36-720: Log-Linear Models: Three-Way Tables

Brian Junker

September 10, 2007

- Hierarchical Specification of Log-Linear Models
- Higher-Dimensional Tables
- Digression: Why Poisson If We Believe Multinomial?
- Product Multinomial Log-linear Models

Hierarchical Specification of Log-Linear Models

A full set of log-linear models for 3-way tables that we have considered so far include:

	Log-Linear Model	Generator
$M^{(-1)}$	$\log m_{ijk} = u$	[0]
$M^{(0)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$	[1][2][3]
$M^{(1)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}$	[1][23]
$M^{(2)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)}$	[2][13]
$M^{(3)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$	[3][12]
$M^{(4)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$	[13][23]
$M^{(5)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)}$	[12][23]
$M^{(6)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$	[12][13]
$M^{(7)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$	[12][13][23]
$M^{(8)}$	$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$	[123]

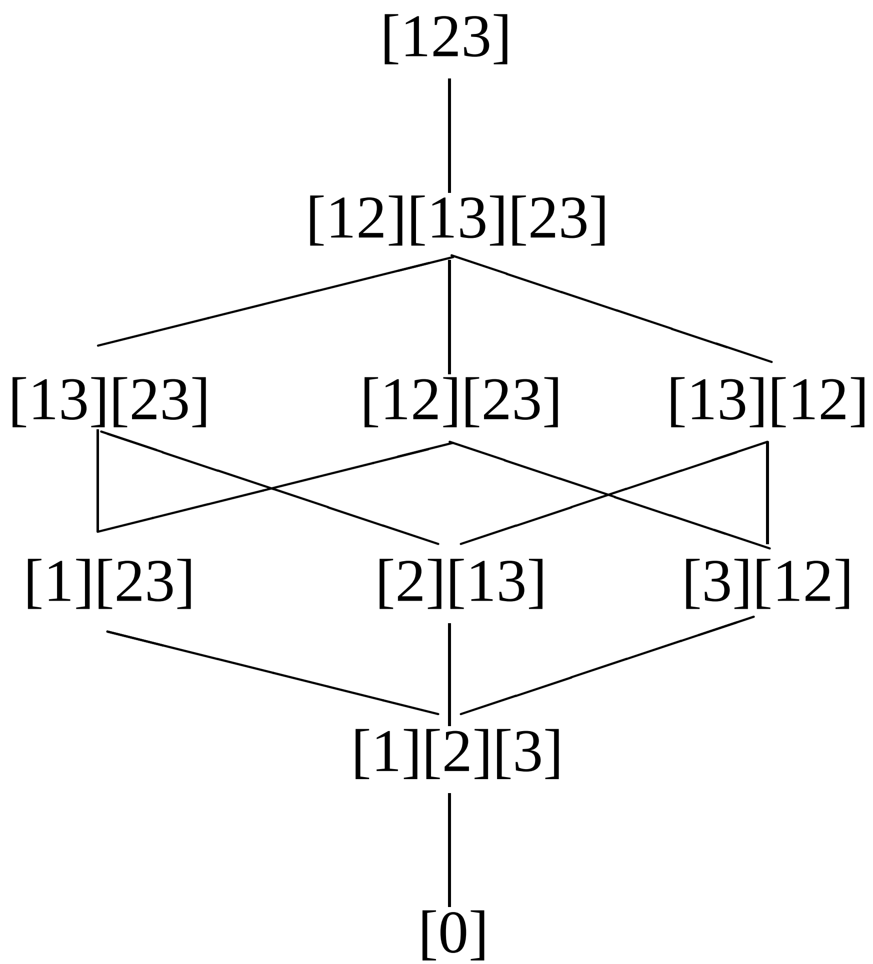
These model specifications have several noteworthy features:

