mean     : 3.176   Mean     : 0.1382   Mean     :0.8902
3rd Qu.:0.00000   3rd Qu.: 4.000   3rd Qu.: 0.0000   3rd Qu.:1.0000
Max.      :1.00000   Max.      :121.000   Max.      :51.0000   Max.      :1.0000
  addon          logl_obs          sex          y84
Min.      :0.00000   Min.      :-16.29824   Min.      :1.000   Min.      :0.0000
1st Qu.:0.00000   1st Qu.: -1.97949   1st Qu.:1.000   1st Qu.:0.0000
Median :0.00000   Median : -0.61100   Median :1.000   Median :0.0000
Mean     :0.01632   Mean     : -1.17735   Mean     :1.480   Mean     :0.1976
3rd Qu.:0.00000   3rd Qu.: -0.08970   3rd Qu.:2.000   3rd Qu.:0.0000
Max.      :1.00000   Max.      : -0.03066   Max.      :2.000   Max.      :1.0000
  y85          y86          y87          y88
Min.      :0.0000   Min.      :0.0000   Min.      :0.0000   Min.      :0.0000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
Mean     :0.1935   Mean     :0.1934   Mean     :0.1870   Mean     :0.2286
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.      :1.0000   Max.      :1.0000   Max.      :1.0000   Max.      :1.0000
  y91          y94          ti          agesq          sample
Min.      :0   Min.      :0   Min.      :1.000   Min.      :0.625   Min.      :1
1st Qu.:0   1st Qu.:0   1st Qu.:4.000   1st Qu.:1.156   1st Qu.:1
Median :0   Median :0   Median :5.000   Median :1.936   Median :1
Mean     :0   Mean     :0   Mean     :5.011   Mean     :2.043   Mean     :1
3rd Qu.:0   3rd Qu.:0   3rd Qu.:6.000   3rd Qu.:2.809   3rd Qu.:1
Max.      :0   Max.      :0   Max.      :7.000   Max.      :4.096   Max.      :1
  doctor          hospital          edlevel          edlevel1
Min.      :0.0000   Min.      :0.00000   Not HS grad:15433   Min.      :0.000
1st Qu.:0.0000   1st Qu.:0.00000   HS grad      : 1153   1st Qu.:1.000
Median :1.0000   Median :0.00000   Coll/Univ    : 1733   Median :1.000
Mean     :0.6139   Mean     :0.08598   Grad School: 1290   Mean     :0.787
3rd Qu.:1.0000   3rd Qu.:0.00000   Max.      :1.00000   3rd Qu.:1.000
Max.      :1.0000   Max.      :1.00000   Max.      :1.00000   Max.      :1.000
  edlevel2          edlevel3          edlevel4          outwork
Min.      :0.0000   Min.      :0.00000   Min.      :0.00000   Min.      :0.0000
1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
Median :0.0000   Median :0.00000   Median :0.00000   Median :0.0000
Mean     :0.0588   Mean     :0.08838   Mean     :0.06579   Mean     :0.3439
3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1.0000
Max.      :1.0000   Max.      :1.00000   Max.      :1.00000   Max.      :1.0000
> head(german)
  id female year age hsat handdum handper  hhninc hhkids educ married haupts
1  1      0 1984  54   8      0      0 3.050000      0   15      1      0
2  1      0 1985  55   8      0      0 4.510050      0   15      1      0
3  1      0 1986  56   7      0      0 3.500000      0   15      1      0
4  2      1 1984  44   7      0      0 3.050000      0    9      1      1

```

