Logistic Regression

Contents

1 Getting Set Up

1.1 Setting chunk options

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE)
knitr::purl("logistic-regression.Rmd")
```

```
##
##
## processing file: logistic-regression.Rmd
```

```
## output file: logistic-regression.R
```

1.2 Installing Packages

```
install.packages("AICcmodavg")
```

1.3 Reading the data

```
donner = read.table("donner-class.txt", row.names = 1, header=TRUE)
attach(donner)
head(donner,10)
```

```
##                          Age Outcome    Sex Family.name Status
## Breen_Edward_             13       1   Male       Breen Family
## Breen_Margaret_Isabella    1       1 Female       Breen Family
## Breen_James_Frederick      5       1   Male       Breen Family
## Breen_John                14       1   Male       Breen Family
## Breen_Margaret_Bulger     40       1 Female       Breen Family
## Breen_Patrick             51       1   Male       Breen Family
## Breen_Patrick_Jr.          9       1   Male       Breen Family
## Breen_Peter                3       1   Male       Breen Family
## Breen_Simon_Preston        8       1   Male       Breen Family
## Donner_Elitha_Cumi        13       1 Female    G_Donner Family
```

```
# Keeping only the variables of interest
donner.na = na.omit(subset(donner,select=c('Age','Outcome','Sex')))
donner.na$fem = as.numeric(donner.na$Sex=="Female")
head(donner.na,10)
```

```
##                          Age Outcome    Sex fem
## Breen_Edward_             13       1   Male   0
## Breen_Margaret_Isabella    1       1 Female   1
## Breen_James_Frederick      5       1   Male   0
## Breen_John                14       1   Male   0
## Breen_Margaret_Bulger     40       1 Female   1
## Breen_Patrick             51       1   Male   0
## Breen_Patrick_Jr.          9       1   Male   0
## Breen_Peter                3       1   Male   0
## Breen_Simon_Preston        8       1   Male   0
## Donner_Elitha_Cumi        13       1 Female   1
```

2 Fitting a logistic regression

```
donner.log = glm(Outcome ~ Age+fem, data=donner.na, family=binomial(link="logit"))
summary(donner.log)
```

```
##
## Call:
```

```
## glm(formula = Outcome ~ Age + fem, family = binomial(link = "logit"),
##     data = donner.na)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8828  -1.0383   0.6511   1.0261   1.7386
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.55382    0.41788   1.325   0.1851
## Age         -0.03561    0.01525  -2.336   0.0195 *
## fem          1.06798    0.48229   2.214   0.0268 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 120.86  on 87  degrees of freedom
## Residual deviance: 108.87  on 85  degrees of freedom
## AIC: 114.87
##
## Number of Fisher Scoring iterations: 4
```

2.1 Odds ratios

```
exp(donner.log$coefficients)
```

```
## (Intercept)         Age         fem
##   1.7398953   0.9650211   2.9094868
```

```
exp(confint(donner.log))
```

```
##                 2.5 %    97.5 %
## (Intercept) 0.7748972 4.0431170
## Age         0.9348223 0.9930661
## fem         1.1543365 7.7529827
```