- They obey the *hierarchy principle*: If the k-way interaction is in the model then every lower order interaction and main effect is also in the model. For example, since $M^{(7)}$ contains $u_{12(ik)}$ terms then we also know it contains $u_{1(i)}$ and $u_{2(j)}$ terms). *Under the hierarchy principle, nested models are obtained by adding or dropping higher-order interactions in the model.*
- Following the hierarchy principle, a model can be completely specified by specifying the indices of the highest-order interactions (the *generators*). E.g.:

$$\begin{aligned} [1][23] \quad &\Rightarrow \quad u_{1(i)} \text{ and } u_{23(jk)} \text{ are in the model} \\ &\Rightarrow \quad u + u_{1(i)} \text{ and } u + u_{2(j)} + u_{3(k)} + u_{23(jk)} \text{ are in the model} \\ &\Rightarrow \quad \log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} \text{ is the model} \end{aligned}$$

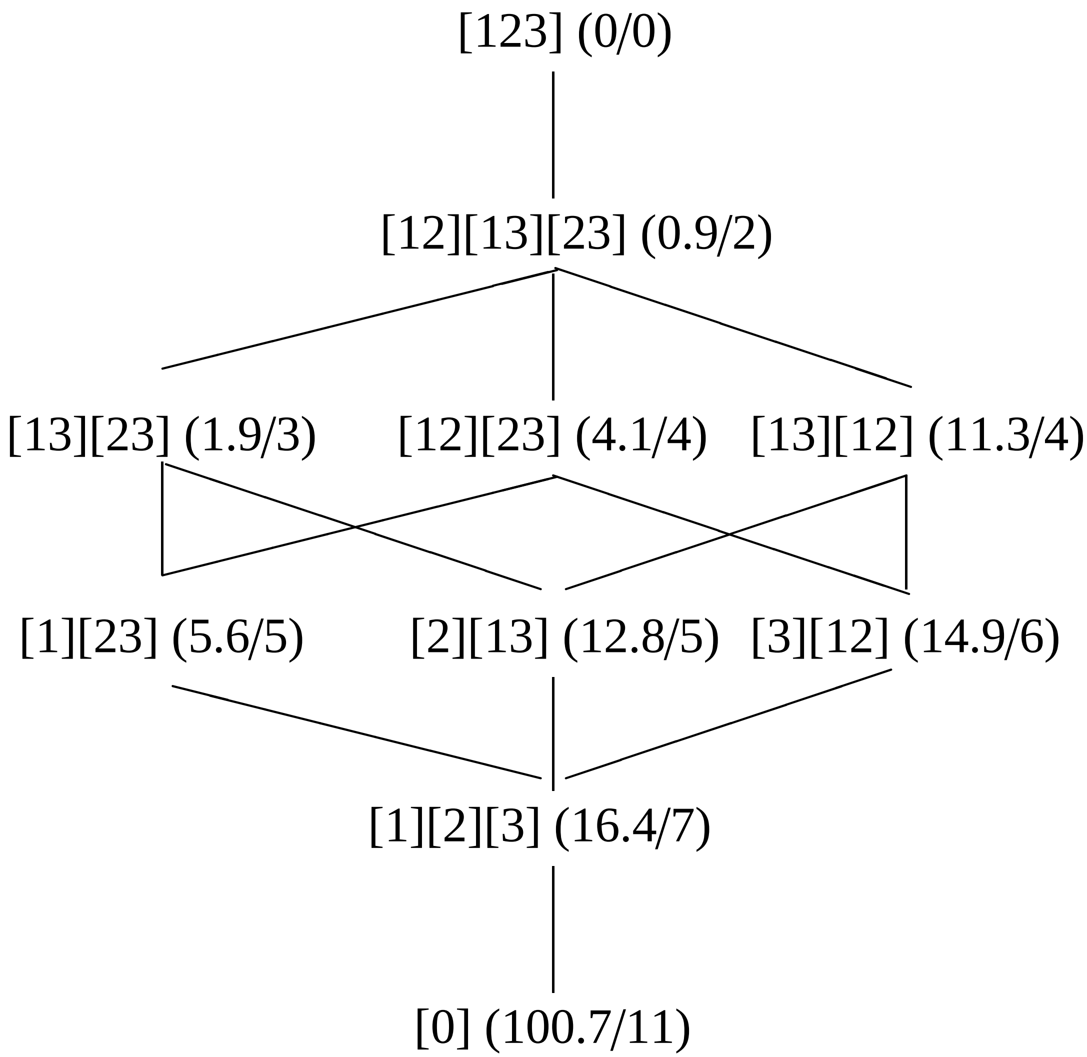
  This is how R's linear modeling notation works also: `Schl + Risk*Beh` expands to `1 + Schl + Risk + Beh + Risk:Beh`.
- The generators are useful mnemonics for
  - What the (conditional) independence model is, e.g. $[12][13] \equiv 2 \perp\!\!\!\perp 3 \mid 1$;
  - What the sufficient statistics are, e.g. for [12][13] the sufficient statistics are $n_{12(ij)}$ and $n_{13(ik)}$.