```

5 2      1 1985 45      8      0      0 3.182779      0      9      1      1
6 2      1 1986 46      7      0      0 3.500000      0      9      1      1
  reals fachhs abitur univ working bluec whitec self beamt docvis hospvis
1 0      1      0      0      1      0      1      0      0      1      0
2 0      1      0      0      1      0      1      0      0      0      0
3 0      1      0      0      1      0      1      0      0      0      0
4 0      0      0      0      0      0      0      0      0      0      0
5 0      0      0      0      0      0      0      0      0      1      0
6 0      0      0      0      0      0      0      0      0      2      0
  public addon      logl_obs sex y84 y85 y86 y87 y88 y91 y94 ti agesq sample
1 1      0 -1.42361009 1 1 0 0 0 0 0 0 3 2.916 1
2 1      0 -0.80305290 1 0 1 0 0 0 0 0 3 3.025 1
3 1      0 -0.97513211 1 0 0 1 0 0 0 0 3 3.136 1
4 1      0 -0.08281561 2 1 0 0 0 0 0 0 4 1.936 1
5 1      0 -0.06861576 2 0 1 0 0 0 0 0 4 2.025 1
6 1      0 -0.08245477 2 0 0 1 0 0 0 0 4 2.116 1
  doctor hospital      edlevel edlevel1 edlevel2 edlevel3 edlevel4 outwork
1 1      0      0 Coll/Univ      0      0      1      0      0
2 0      0      0 Coll/Univ      0      0      1      0      0
3 0      0      0 Coll/Univ      0      0      1      0      0
4 0      0      0 Not HS grad      1      0      0      0      1
5 1      0      0 Not HS grad      1      0      0      0      1
6 1      0      0 Not HS grad      1      0      0      0      1
> str(german)
'data.frame': 19609 obs. of 45 variables:
 $ id      : int  1 1 1 2 2 2 2 3 3 3 ...
 $ female  : int  0 0 0 1 1 1 1 1 1 1 ...
 $ year    : int  1984 1985 1986 1984 1985 1986 1988 1984 1986 1987 ...
 $ age     : int  54 55 56 44 45 46 48 58 60 61 ...
 $ hsat    : num  8 8 7 7 8 7 8 10 9 10 ...
 $ handdum : num  0 0 0 0 0 ...
 $ handper : num  0 0 0 0 0 ...
 $ hhninc  : num  3.05 4.51 3.50 3.05 3.18 ...
 $ hhkids  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ educ    : num  15 15 15 9 9 9 9 11 11 11 ...
 $ married : int  1 1 1 1 1 1 1 0 0 0 ...
 $ haupts  : int  0 0 0 1 1 1 1 1 1 1 ...
 $ reals   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ fachhs  : int  1 1 1 0 0 0 0 0 0 0 ...
 $ abitur  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ univ    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ working : int  1 1 1 0 0 0 0 0 0 0 ...
 $ bluec   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ whitec  : int  1 1 1 0 0 0 0 0 0 0 ...
 $ self    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ beamt   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ docvis  : int  1 0 0 0 1 2 1 0 0 10 ...
 $ hospvis : int  0 0 0 0 0 0 0 0 0 0 ...
 $ public  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ addon   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ logl_obs : num  -1.4236 -0.8031 -0.9751 -0.0828 -0.0686 ...
 $ sex     : int  1 1 1 2 2 2 2 2 2 2 ...
 $ y84     : int  1 0 0 1 0 0 0 1 0 0 ...
 $ y85     : int  0 1 0 0 1 0 0 0 0 0 ...
 $ y86     : int  0 0 1 0 0 1 0 0 1 0 ...
 $ y87     : int  0 0 0 0 0 0 0 0 0 1 ...
 $ y88     : int  0 0 0 0 0 0 1 0 0 0 ...
 $ y91     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ y94     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ ti      : int  3 3 3 4 4 4 4 4 4 4 ...
 $ agesq   : num  2.92 3.03 3.14 1.94 2.03 ...
 $ sample  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ doctor  : int  1 0 0 0 1 1 1 0 0 1 ...

```