```
exp(cbind(OR=donner.log$coefficients, confint(donner.log)))
```

```
##                    OR     2.5 %    97.5 %
## (Intercept) 1.7398953 0.7748972 4.0431170
## Age         0.9650211 0.9348223 0.9930661
## fem         2.9094868 1.1543365 7.7529827
```

2.2 Odd ratio for Survival with 10 year increase

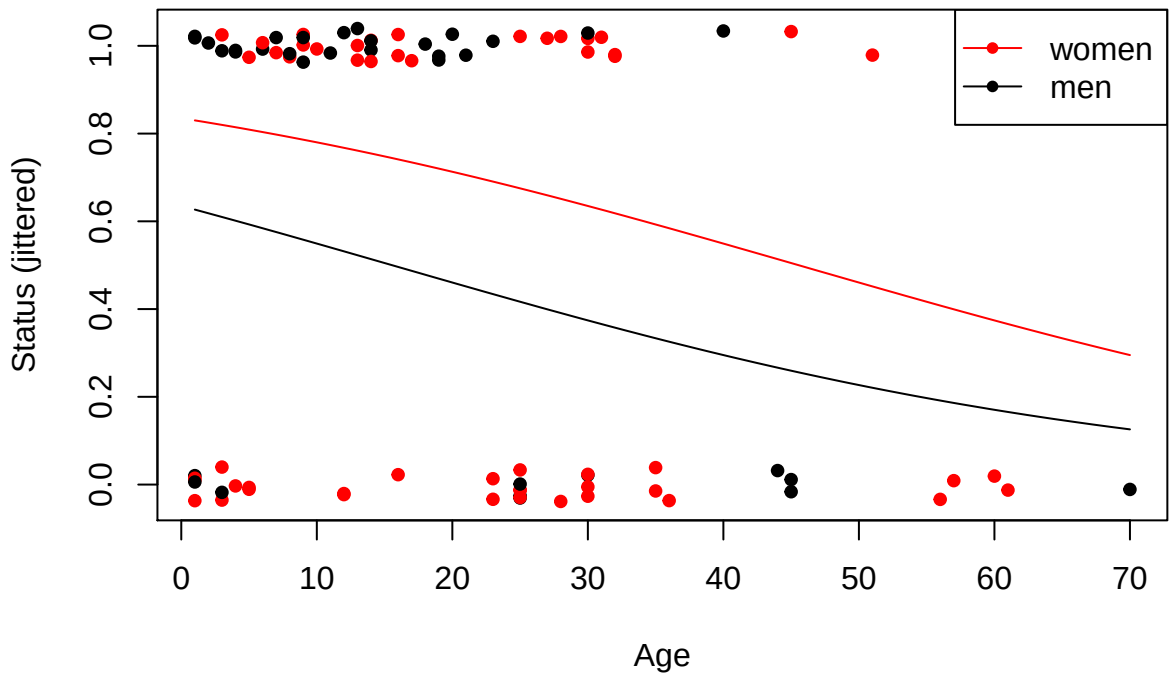
```
#exp(donner.log$coefficients*10)
exp(c(OR=donner.log$coefficients[2]*10, confint(donner.log)[2,]*10))
```

```
##    OR.Age     2.5 %    97.5 %
## 0.7004356 0.5096720 0.9327850
```

2.3 Plotting the logit curve

```
logit = function(x){
  log(x/(1-x))
}
ilogit = function(x,a,b){
  exp(a+b*x)/(1+exp(a+b*x))
}

# Plotting survival for men vs women
cl = coef(donner.log)
plot(donner.na$Age, jitter(donner.na$Outcome,.2), col=Sex, pch=20, cex=1.2, xlab="Age", ylab="Status (j
curve(ilogit(cl[1]+cl[2]*x+cl[3]*0, 0, 1), add=T)
curve(ilogit(cl[1]+cl[2]*x+cl[3]*1, 0, 1), add=T, col="red")
legend("topright", pch=20, lty="solid", col=c("red","black"), c("women","men"))
```



2.4 Predicted probabilities of survival

```
newdata2 = data.frame(fem=1, Age=mean(donner.na$Age))
newdata2$greP = predict(donner.log, newdata=newdata2,type="response")
newdata2
```

```
##   fem      Age     greP
```

```
## 1   1 20.22727 0.711279
```

```
newdata3 = data.frame(fem=0, Age=mean(donner.na$Age))
newdata3$greP = predict(donner.log, newdata=newdata3,type="response")
newdata3
```

```
##   fem      Age      greP
## 1   0 20.22727 0.4585025
```

```
newdata4 = data.frame(fem=c(0,1), Age=mean(donner.na$Age))
newdata4$greP = predict(donner.log, newdata=newdata4,type="response")
newdata4
```