These ten hierarchical models can be organized by nesting as follows:



Example: Residual Deviances (LR vs [123]) in boys' deviance data

Take 1=Beh, 2=Risk, 3=Schl:



Higher-Dimensional Tables

Christensen provides data on the relationship between two drugs ($k = 1, 2$) and muscle tensions ($\ell = 1, 2$) for two weights ($i = 1, 2$) and types ($j = 1, 2$) of muscles in mice.

			Drug (k)	
Tension (ℓ)	Weight (i)	Muscle (j)	Drug 1	Drug 2
High	High	Type 1	3	21
		Type 2	23	11
	Low	Type 1	22	32
		Type 2	4	12
Low	High	Type 1	3	10
		Type 2	41	21
	Low	Type 1	45	23
		Type 2	6	22

Which log-linear models best describe this data?

It is already painful to write down the saturated model [1234] in hierarchical log-linear form:

$$\begin{aligned}\log m_{ijk} &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(\ell)} \\ &+ u_{12(ij)} + u_{13(ik)} + u_{14(i\ell)} + u_{23(jk)} + u_{24(j\ell)} + u_{34(j\ell)} \\ &+ u_{123(ijk)} + u_{124(ij\ell)} + u_{134(ik\ell)} + u_{234(jk\ell)} \\ &+ u_{1234(ijk\ell)}\end{aligned}$$

However, the "generator notation" makes model specification easier:

- The independence model [1][2][3][4] consists of the first line above.
- The model of no three-way interaction [12][13][14][23][24][34] consists of the first two lines above.
- The model of no four-way interaction [123][124][134][234] consists of the first three lines above.
- The model [123][4] (or [WMD][T]) speficies that muscle weight, type and drug are independent of tension.
- The model [TMD][WMD] specifies that (tension)$\perp\!\!\!\perp$(weight) | (type,drug).

As an illustration we fit three models

```
n <- scan() ...
Wt <- rep(rep(c("H","L"),c(4,4)),2) # [W]eight
Mt <- rep(rep(c("1","2"),c(2,2)),4) # [M]uscle type
Dt <- rep(c("1","2"),8)             # [D]rug type
Tn <- rep(c("H","L"),c(8,8))        # [T]ension
musc <- data.frame(n,Wt,Mt,Dt,Tn)
```

- Independence: [T][W][M][D]
  `glm(n ~ Tn + Wt + Mt + Dt, data = musc, family = poisson)`
- No three-way interaction: [TW][TM][TD][WM][WD][MD]
  `glm(n~Tn*Wt + Tn*Mt + Tn*Dt + Wt*Mt + Wt*Dt + Mt*Dt, ...)`
- No four-way interaction: [TWM][TWD][TMD][WMD]
  `glm(n ~ Tn*Wt*Mt*Dt - Tn:Wt:Mt:Dt, ...)`

Model	Resid. Df	Resid. Dev	$P[\chi^2_{df} > \text{Dev}]$
Independence	11	127.351	0.00
No three-way	5	47.669	4.15e-09
No four-way	1	0.111	0.74

Digression: Why Poisson If We Believe Multinomial?

Maximum likelihood for the Poisson model

We re-index a table such as n_{ijk} to be just n_c, and we suppose that n_c, $c = 1, \ldots, C$ are independent Poisson counts with means m_c. Then

$$\begin{aligned} f(n_c|m_c) &= \frac{e^{-m_c} {m_c}^{n_c}}{n_c!} \\ &= \exp[n_c \log m_c - m_c - \log n_c!] \\ &= \exp[n_c \theta_c - \exp \theta_c - \log n_c!] \end{aligned}$$

which is an exponential family model with natural parameter $\theta_c = \log m_c$. If we model $\theta_c = \log m_c$ linearly as

$$\theta_{C\times 1} = [\log m_c]_{C\times 1} = X_{C\times D}\beta_{D\times 1}$$

then the above is a *generalized linear model* for Poisson counts with the log link function $\theta_c = \log \mu_c$.