```

$ hospital: int  0 0 0 0 0 0 0 0 0 0 ...
$ edlevel  : Factor w/ 4 levels "Not HS grad",...: 3 3 3 1 1 1 1 1 1 1 ...
$ edlevel1: int  0 0 0 1 1 1 1 1 1 1 ...
$ edlevel2: int  0 0 0 0 0 0 0 0 0 0 ...
$ edlevel3: int  1 1 1 0 0 0 0 0 0 0 ...
$ edlevel4: int  0 0 0 0 0 0 0 0 0 0 ...
$ outwork  : num  0 0 0 1 1 1 1 1 1 1 ...
- attr(*, "datalabel")= chr ""
- attr(*, "time.stamp")= chr "29 Jun 2008 11:54"
- attr(*, "formats")= chr  "%8.0g" "%8.0g" "%8.0g" "%8.0g" ...
- attr(*, "types")= int   252 251 252 251 254 254 254 251 254 ...
- attr(*, "val.labels")= chr  "" "" "" "" "" ...
- attr(*, "var.labels")= chr  "" "" "" "" "" ...
- attr(*, "version")= int  8
- attr(*, "label.table")=List of 1
..$ edlevel: Named num  1 2 3 4
.. ..- attr(*, "names")= chr  "Not HS grad" "HS grad" "Coll/Univ" "Grad
School"

> fitl3_1a<- glm(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
+ data=german, family=binomial(link='logit'))
> summary(fitl3_1a)

Call:
glm(formula = outwork ~ female + married + edlevel2 + edlevel3 +
    edlevel4, family = binomial(link = "logit"), data = german)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.420   -0.600   -0.593    1.046    2.416

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.86068     0.04503 -41.323  < 2e-16 ***
female       1.96683     0.03525  55.790  < 2e-16 ***
married      0.21174     0.04056   5.220 1.79e-07 ***
edlevel2    -0.32737     0.07031  -4.656 3.23e-06 ***
edlevel3     0.23704     0.06015   3.941 8.13e-05 ***
edlevel4    -1.00256     0.08811 -11.378 < 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: 25241  on 19608  degrees of freedom
Residual deviance: 21306  on 19603  degrees of freedom
AIC: 21318

Number of Fisher Scoring iterations: 4

> exp(coef(fitl3_1a)) #ORs
(Intercept)      female    married    edlevel2    edlevel3    edlevel4
  0.1555669    7.1479921    1.2358241    0.7208191    1.2674923    0.3669405

> library('gee')
> fitl3_1b<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
+ data=german, id=id, corstr='independence', family=binomial)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)      female    married    edlevel2    edlevel3    edlevel4
 -1.8606796    1.9668315    0.2117381   -0.3273671    0.2370404   -1.0025556
> summary(fitl3_1b)

```

```

GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link:                               Logit
Variance to Mean Relation: Binomial
Correlation Structure:              Independent

Call:
gee(formula = outwork ~ female + married + edlevel2 + edlevel3 +
    edlevel4, id = id, data = german, family = binomial, corstr =
    "independence")

Summary of Residuals:
      Min       1Q   Median       3Q      Max
-0.6352787 -0.1647036 -0.1612520  0.4211901  0.9459988

Coefficients:
              Estimate Naive S.E.    Naive z Robust S.E.    Robust z
(Intercept) -1.8606796  0.04528027 -41.092504  0.07861377 -23.668624
female       1.9668315  0.03545194  55.478812  0.05916917  33.240814
married      0.2117381  0.04078799   5.191187  0.06950145   3.046527
edlevel2     -0.3273671  0.07070579  -4.629991  0.11499083  -2.846898
edlevel3     0.2370404  0.06049057   3.918634  0.10222682   2.318769
edlevel4     -1.0025557  0.08861313 -11.313850  0.13549269  -7.399334

Estimated Scale Parameter:  1.011225
Number of Iterations:  1

Working Correlation
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0    0
[2,]    0    1    0    0    0
[3,]    0    0    1    0    0
[4,]    0    0    0    1    0
[5,]    0    0    0    0    1
> exp(coef(fit13_1b))
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
  0.1555669   7.1479921   1.2358241   0.7208191   1.2674923   0.3669405

