4050_A4_Q1.R

SOPH

Mon Mar 27 19:31:16 2017

```
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 3.2.3
```

```
data(femsmoke)
attach(femsmoke)

# Fitting a binomial glm
Dead<-subset(femsmoke, dead == 'yes')
Alive<-subset(femsmoke, dead =='no')

totalDead = apply(Dead,1,function(x) sum(Dead$y[(Dead$age ==x[4] & Dead$smoker ==x[2])]))
totalAlive = apply(Alive,1,function(x) sum(Alive$y[(Alive$age ==x[4] & Alive$smoker ==x[2])]))
totals = (totalDead+totalAlive)

Smoke<-data.frame(Dead,totals)
age<-as.factor(age)
smoker<-as.factor(smoker)

# Identify the best fitting model
binmod.0<-glm(cbind(y,totals-y)~age + smoker + age*smoker, family = binomial, data = Smoke)
summary(binmod.0)
```

```
##
## Call:
## glm(formula = cbind(y, totals - y) ~ age + smoker + age * smoker,
##     family = binomial, data = Smoke)
##
## Deviance Residuals:
##  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -3.277e+00  7.203e-01  -4.550 5.38e-06 ***
## age25-34          -4.200e-01  9.276e-01  -0.453  0.65068
## age35-44           1.362e+00  7.751e-01   1.758  0.07882 .
## age45-54           1.938e+00  7.521e-01   2.577  0.00996 **
## age55-64           3.050e+00  7.444e-01   4.097 4.18e-05 ***
## age65-74           4.699e+00  8.344e-01   5.631 1.79e-08 ***
## age75+             2.863e+01  5.387e+04   0.001  0.99958
## smokerno          -8.337e-01  1.239e+00  -0.673  0.50103
## age25-34:smokerno  1.116e+00  1.443e+00   0.773  0.43923
## age35-44:smokerno -4.174e-02  1.330e+00  -0.031  0.97496
## age45-54:smokerno  4.679e-01  1.296e+00   0.361  0.71816
## age55-64:smokerno  3.552e-01  1.268e+00   0.280  0.77937
## age65-74:smokerno  6.953e-01  1.326e+00   0.524  0.60004
```

```
## age75+:smokerno    2.352e+00  7.478e+04   0.000  0.99997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6.4150e+02  on 13  degrees of freedom
## Residual deviance: 5.2696e-10  on  0  degrees of freedom
## AIC: 74.996
##
## Number of Fisher Scoring iterations: 22
```

```
drop1(binmod.0, test = 'Chi')
```

```
## Single term deletions
##
## Model:
## cbind(y, totals - y) ~ age + smoker + age * smoker
##            Df Deviance    AIC    LRT Pr(>Chi)
## <none>          0.0000 74.996
## age:smoker  6   2.3809 65.377 2.3809   0.8815
```

```
binmod.1<-glm(cbind(y,totals-y)~smoker + age, family = binomial, data = Smoke)
summary(binmod.1)
```

```
##
## Call:
## glm(formula = cbind(y, totals - y) ~ smoker + age, family = binomial,
##     data = Smoke)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.72545  -0.22836   0.00005   0.19146   0.68162
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.4327     0.5901  -5.817 6.00e-09 ***
## smokerno       -0.4274     0.1770  -2.414 0.015762 *
## age25-34        0.1201     0.6865   0.175 0.861178
## age35-44        1.3411     0.6286   2.134 0.032874 *
## age45-54        2.1134     0.6121   3.453 0.000555 ***
## age55-64        3.1808     0.6006   5.296 1.18e-07 ***
## age65-74        5.0880     0.6195   8.213  < 2e-16 ***
## age75+         27.8073 11293.1430   0.002 0.998035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 641.4963  on 13  degrees of freedom
## Residual deviance:   2.3809  on  6  degrees of freedom
## AIC: 65.377
##
## Number of Fisher Scoring iterations: 20
```

```r
drop1(binmod.1, test = 'Chi')
```

```
## Single term deletions
##
## Model:
## cbind(y, totals - y) ~ smoker + age
##        Df Deviance    AIC    LRT Pr(>Chi)
## <none>        2.38  65.38
## smoker  1     8.33  69.32   5.95  0.01475 *
## age     6   632.30 683.29 629.92  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
binmod.2<-glm(cbind(y,totals-y)~age, family = binomial, data = Smoke)
summary(binmod.2)
```

```
##
## Call:
## glm(formula = cbind(y, totals - y) ~ age, family = binomial,
##     data = Smoke)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.36174  -0.45445   0.00013   0.42093   1.27575
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.6376     0.5849  -6.219 5.00e-10 ***
## age25-34      0.1076     0.6861   0.157 0.875435
## age35-44      1.3398     0.6281   2.133 0.032920 *
## age45-54      2.1712     0.6113   3.552 0.000382 ***
## age55-64      3.1717     0.6000   5.286 1.25e-07 ***
## age65-74      4.9498     0.6151   8.047 8.49e-16 ***
## age75+       25.7806  4445.2217   0.006 0.995373
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 641.4963  on 13  degrees of freedom
## Residual deviance:   8.3269  on  7  degrees of freedom
## AIC: 69.323
##
## Number of Fisher Scoring iterations: 18
```

```r
drop1(binmod.2, test = 'Chi')
```

```
## Single term deletions
##
## Model:
## cbind(y, totals - y) ~ age
##        Df Deviance    AIC    LRT  Pr(>Chi)
```

```
## <none>        8.33  69.32
## age     6   641.50 690.49 633.17 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
binmod.3<-glm(cbind(y,totals-y)~smoker, family = binomial, data = Smoke)
summary(binmod.3)
```

```
##
## Call:
## glm(formula = cbind(y, totals - y) ~ smoker, family = binomial,
##     data = Smoke)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -9.052  -5.674  -1.869   5.776  12.173
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.15910    0.09722 -11.923  < 2e-16 ***
## smokerno     0.37858    0.12566   3.013  0.00259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 641.5  on 13  degrees of freedom
## Residual deviance: 632.3  on 12  degrees of freedom
## AIC: 683.29
##
## Number of Fisher Scoring iterations: 4
```

```
drop1(binmod.3, test = 'Chi')
```

```
## Single term deletions
##
## Model:
## cbind(y, totals - y) ~ smoker
##        Df Deviance    AIC    LRT Pr(>Chi)
## <none>       632.3 683.29
## smoker  1    641.5 690.49 9.2003  0.00242 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```