The log-likelihood may be written

$$
\begin{aligned}
L(n|\beta) = \prod_c f(n_c|\beta) &= \exp\left[\sum_c (n_c \log m_c - m_c) - \sum_c \log n_c!\right] \\
&= \exp\left[n^T X\beta - \sum_c \exp([X\beta]_c) + g(data)\right]
\end{aligned}
$$

and, either explicitly setting derivatives equal to zero, or by using the general theory of glm's, we see that the likelihood equations reduce to

$$\sum_c (n_c - m_c)x_{cd} = 0\ , \quad d = 1, \ldots, D$$

This is the usual result that the MLE in an exponential family model equates observed and expected sufficient statistics

$$n^T X = m^T X$$

In log-linear models for tables, these are invariably appropriate marginal totals for the table (see example, next slide).

Sub-digression: Where does X come from?

Consider the table n_{ijk} with $i = 1, 2,\ j = 1, 2$, and $k = 1, 2$. If we lay out the log expected cell counts in a column, the model of independence $\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)}$ looks like this:

$$\theta = \begin{bmatrix} \log m_{111} \\ \log m_{112} \\ \log m_{121} \\ \log m_{122} \\ \log m_{211} \\ \log m_{212} \\ \log m_{221} \\ \log m_{222} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ u_{1(2)} \\ u_{2(2)} \\ u_{3(2)} \end{bmatrix} = X\beta$$

Clearly the observed sufficient statistics for the u parameters are

$$n^T X = (n_{+++}, n_{2++}, n_{+2+}, n_{++2})$$

which are equated by ML to the expected sufficient statistics

$$m^T X = (m_{+++}, m_{2++}, m_{+2+}, m_{++2})$$

Back to the Poisson Likelihood

- Further differentiation (or application of general results from glm's) shows that the information matrix for β is

$$\left[-\frac{\partial^2}{\partial\beta_{d_1}\partial\beta_{d_2}}\right]_{D\times D} = X^T \hat{W} X$$

where $W = diag(\hat{m})$ and hence

$$\widehat{\text{Var}}\,(\beta) = [X^T \hat{W} X]^{-1}$$

- Since the MLE's under the saturated model (no relations among the m_i's) are $\hat{m}_i = n_i$, the log-LR statistic for testing against the saturated model is

$$\begin{aligned} -2[\log L(n|\beta) - \log L(n|\text{saturated})] &= 2\sum_c [n_c \log(n_c/\hat{m}_c) - n_c + \hat{m}_c] \\ &= 2\sum_c n_c \log(n_c/\hat{m}_c) \end{aligned}$$

since $n_+ = \hat{m}_+$ as long as the log-linear model has the intercept u in it.

- Since the individual cells counts $n_i \sim Poiss(m_i)$, with $E[n_i] = \text{Var}\,(n_i) = m_i$, it follows that the Pearson residuals

$$r_i = (n_i - \hat{m}_i)/\sqrt{\hat{m}_i}$$

are approximately mean 0, variance 1 (and should be approximately normal for large n_i). This is why they are sensible residuals to use.

- Recall that the model deviance (LR statistic for testing against the saturated model) is

$$G^2 = 2\sum_i n_i \log(n_i/\hat{m}_i) = \sum_i d_i$$

The *deviance residuals* are

$$r_i^D = \text{sgn}\,(n_i - \hat{m}_i) \cdot \sqrt{|d_i|}$$

and these are approximately normally distributed (since they are signed square roots of approximately 1-df contributions to G^2).

