MATH 523: EXAMPLE - MODELS FOR SQUARE CONTINENCY TABLES

The following data represent recorded voting for a cohort of 1651 voters in successive elections; each member of the cohort voted for one of four parties in each election:

		Election 2			
		1	2	3	4
Election 1	1	850	35	25	6
	2	9	286	21	6
	3	3	26	185	5
	4	3	26	27	138

We can consider independence, quasi-independence, symmetry and quasi-symmetry models; perhaps the last two address the most interesting questions in research terms. The following code can be used to fit these models using the `gnm` library in `R`:

```
library(gnm)

elect.count<-c(850,35,25,6,9,286,21,6,3,26,185,5,3,26,27,138)
elect.one<-gl(4,4)
elect.two<-gl(4,1,16)

elections<-data.frame(count=elect.count,Elect1=elect.one,Elect2=elect.two)

independence<-glm(count~Elect1+Elect2,family=poisson,data=elections)
quasi.indep<-glm(count~Elect1+Elect2+Diag(Elect1,Elect2),family=poisson,data=elections)
symmetry<-glm(count~Symm(Elect1,Elect2),family=poisson,data=elections)
quasi.symm<-glm(count~Elect1+Elect2+Symm(Elect1,Elect2),family=poisson,data=elections)
comparison1<-anova(independence,symmetry,quasi.indep,quasi.symm)
print(comparison1)

> print(comparison1)
Analysis of Deviance Table

Model 1: count ~ Elect1 + Elect2
Model 2: count ~ Elect1 + Elect2 + Diag(Elect1, Elect2)
Model 3: count ~ Symm(Elect1, Elect2)
Model 4: count ~ Elect1 + Elect2 + Symm(Elect1, Elect2)
  Resid. Df Resid. Dev Df Deviance
1         9    2461.65
2         5       5.18  4  2456.48
3         6      67.81 -1   -62.64
4         3       2.49  3    65.32
```

The following results are obtained for the fits of the models (p is the number of parameters fitted):

Model	Deviance	Res. Df.	p
Independence	2461.655	9	7
Quasi-Independence	5.175	5	11
Symmetry	67.814	6	10
Quasi-symmetry	2.494	3	13

On deviance terms, clearly the two quasi- models fit better, and each appears to be an adequate model given the deviance/residual degrees of freedom comparison. If a direct comparison were possible,

then comparing the quasi-independence model (simpler) with the quasi-symmetry model (more complex) would be based on $\triangle$Dev. $= 2.682$, $\triangle$df $= 2$, implying a conclusion that we cannot reject the hypothesis that the simpler model is an adequate simplification of the more complex model. However, **these models are not nested**.

Recall the definitions of the quasi- models: taking the $j = k = 1$ case as the baseline group, and utilizing the usual identifiability constraints, we have

- **Quasi-Independence:**

$$\log \mu_{jk} = \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_j \mathbb{1}\{j = k\} = \begin{cases} \beta + \gamma_1 & j = 1, k = 1 \\ \beta + \beta_j^{(R)} & j = 2, 3, 4, k = 1 \\ \beta + \beta_k^{(C)} & j = 1, k = 2, 3, 4 \\ \beta + \beta_j^{(R)} + \beta_j^{(C)} + \gamma_j & j = k = 2, 3, 4 \\ \beta + \beta_j^{(R)} + \beta_k^{(C)} & j = 2, 3, 4, j = 2, 3, 4, j \neq k. \end{cases}$$

  Thus the eleven parameters are

$$\beta, \beta_2^{(C)}, \beta_3^{(C)}, \beta_4^{(C)}, \beta_2^{(R)}, \beta_3^{(R)}, \beta_4^{(R)}, \gamma_1, \gamma_2, \gamma_3, \gamma_4$$

- **Quasi-Symmetry:** $\log \mu_{jk} = \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_{jk}$, where $\gamma_{jk} = \gamma_{kj}$ for $j \neq k$. That is,

$$\log \mu_{jk} = \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_{jk} = \begin{cases} \beta & j = 1, k = 1 \\ \beta + \beta_j^{(R)} + \gamma_{jk} & j = 2, 3, 4, k = 1 \\ \beta + \beta_k^{(C)} + \gamma_{jk} & j = 1, k = 2, 3, 4 \\ \beta + \beta_j^{(R)} + \beta_k^{(C)} + \gamma_{jk} & j, k = 2, 3, 4 \end{cases}$$

  Thus the thirteen parameters are

$$\beta, \beta_2^{(C)}, \beta_3^{(C)}, \beta_4^{(C)}, \beta_2^{(R)}, \beta_3^{(R)}, \beta_4^{(R)}, \gamma_{12}, \gamma_{13}, \gamma_{14}, \gamma_{23}, \gamma_{24}, \gamma_{34}$$

It is not possible to restrict the quasi-symmetry model in such a way as to produce the quasi-independence model. On the basis of these results alone, it is not possible to assess which model is more appropriate.