> fit13_1c<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
+   data=german, id=id, corstr='exchangeable', family=binomial)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
 -1.8606796   1.9668315   0.2117381  -0.3273671   0.2370404  -1.0025556
> summary(fit13_1c)

GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link:                               Logit
Variance to Mean Relation: Binomial
Correlation Structure:              Exchangeable

Call:
gee(formula = outwork ~ female + married + edlevel2 + edlevel3 +
    edlevel4, id = id, data = german, family = binomial, corstr =
    "exchangeable")

Summary of Residuals:

```

	Min	1Q	Median	3Q	Max
	-0.6298875	-0.2034342	-0.1810573	0.4289870	0.9345034

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	-1.6107157	0.05887832	-27.356687	0.06633523	-24.281452
female	1.7951863	0.05272514	34.048014	0.05473564	32.797391
married	0.1015149	0.04848924	2.093556	0.05674723	1.788897
edlevel2	-0.2192355	0.09900304	-2.214432	0.11182973	-1.960441
edlevel3	0.2457486	0.08308821	2.957683	0.10259829	2.395250
edlevel4	-1.0473016	0.12570230	-8.331603	0.14270261	-7.339050

Estimated Scale Parameter: 0.9515808

Number of Iterations: 3

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.0000000	0.6423391	0.6423391	0.6423391	0.6423391
[2,]	0.6423391	1.0000000	0.6423391	0.6423391	0.6423391
[3,]	0.6423391	0.6423391	1.0000000	0.6423391	0.6423391
[4,]	0.6423391	0.6423391	0.6423391	1.0000000	0.6423391
[5,]	0.6423391	0.6423391	0.6423391	0.6423391	1.0000000

> exp(coef(fit13_1c)) #ORs

	female	married	edlevel2	edlevel3	edlevel4
(Intercept)	0.1997446	6.0205963	1.1068464	0.8031325	1.2785781
					0.3508833

> library('geepack') #another package for geeglm() and geese()

> fit13_1d<- geeglm(outwork ~ female + married + edlevel2 + edlevel3 +
edlevel4,

+ data=german, id=id, corstr='ar1', family=binomial(link='logit'))

> summary(fit13_1d)

Call:

```
geeglm(formula = outwork ~ female + married + edlevel2 + edlevel3 +  
edlevel4, family = binomial(link = "logit"), data = german,  
id = id, corstr = "ar1")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)
(Intercept)	-1.6056	0.0664	584.14	< 2e-16 ***
female	1.7985	0.0546	1086.27	< 2e-16 ***
married	0.0937	0.0574	2.67	0.103
edlevel2	-0.2095	0.1165	3.23	0.072 .
edlevel3	0.2280	0.1080	4.46	0.035 *
edlevel4	-1.0212	0.1477	47.79	4.8e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	0.95	0.0293

Correlation: Structure = ar1 Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.781	0.0159

Number of clusters: 6127 Maximum cluster size: 5

> exp(coef(fit13_1d)) #ORs

	female	married	edlevel2	edlevel3	edlevel4
(Intercept)	0.201	6.041	1.098	0.811	1.256
					0.360

```
> fit13_1e<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
+   data=german, id=id, corstr='unstructured', family=binomial)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
      -1.861      1.967      0.212      -0.327      0.237      -1.003
> summary(fit13_1e)
```