Maximum likelihood for the Multinomial model

We again take a table like n_{ijk} and re-index the cells as $n_1, \ldots, n_C$. *The key observation here is that if we condition on* $n_+ = \sum_c n_c$, *C independent Poisson's become a C-cell multinomial.*

Begin with a Poisson log-linear model with an intercept α,

$$\log m_c = \alpha + x_c^T \beta \qquad (= u + [\text{other } u\text{-terms}])$$

where $(1, x_c^T)$ is the c^{th} row of X. Up to a function of the data $g(data) = \sum_c \log n_c!$, the Poisson log-likelihood is

$$\begin{aligned}
\log L(n|\alpha, \beta) &= \sum_c n_c \log m_c - \sum_c m_c \\
&= \sum_c n_c(\alpha + x_c^T \beta) - \sum_c \exp(\alpha + x_c^T \beta) \\
&= n_+ \alpha + \sum_c n_c x_c^T \beta - \tau
\end{aligned}$$

where $\tau = \sum_c m_c = \sum_c \exp(\alpha + x_c^T \beta) = e^\alpha \sum_c \exp(x_c^T \beta)$.

Now since $\log \tau = \alpha + \log \sum_c \exp(x_c^T \beta)$, it follows that

$$\begin{aligned}
\log L(n|\alpha,\beta) &= \log L(n|\tau,\beta) \\
&= \left\{ \sum_c n_c x_c^T \beta - n \log \sum_c \exp(x_c^T \beta) \right\} + [n_+ \log \tau - \tau] \\
&= \left\{ \sum_c n_c x_c^T \beta - \sum_c n_c \log \sum_{c'} \exp(x_{c'}^T \beta) \right\} + [n_+ \log \tau - \tau] \\
&= \left\{ \sum_c n_c \log p_c \right\} + [n_+ \log \tau - \tau]
\end{aligned}$$

where $p_c = m_c/m_+ = \exp(\alpha + x_c^T \beta)/ \sum_{c'} \exp(\alpha + x_{c'}^T \beta) =$ $\exp(x_c^T \beta)/ \sum_{c'} \exp(x_{c'}^T \beta)$. The term in brackets is the multinomial log-likelihood, up to a function of the data $h(data) = \log n_+! - \sum_c \log n_c!$.

Thus, as log-likelihoods:

$$\text{Poisson}_{\alpha,\beta}(n) = \{ \text{ Multinomial}_{\beta}(n) \ | \ n_+ \} + [\text{ Poisson}_{\tau}(n_+) \ ].$$

That is worth repeating: For the log-linear model $\log m_c = \alpha + x_c^T \beta$,

$$\text{Poisson}_{\alpha,\beta}(n) = \{ \text{ Multinomial}_\beta(n) \ \mid \ n_+ \} + [\text{ Poisson}_\tau(n_+) \ ].$$

as log-likelihoods. Therefore:

- There is a **1-1 correspondence** between *Multinomial log-linear models* and *Poisson log-linear models* with the same log-linear form. The Poisson model requires one more parameter (τ), corresponding to the grand total n_+.
- The (non-intercept) parameters β in the two models have **the same MLE's** and **the same variance-covariance matrix** (since the β's only enter in the bracketed expressions on the previous slide).
- The **intercept** parameter α **will be different** in the two models.
- The **Pearson residuals** are still variance-stabilized, and as long as the grand total n_+ is large relative to the cell count n_i, each r_i will be **roughly** $\boldsymbol{N(0,1)}$.

You can also show that the computation and interpretation of G^2 and X^2 also do not change between the models.

Product Multinomial Log-linear Models

A similar argument can be used to derive product-multinomial MLE's from Poisson MLE's.

- The key to the equivalence between Poisson MLE's $\hat{\beta}$ and single-Multinomial MLE's $\hat{\beta}$ was to *include the intercept α corresponding to the fixed grand total n_+* in the log-linear model: $\log m_c = \alpha + x_c^T \beta$. Then as log-likelihoods

$$\text{Poisson}_{\alpha,\beta}(n) = \{\ \text{Multinomial}_\beta(n)\ |\ n_+\ \} + [\ \text{Poisson}_\tau(n_+)\ ].$$

- The key to equivalence between Poisson MLE's and Product Multinomial MLE's will be to *include log-linear terms $\alpha^{(1)}, \ldots, \alpha^{(H)}$ corresponding to the fixed margins $n_+^{(1)}, \ldots, n_+^{(H)}$*. Then as log-likelihoods

$$\text{Poisson}_{\alpha,\beta}(n) = \sum_{h=1}^{H} \{\ \text{Multinomial}_\beta(n^{(h)})\ |\ n_+^{(h)}\ \} + [\ \text{Poisson}_\tau(n_+^{(h)})\ ].$$

These results have been known for some time (e.g. Birch, 1963, *JRSSB*); a recent update/generalization is Lang (1996, *JRSSB*).

