The Berkeley Graduate Admissions Data

```
> UCB
, , Admit = Admitted

    Gender
Dept Female Male
   A     89  512
   B     17  353
   C    202  120
   D    131  138
   E     94   53
   F     24   22

, , Admit = Rejected

    Gender
Dept Female Male
   A     19  313
   B      8  207
   C    391  205
   D    244  279
   E    299  138
   F    317  351

> is.table(UCB) # T
[1] TRUE
> summary(UCB)  # X2 for complete independence = 2000.3, df=16
Call: xtabs(formula = Freq ~ Dept + Gender + Admit, data = berkeley)
Number of cases in table: 4526
Number of factors: 3
Test for independence of all factors:
 Chisq = 2000.3, df = 16, p-value = 0
```

```
> all2ways <- loglin(UCB,margin=list(c(1,2),c(1,3),c(2,3))); all2ways
7 iterations: deviation 0.04308377
$lrt
[1] 20.20428

$pearson
[1] 18.82298

$df
[1] 5

$margin
$margin[[1]]
[1] "Dept"   "Gender"

$margin[[2]]
[1] "Dept"  "Admit"

$margin[[3]]
[1] "Gender" "Admit"


> 1-pchisq(all2ways$lrt,df=all2ways$df) # p-value for H0: mu123=0
[1] 0.001144076
> # So the relationship between gender and admission DEPENDS on department
>
```

Let's look at some 2-dimensional marginal tables

```
> sex_by_admit = xtabs(Freq ~ Gender + Admit, data = berkeley)
> sex_by_admit
        Admit
Gender   Admitted Rejected
  Female      557     1278
  Male       1198     1493
> round(100*prop.table(sex_by_admit,1),2) # Row percentages
        Admit
Gender   Admitted Rejected
  Female    30.35    69.65
  Male      44.52    55.48
> summary(sex_by_admit)
Call: xtabs(formula = Freq ~ Gender + Admit, data = berkeley)
Number of cases in table: 4526
Number of factors: 2
Test for independence of all factors:
 Chisq = 92.21, df = 1, p-value = 7.814e-22

>
> sex_by_dept = xtabs(Freq ~ Gender + Dept, data = berkeley)
> sex_by_dept
        Dept
Gender     A   B   C   D   E   F
  Female 108  25 593 375 393 341
  Male   825 560 325 417 191 373
> round(100*prop.table(sex_by_dept,1),2) # Row percentages
        Dept
Gender       A     B     C     D     E     F
  Female  5.89  1.36 32.32 20.44 21.42 18.58
  Male   30.66 20.81 12.08 15.50  7.10 13.86
> summary(sex_by_dept)
Call: xtabs(formula = Freq ~ Gender + Dept, data = berkeley)
Number of cases in table: 4526
Number of factors: 2
Test for independence of all factors:
 Chisq = 1068.4, df = 5, p-value = 9.444e-229
>
> dept_by_admit = xtabs(Freq ~ Dept + Admit, data = berkeley)
> dept_by_admit
    Admit
Dept Admitted Rejected
   A      601      332
   B      370      215
   C      322      596
   D      269      523
   E      147      437
   F       46      668
```

```
> round(100*prop.table(dept_by_admit,1),2) # Row percentages
    Admit
Dept Admitted Rejected
   A    64.42    35.58
   B    63.25    36.75
   C    35.08    64.92
   D    33.96    66.04
   E    25.17    74.83
   F     6.44    93.56
> summary(dept_by_admit)
Call: xtabs(formula = Freq ~ Dept + Admit, data = berkeley)
Number of cases in table: 4526
Number of factors: 2
Test for independence of all factors:
 Chisq = 778.9, df = 5, p-value = 4.23e-166
>
> # What is going on here? Assemble a good table.
> admitper <- round(100*prop.table(dept_by_admit,1),2)
> genderper <- round(100*prop.table(sex_by_dept,1),2)
> cbind(admitper[,1],t(genderper))
        Female  Male
A 64.42   5.89 30.66
B 63.25   1.36 20.81
C 35.08  32.32 12.08
D 33.96  20.44 15.50
E 25.17  21.42  7.10
F  6.44  18.58 13.86
>
```

```
> # Look at gender by admit controlling for department
> ucb <- xtabs(Freq ~ Gender + Admit + Dept, data = berkeley)
> # That's 6 2x2 tables -- hard to look at
> dept <- dimnames(ucb)$Dept; dept
[1] "A" "B" "C" "D" "E" "F"
> totalgsq <- 0
> for(k in 1:6)
+     {
+     cat("\n", "   Department ",dept[k],"\n")
+     cat("    ------------- \n\n")
+     freq <- ucb[,,k]
+     rowper <- round(100*prop.table(freq,1),2)
+     llm <- loglin(freq,margin=list(1,2),print=F) # Don't print iterations
+     g2 <- llm$lrt; df = llm$df; pval = 1-pchisq(g2,df)
+     cat("   Observed Frequencies \n\n")
+     print(freq)
+     cat("\n    Row Percentages \n\n")
+     print(rowper)
+     cat("\n G-squared = ",g2,", df = ",df,", p = ",pval,"\n")
+     totalgsq = totalgsq + g2
+     }

    Department  A
    -------------

   Observed Frequencies

        Admit
Gender   Admitted Rejected
  Female       89       19
  Male        512      313

    Row Percentages

        Admit
Gender   Admitted Rejected
  Female    82.41    17.59
  Male      62.06    37.94

 G-squared =  19.05401 , df =  1 , p =  1.270705e-05
```

Department B

Observed Frequencies

Gender	Admit: Admitted	Rejected
Female	17	8
Male	353	207

Row Percentages

Gender	Admit: Admitted	Rejected
Female	68.00	32.00
Male	63.04	36.96

G-squared = 0.2586429 , df = 1 , p = 0.611054

Department C

Observed Frequencies

Gender	Admit: Admitted	Rejected
Female	202	391
Male	120	205

Row Percentages

Gender	Admit: Admitted	Rejected
Female	34.06	65.94
Male	36.92	63.08

G-squared = 0.7509844 , df = 1 , p = 0.3861648

Department D

Observed Frequencies

Gender	Admit: Admitted	Rejected
Female	131	244
Male	138	279

Row Percentages

Gender	Admit: Admitted	Rejected
Female	34.93	65.07
Male	33.09	66.91

G-squared = 0.2978665 , df = 1 , p = 0.585223

Department E

Observed Frequencies

Gender	Admit: Admitted	Rejected
Female	94	299
Male	53	138

Row Percentages

Gender	Admit: Admitted	Rejected
Female	23.92	76.08
Male	27.75	72.25

G-squared = 0.9903864 , df = 1 , p = 0.3196480

```
    Department  F
    -------------

   Observed Frequencies

        Admit
Gender   Admitted Rejected
  Female       24      317
  Male         22      351

    Row Percentages

        Admit
Gender   Admitted Rejected
  Female     7.04    92.96
  Male       5.90    94.10

 G-squared =  0.3836167 , df =  1 , p =  0.535674
>
> # Model of conditional independence should not fit, with
> # G-squared = totalgsq
> loglin(ucb,margin=list(c("Gender","Dept"),c("Dept","Admit")))$lrt
2 iterations: deviation 5.684342e-14
[1] 21.73551
> totalgsq
[1] 21.73551
> 1-pchisq(totalgsq,6)
[1] 0.001351993
```