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

```
Model:
Link:                               Logit
Variance to Mean Relation: Binomial
Correlation Structure:              Unstructured
```

```
Call:
gee(formula = outwork ~ female + married + edlevel2 + edlevel3 +
      edlevel4, id = id, data = german, family = binomial, corstr =
      "unstructured")
```

```
Summary of Residuals:
      Min      1Q  Median      3Q      Max
-0.626 -0.189 -0.174  0.428  0.937
```

```
Coefficients:
              Estimate Naive S.E. Naive z Robust S.E. Robust z
(Intercept)   -1.683      0.0567  -29.66      0.0683   -24.65
female         1.843      0.0478   38.53      0.0557   33.08
married        0.128      0.0492    2.61      0.0582    2.20
edlevel2       -0.281      0.0930   -3.03      0.1097   -2.56
edlevel3        0.228      0.0786    2.90      0.0993    2.29
edlevel4       -1.025      0.1179   -8.70      0.1347   -7.61
```

```
Estimated Scale Parameter: 0.968
Number of Iterations: 3
```

```
Working Correlation
      [,1] [,2] [,3] [,4] [,5]
[1,] 1.000 0.653 0.450 0.282 0.135
[2,] 0.653 1.000 0.525 0.339 0.170
[3,] 0.450 0.525 1.000 0.390 0.187
[4,] 0.282 0.339 0.390 1.000 0.223
[5,] 0.135 0.170 0.187 0.223 1.000
```

```
> exp(coef(fit13_1e)) #ORs
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
      0.186      6.317      1.137      0.755      1.256      0.359
```

```
> fit13_1f<- geeglm(outwork ~ female + married + edlevel2 + edlevel3 +
+   edlevel4,
+   data=german, id=id, corstr='unstructured', family=binomial(link='logit'))
> summary(fit13_1f)
```

```
Call:
geeglm(formula = outwork ~ female + married + edlevel2 + edlevel3 +
      edlevel4, family = binomial(link = "logit"), data = german,
      id = id, corstr = "unstructured")
```

```
Coefficients:
              Estimate Std.err      Wald Pr(>|W|)
(Intercept)  -1.6197   0.0653   615.14 < 2e-16 ***
female        1.8017   0.0545  1094.00 < 2e-16 ***
```

```

married      0.1082  0.0552   3.84   0.050 .
edlevel2     -0.2146  0.1114   3.71   0.054 .
edlevel3      0.2468  0.1027   5.78   0.016 *
edlevel4     -1.0326  0.1419  52.98  3.4e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Estimated Scale Parameters:
              Estimate Std.err
(Intercept)   0.952   0.0295

Correlation: Structure = unstructured Link = identity

Estimated Correlation Parameters:
              Estimate Std.err
alpha.1:2     0.704   0.0256
alpha.1:3     0.607   0.0249
alpha.1:4     0.523   0.0254
alpha.1:5     0.455   0.0289
alpha.2:3     0.707   0.0266
alpha.2:4     0.627   0.0270
alpha.2:5     0.574   0.0307
alpha.3:4     0.720   0.0286
alpha.3:5     0.630   0.0321
alpha.4:5     0.750   0.0344
Number of clusters: 6127 Maximum cluster size: 5
> exp(coef(fit13_1f)) #ORs
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
      0.198         6.060         1.114         0.807         1.280         0.356

> fit13_1g<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
+ data=german, id=id, corstr='stat_M_dep', Mv=1, family=binomial)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
      -1.861         1.967         0.212        -0.327         0.237        -1.003
Error in gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4, :
  cgee: M-dependence, M=1, but clustsize=1
fatal error for this model
> summary(fit13_1g)
Error in summary(fit13_1g) : object "fit13_1g" not found
> exp(coef(fit13_1g)) #ORs
Error in coef(fit13_1g) : object "fit13_1g" not found

> fit13_1h<- gee(outwork ~ female + married + edlevel2 + edlevel3 + edlevel4,
+ data=german, id=id, corstr='non_stat_M_dep', Mv=1, family=binomial)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)      female      married      edlevel2      edlevel3      edlevel4
      -1.861         1.967         0.212        -0.327         0.237        -1.003
> summary(fit13_1h)

GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
Link:                               Logit
Variance to Mean Relation: Binomial
Correlation Structure:               Non-Stationary M-dependent , M = 1

Call:
gee(formula = outwork ~ female + married + edlevel2 + edlevel3 +