Example

Consider again the Aspirin and Heart Attack data:

	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	Total
Placebo	$n_{11} = 18$	$n_{12} = 171$	$n_{1+} = 189$
Aspirin	$n_{21} = 5$	$n_{22} = 99$	$n_{2+} = 104$
Total	$n_{+1} = 23$	$n_{+2} = 270$	$n_{++} = 293$

- If we consider this to be observational data with only n_{++} fixed and a multinomial model for $n_{ij}|n_{++}$, we can use `glm(..., family=poisson)` to fit log-linear models to this data *as long as we include the intercept u in each log-linear model*.

- If we consider this to be a designed experiment so that the cells are product multinomial with fixed row totals n_{1+} and n_{2+}, then we can use `glm(..., family=poisson)` to fit log-linear models to this data *as long as we include the $u_{1(i)}$ terms in each log-linear model*.

- Testing independence with the Single Multinomial Model:

```
n <- scan()
18 171 10845
5 99 10993

Tx <- rep(c("Placebo","Aspirin"),c(3,3))
Obs <- rep(c("Fatal","NonFatal","NoAttack"),2)
aha.data <- data.frame(n,Tx,Obs)

print(fit <- glm(n ~ Tx + Obs,data=aha.data,family=poisson))
```

  $G^2 = 28.058$ on 2 d.f.; $p \approx 8 \times 10^{-8}$, so again we reject independence.

- Testing independence with the Product Multinomial Model:

  Since the model of independence already has log-linear terms (`Tx`, i.e. $u_{1(i)}$) corresponding to the fixed row totals n_{i+}, the same Poisson fit above also gives the results for the Product Multinomial.

In general, when H_0 already contains terms corresponding to the fixed margins in the table, testing H_0 is <u>identical</u> under the Poisson, Multinomial, and Product Multinomial sampling models.

Example

Back to the muscle study in mice:

Tension (ℓ)	Weight (i)	Muscle (j)	Drug (k) Drug 1	Drug 2
High	High	Type 1	3	21
		Type 2	23	11
	Low	Type 1	22	32
		Type 2	4	12
Low	High	Type 1	3	10
		Type 2	41	21
	Low	Type 1	45	23
		Type 2	6	22

We treated this data before as Poisson or single multinomial. In fact, it was a designed study with the total number of muscles of each *type* fixed in advance. *So every log-linear model should have the* `M` *terms* ($u_{2(j)}$) *in it.*

An advantage of *hierarchical principle* is that the interesting models automatically contain the main effects, so if the totals in one dimension are fixed, the sampling scheme “doesn’t matter” for model fit/comparison:

Model	Resid. Df	Resid. Dev	P[χ^2_{df} > Dev]
[T][W][M][D]	11	127.351	0.00
[TW][TM][TD][WM][WD][MD]	5	47.669	4.15e-09
[TWM][TWD][TMD][WMD]	1	0.111	0.74

In the study, [T]ension and [W]eight of muscle are measured on each combination of [M]uscle type and [D]rug. Two more models of interest might be

1. [TMD][WMD]: (tension)$\perp\!\!\!\perp$(weight) | (muscle type,drug)
2. [TWM][DM]: (tension,weight)$\perp\!\!\!\perp$(drug) | (muscle type)

- In the real study, only the totals for muscle [T]ype were fixed, so either of these models could be fitted as well.

- In another study of this type, perhaps muscle [W]eight and [D]rug type would be fixed in advance.
  - In that case all models would need to include the [WD] interaction: we could fit and interpret [TMD][WMD] but not [TWM][DM]
  - (without the [WD] interaction, the totals for (weight)$\times$(drug) combinations would not be fixed by the log-linear model; the product multinomial model with [WD] margins fixed could not be represented)