STA303/1002 - Week 5 R Markdown

February 4-8, 2018

Case Study 3: The Data

Get the data (from R library):

```
#load Sleuth3 R data library; see case2001
library(Sleuth3)
#Donner party survival data
donner = case2001
str(donner)
```

```
## 'data.frame':    45 obs. of  3 variables:
##  $ Age   : int  23 40 40 30 28 40 45 62 65 45 ...
##  $ Sex   : Factor w/ 2 levels "Female","Male": 2 1 2 2 2 2 1 2 2 1 ...
##  $ Status: Factor w/ 2 levels "Died","Survived": 1 2 2 1 1 1 1 1 1 1 ...
```

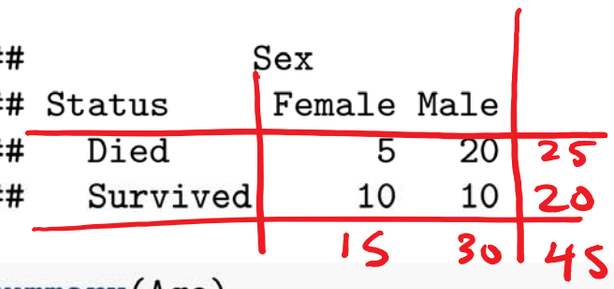
```
attach(donner)
head(donner)
```

```
##   Age    Sex   Status
## 1  23   Male     Died
## 2  40 Female Survived
## 3  40   Male Survived
## 4  30   Male     Died
## 5  28   Male     Died
## 6  40   Male     Died
```

Case Study 3: Summarizing the data

```
#two-way contingency table for status by sex
#check that cell counts>0
xtabs(~Status+Sex, data=donner)
```

```
##            Sex
## Status      Female Male
##   Died           5   20
##   Survived      10   10
```




```
summary(Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    15.0    24.0    28.0    31.8    40.0    65.0
```

Case Study 3: Marginal Mean Ages

```
tapply(Age, Status, mean)
```

```
##      Died Survived
##     35.48    27.20
```

```
tapply(Age, Sex, mean)
```

```
##   Female     Male
## 31.06667 32.16667
```

```
fita<-glm(Status~Age+Sex, family=binomial, data=donner)
```

Case Study 2: Additve model summary

```
##
## Call:
## glm(formula = Status ~ Age + Sex, family = binomial, data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7445  -1.0441  -0.3029   0.8877   2.0472
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.23041    1.38686   2.329   0.0198 *
## Age         -0.07820    0.03728  -2.097   0.0359 *
## SexMale     -1.59729    0.75547  -2.114   0.0345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4
```

Case Study 3: ANOVA table

```r
anova(fita)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Status
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev
## NULL                    44     61.827
## Age   1   5.5358        43     56.291
## Sex   1   5.0344        42     51.256
```

Case Study 3: Modelling "Died"

```
status=relevel(Status, ref="Survived")
fitad<-glm(status~Age+Sex, family=binomial, data=donner)
summary(fitad)
```

```
##
## Call:
## glm(formula = status ~ Age + Sex, family = binomial, data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0472  -0.8877   0.3029   1.0441   1.7445
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.23041    1.38686  -2.329   0.0198 *
## Age          0.07820    0.03728   2.097   0.0359 *
## SexMale      1.59729    0.75547   2.114   0.0345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
```

$\pi_i = P(\text{"died"})$

Case Study 3: Sex Reference group as "Male"

```
sex=relevel(Sex, ref="Male")
fitadf<-glm(status~Age+sex, family=binomial, data=donner)
summary(fitadf)
```

```
##
## Call:
## glm(formula = status ~ Age + sex, family = binomial, data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0472  -0.8877   0.3029   1.0441   1.7445
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.63312    1.11018  -1.471   0.1413
## Age          0.07820    0.03728   2.097   0.0359 *
## sexFemale   -1.59729    0.75547  -2.114   0.0345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
```

$\pi = P(\text{"died"})$

$1_F = 1$

Case Study 3: Additive model for Survived

```
fitasf<-glm(Status~Age+sex, family=binomial, data=donner)
summary(fitasf)
```

```
##
## Call:
## glm(formula = Status ~ Age + sex, family = binomial, data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7445  -1.0441  -0.3029   0.8877   2.0472
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.63312    1.11018   1.471   0.1413
## Age         -0.07820    0.03728  -2.097   0.0359 *
## sexFemale    1.59729    0.75547   2.114   0.0345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 51.256  on 42  degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4
```

$\pi = P(\text{"survived"})$

1 F

Case Study 3: Higher Order Model with 3 higher order/interaction terms

```
fitfull<-glm(Status~Age+sex+Age:sex+I(Age^2)+I(Age^2):sex, family=binomial, dat
summary(fitfull)
```

```
##
## Call:
## glm(formula = Status ~ Age + sex + Age:sex + I(Age^2) + I(Age^2):sex,
##     family = binomial, data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3396  -0.9757  -0.3438   0.5269   1.5901
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -3.318484   3.940184  -0.842    0.400
## Age                 0.183031   0.226632   0.808    0.419
## sexFemale           0.265286  10.455222   0.025    0.980
## I(Age^2)           -0.002803   0.002985  -0.939    0.348
## Age:sexFemale       0.299877   0.696050   0.431    0.667
## sexFemale:I(Age^2) -0.007356   0.010689  -0.688    0.491
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 45.361  on 39  degrees of freedom
## AIC: 57.361
```

Case Study 3: Interaction Model, Age*Sex

```
fitas<-glm(Status~Age*sex, family=binomial, data=donner)
summary(fitas)
```

```
##
## Call:
## glm(formula = Status ~ Age * sex, family = binomial, data = donner)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2279  -0.9388  -0.5550   0.7794   1.6998
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.31834    1.13103   0.281   0.7784
## Age           -0.03248    0.03527  -0.921   0.3571
## sexFemale      6.92805    3.39887   2.038   0.0415 *
## Age:sexFemale -0.16160    0.09426  -1.714   0.0865 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.827  on 44  degrees of freedom
## Residual deviance: 47.346  on 41  degrees of freedom
## AIC: 55.346
##
```