```
##   fem      Age      greP
## 1   0 20.22727 0.4585025
## 2   1 20.22727 0.7112790
```

2.5 Interaction model

```
m4 = glm(Outcome ~ Age*fem, data=donner.na, family=binomial(link="logit"))
summary(m4)
```

```
##
## Call:
## glm(formula = Outcome ~ Age * fem, family = binomial(link = "logit"),
##     data = donner.na)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9888  -1.0532   0.5961   1.0727   1.6317
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.39779    0.48139   0.826    0.409
## Age         -0.02789    0.01911  -1.460    0.144
## fem          1.47859    0.82469   1.793    0.073 .
## Age:fem     -0.01977    0.03166  -0.624    0.532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 120.86  on 87  degrees of freedom
## Residual deviance: 108.47  on 84  degrees of freedom
## AIC: 116.47
##
## Number of Fisher Scoring iterations: 4
```

3 Model Selection

3.1 Fitting the models

```
donner.list=list()

donner.list[[1]] = glm(Outcome ~ Age,data=donner.na, family=binomial(link="logit"))
donner.list[[2]] = glm(Outcome ~ fem,data=donner.na, family=binomial(link="logit"))
donner.list[[3]] = glm(Outcome ~ Age+fem, data=donner.na, family=binomial(link="logit"))
donner.list[[4]] = glm(Outcome ~ Age*fem, data=donner.na, family=binomial(link="logit"))

donner.modnames = c("Age", "Sex", "Age+Sex", "Age+Sex+Age:Sex")
```

3.2 Akaike Weights

```
library('AICcmodavg')
donner.aictab=aictab(cand.set = donner.list, modnames = donner.modnames)
donner.aictab
```

```
##
## Model selection based on AICc:
##
##                 K   AICc Delta_AICc AICcWt Cum.Wt     LL
## Age+Sex         3 115.15       0.00   0.56   0.56 -54.43
## Age+Sex+Age:Sex 4 116.95       1.80   0.23   0.79 -54.23
## Age             2 118.16       3.01   0.13   0.92 -57.01
## Sex             2 119.02       3.87   0.08   1.00 -57.44
```

4 Model Averaging

```
modavg(cand.set=donner.list, parm="Age", second.ord=TRUE, modnames=donner.modnames, uncond.se="revised"
```

```
##
## Multimodel inference on "Age" based on AICc
##
## AICc table used to obtain model-averaged estimate:
##
##         K   AICc Delta_AICc AICcWt Estimate   SE
## Age     2 118.16       3.01   0.18    -0.04 0.01
## Age+Sex 3 115.15       0.00   0.82    -0.04 0.02
##
## Model-averaged estimate: -0.04
## Unconditional SE: 0.02
## 95% Unconditional confidence interval: -0.07, -0.01
```

```
modavg(cand.set=donner.list, parm="fem", second.ord=TRUE, modnames=donner.modnames, uncond.se="revised"
```

```
##
## Multimodel inference on "fem" based on AICc
##
## AICc table used to obtain model-averaged estimate:
##
##         K   AICc Delta_AICc AICcWt Estimate   SE
## Sex     2 119.02       3.87   0.13     1.11 0.46
## Age+Sex 3 115.15       0.00   0.87     1.07 0.48
##
## Model-averaged estimate: 1.07
## Unconditional SE: 0.48
## 95% Unconditional confidence interval: 0.13, 2.01
```

5 Odds Ratio with interaction model

```
x = seq(1, 70, 0.01)
y = exp(coef(m4)[3]+coef(m4)[4]*x)

plot(x, y, type="n", ylim=c(0.7,4.5), xlab="Age", ylab="Odds ratio", main="Odds ratio for gender")
lines(x, y, lty=1, col="red")
abline(h=1)
```