Categorical Data Analysis: Cochran-Mantel-Haenzsel Analysis for Stratified 2x2 tables

```
> mydat <-
read.csv(file="http://sites.williams.edu/bklingen/files/2012/02/smoking.csv")
> mydat
  Agegroup S.alive S.dead NS.alive NS.dead
1    18-34     174      5      213       6
2    35-54     198     41      180      19
3    55-64      64     51       81      40
4      65+       7     42       28     165
> rownames(mydat) <- mydat[,1]
> mydat <- mydat[,-1]
> mydat
      S.alive S.dead NS.alive NS.dead
18-34     174      5      213       6
35-54     198     41      180      19
55-64      64     51       81      40
65+         7     42       28     165
> ## marginal table:
> colSums(mydat)
 S.alive   S.dead NS.alive  NS.dead
     443      139      502      230
> ## Preparations for CMH test:
> mydat1 <- mydat[,c(1,3,2,4)]
> my3dtable <- array(unlist(t(mydat1)), dim=c(2,2,4),
+ dimnames=list(Smoking=c("Yes", "No"),
+               Survival=c("Yes","No"),
+               Age=c("18-34", "35-54", "55-64", "65+")))
> mantelhaen.test(my3dtable, correct=FALSE)

        Mantel-Haenszel chi-squared test without continuity correction

data:  my3dtable
Mantel-Haenszel X-squared = 6.5896, df = 1, p-value = 0.01026
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.4546701 0.9024477
sample estimates:
common odds ratio
        0.6405592

> ################
> table.long <- data.frame(ftable(my3dtable))
> table.long
   Smoking Survival   Age Freq
1      Yes      Yes 18-34  174
2       No      Yes 18-34  213
3      Yes       No 18-34    5
4       No       No 18-34    6
5      Yes      Yes 35-54  198
6       No      Yes 35-54  180
…
15     Yes       No   65+   42
16      No       No   65+  165
> index <- table.long$Survival=="Yes"
> table1 <- cbind(table.long[index,-2], table.long[!index,4])
```

```
> names(table1)[3:4] <- c("S","F")
> table1
   Smoking   Age   S   F
1      Yes 18-34 174   5
2       No 18-34 213   6
5      Yes 35-54 198  41
6       No 35-54 180  19
9      Yes 55-64  64  51
10      No 55-64  81  40
13     Yes   65+   7  42
14      No   65+  28 165
> logistic.fit <- glm(cbind(S,F)~ Age + (Smoking=="Yes"), family="binomial",
data=table1)
> summary(logistic.fit)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)           3.7868     0.3212  11.790  < 2e-16 ***
Age35-54             -1.6826     0.3364  -5.001  5.7e-07 ***
Age55-64             -3.0958     0.3343  -9.260  < 2e-16 ***
Age65+               -5.4842     0.3635 -15.088  < 2e-16 ***
Smoking == "Yes"TRUE -0.4500     0.1757  -2.561   0.0104 *
---
> anova(logistic.fit, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(S, F)

Terms added sequentially (first to last)


                Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                7     603.67
Age              3   595.03         4       8.64 < 2.2e-16 ***
Smoking == "Yes" 1     6.71         3       1.93  0.009579 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1



> ### Is the assumption about a common odds in each partial table plausible?
> ### Test the current model against a more complicated one
> ### that allows an interaction between age group and smoking:
> logistic.fit1 <- glm(cbind(S,F)~ Age*(Smoking=="Yes"), family="binomial",
data=table1)
> summary(logistic.fit1)

Call:
glm(formula = cbind(S, F) ~ Age * (Smoking == "Yes"), family = "binomial",
    data = table1)

Deviance Residuals:
[1]  0  0  0  0  0  0  0  0
```

```
Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                     3.5695     0.4140    8.62  < 2e-16 ***
Age35-54                       -1.3210     0.4791   -2.76   0.0058 **
Age55-64                       -2.8640     0.4568   -6.27  3.6e-10 ***
Age65+                         -5.3433     0.4617  -11.57  < 2e-16 ***
Smoking == "Yes"TRUE           -0.0199     0.6141   -0.03   0.9741
Age35-54:Smoking == "Yes"TRUE  -0.6539     0.6817   -0.96   0.3375
Age55-64:Smoking == "Yes"TRUE  -0.4586     0.6706   -0.68   0.4941
Age65+:Smoking == "Yes"TRUE     0.0019     0.7652    0.00   0.9980
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.0367e+02  on 7  degrees of freedom
Residual deviance: 9.0816e-14  on 0  degrees of freedom
AIC: 52.13

Number of Fisher Scoring iterations: 3

> anova(logistic.fit1, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(S, F)

Terms added sequentially (first to last)


                     Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                                     7        604
Age                   3      595         4          9   <2e-16 ***
Smoking == "Yes"      1        7         3          2   0.0096 **
Age:Smoking == "Yes"  3        2         0          0   0.5878
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>

> #### Estimating Common Odds Ratio
> confint(logistic.fit2)
Waiting for profiling to be done...
                            2.5      97.5
(Intercept)           3.2052621  4.4752039
Age35-54             -2.3939910 -1.0625275
Age55-64             -3.8045653 -2.4820767
Age65+               -6.2466077 -4.8125639
Smoking == "Yes"TRUE -0.7984353 -0.1087374

> exp(c(-0.7984353, -0.1087374))
[1] 0.4500326 0.8969659
> exp(-0.45)
[1] 0.6376282
```