```



```

    edlevel4, id = id, data = german, family = binomial, corstr =
"non_stat_M_dep",
    Mv = 1)

```

Summary of Residuals:

Min	1Q	Median	3Q	Max
-0.635	-0.178	-0.165	0.421	0.942

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	-1.768	0.0518	-34.13	0.0740	-23.90
female	1.935	0.0428	45.20	0.0584	33.11
married	0.150	0.0452	3.33	0.0641	2.35
edlevel2	-0.298	0.0837	-3.56	0.1177	-2.53
edlevel3	0.236	0.0716	3.30	0.1049	2.25
edlevel4	-1.023	0.1069	-9.57	0.1393	-7.35

Estimated Scale Parameter: 0.993

Number of Iterations: 3

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1.000	0.649	0.000	0.000	0.000
[2,]	0.649	1.000	0.521	0.000	0.000
[3,]	0.000	0.521	1.000	0.387	0.000
[4,]	0.000	0.000	0.387	1.000	0.222
[5,]	0.000	0.000	0.000	0.222	1.000

> exp(coef(fit13_1h)) #ORs

(Intercept)	female	married	edlevel2	edlevel3	edlevel4
0.171	6.922	1.162	0.742	1.266	0.359

> heart<- read.dta('heart01.dta') #read Stata format file

> heart\$anterior <- heart\$anterior[,drop=TRUE] #drop empty levels that
complicate tabulations

> heart\$center<- factor(heart\$center) #convert to factor from numeric levels

> heart\$killip<- factor(heart\$killip) #convert to factor from numeric levels

> fit13_1i<- gee(death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 + age4,
+ data=heart, id=center, corstr='exchangeable', family=binomial)

Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27

running glm to get initial regression estimate

(Intercept)	anterior	Anterior	hcabg	kk2
-4.816		0.639	0.786	0.825
	kk3	kk4	age3	age4
0.797		2.684	1.267	1.941

> fit13_1i

GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:

Link: Logit
Variance to Mean Relation: Binomial
Correlation Structure: Exchangeable

Call:

gee(formula = death ~ anterior + hcabg + kk2 + kk3 + kk4 + age3 +
age4, id = center, data = heart, family = binomial, corstr =
"exchangeable")

Number of observations : 4503

Maximum cluster size : 215

Coefficients:

(Intercept)	anteriorAnterior	hcabg	kk2
-4.816	0.639	0.785	0.825
kk3	kk4	age3	age4
0.796	2.682	1.267	1.941

Estimated Scale Parameter: 0.937

Number of Iterations: 2

Working Correlation[1:4,1:4]

	[,1]	[,2]	[,3]	[,4]
[1,]	1.000000	-0.000194	-0.000194	-0.000194
[2,]	-0.000194	1.000000	-0.000194	-0.000194
[3,]	-0.000194	-0.000194	1.000000	-0.000194
[4,]	-0.000194	-0.000194	-0.000194	1.000000

Returned Error Value:

[1] 0

```
> heartgrp<- read.dta('heart02grp.dta')
```

```
> head(heartgrp)
```

	death	anterior	hcabg	kk1	kk2	kk3	kk4	center	age1	age2	age3	age4	grp	cases
1	0	Inferior	0	0	0	0	1	1489	0	0	0	1	1	1
2	0	Inferior	0	0	0	0	1	7048	0	0	0	1	2	1
3	0	Inferior	0	0	0	0	1	8332	0	0	0	1	3	1
4	0	Inferior	0	0	0	0	1	4012	0	0	1	0	4	1
5	1	Inferior	0	0	0	0	1	7048	0	0	1	0	5	1
6	0	Inferior	0	0	0	0	1	8542	0	0	1	0	6	1

	dead	kk23	killip
1	0	0	4
2	0	0	4
3	0	0	4
4	0	0	4
5	1	0	4
6	0	0	4

```
> ix<- order(heartgrp$center) #get indices for center order
```

```
> heartgrp2<- heartgrp[ix,] #sort by center
```

```
> head(heartgrp2)
```

	death	anterior	hcabg	kk1	kk2	kk3	kk4	center	age1	age2	age3	age4	grp	cases
13	0	Inferior	0	0	0	0	1	1255	1	0	0	0	13	1
18	0	Inferior	0	0	0	1	0	1255	0	0	0	1	18	2
30	0	Inferior	0	0	0	1	0	1255	0	0	1	0	30	1
76	1	Inferior	0	0	1	0	0	1255	0	0	0	1	76	1
104	0	Inferior	0	0	1	0	0	1255	0	0	1	0	104	1
173	0	Inferior	0	0	1	0	0	1255	1	0	0	0	173	5

	dead	kk23	killip
13	0	0	4
18	0	1	3
30	0	1	3
76	1	1	2
104	0	1	2
173	0	1	2

```
> fit13_1j<- gee(cbind(death, cases-death)~ anterior + hcabg + kk2 + kk3 + kk4 +  
age3 + age4,
```

```
+ data=heartgrp2, id=center, corstr='exchangeable', family=binomial)
```

```
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
Error in gee(cbind(death, cases - death) ~ anterior + hcabg + kk2 + kk3 + :
```

```
rank-deficient model matrix
```

SECTION 13.1.6 - no output

* My thanks to Jichun Xie, Univ. of Pennsylvania, for her contribution to this section of R code.

```
setwd("C:/My Documents/Joe")

library('foreign') #foreign import package library
library('qlspack')
rwm<- read.dta('rwm_1980.dta') #works for STATA v.5-9

attach(rwm)

id.list<- cluster.size(id)
n<- id.list$n
cnt<- rep(n,n)
length(cnt) #number of observations
summary(cnt) #range, mean and quantiles
sd(cnt) #standard error

qls.ar1<- qls(formula = outwork ~ female + married + edlevel2+ edlevel3+edlevel4, data = rwm,
              id = id, family = binomial, time = year, correlation = "ar1")
              #fit the model with the AR1 structure
summary(qls.ar1) #show more details

alpha<- qls.ar1$geese$alpha
corr.mat<- cormax.ar1(alpha,id,time="NA")
corr.mat #get the correlation matrix
eigen(corr.mat) #get the eigenvalues for the correlation matrix

n.max<- max(n)
a = -1/(2*sin((n.max-1)/(n.max + 1)*pi/2))
b = 1/(2*sin((n.max-1)/(n.max+ 1)*pi/2))
scope<- c(a,b)
names(scope)<- c("a","b")
print(scope) #get the boundary

qls.markov<- qls(formula = outwork ~ female + married + edlevel2+ edlevel3+edlevel4,
                 data = rwm, id = id, family = binomial, time = year, correlation = "markov")
                 #fit the model with the Markov structure
summary(qls.markov) #show more details

alpha<- qls.markov$geese$alpha
corr.mat<- cormax.markov(alpha,id,time="NA")
eigen(corr.mat)

qls.tri<- qls(formula = outwork ~ female + married +edlevel2+ edlevel3+edlevel4 data = rwm,
              id = id, family = binomial, time = year, correlation = "tridiagonal") #fit the model with the
              Markov structure
summary(qls.tri) #show more details

alpha<- qls.tri$geese$alpha
corr.mat<- cormax.tri(alpha,id,time="NA")
eigen(corr.mat)
```

#SECTION 2: UNCONDITIONAL FIXED EFFECTS LOGISTIC REGRESSION

#SECTION 3: CONDITIONAL FIXED EFFECTS LOGSITIC REGRESSION

**#SECTION 4: RANDOM EFFECTS AND MIXED EFFECTS LOGISTIC
REGRESSION**

CHAPTERS 14 and 15

No code is available for these chapter. Very few examples have R support.