Log-Linear Models

Ronald Christensen
Department of Mathematics and Statistics
University of New Mexico
Albuquerque, NM 87131
e-mail: fletcher@stat.unm.edu

1. Introduction

1.1 Odds and Odds Ratios

Example *1990 Superbowl: Harrah's Odds*

Team	Odds	Odds	p
San Francisco Forty-Niners	even (1 to 1)	1	.50
Denver Broncos	5 to 2	2.5	.29
New York Giants	3 to 1	3	.25
Cleveland Browns	9 to 2	4.5	.18
Los Angeles Rams	5 to 1	5	.17
Minnesota Vikings	6 to 1	6	.14
Buffalo Bills	8 to 1	8	.11
Pittsburgh Steelers	10 to 1	10	.09

$$Odds = \frac{p}{1-p} = \frac{\Pr(\text{Event Occurs})}{\Pr(\text{Event Does Not Occur})}.$$

$$p = \frac{O}{O+1}.$$

$$\Pr(\text{Vikings do not win}) = \frac{6}{6+1} = \frac{6}{7}$$

$$\Pr(\text{Vikings win}) = \frac{\frac{1}{6}}{\frac{1}{6}+1} = \frac{1}{1+6} = \frac{1}{7}.$$

$$\Pr(\text{Broncos do not win}) = \frac{\frac{5}{2}}{\frac{5}{2}+1} = \frac{5}{5+2} = \frac{5}{7}$$

$$\Pr(\text{Broncos win}) = \frac{\frac{2}{5}}{\frac{2}{5}+1} = \frac{2}{5+2} = \frac{2}{7}.$$

Team	Odds	Odds	p
San Francisco Forty-Niners	even (1 to 1)	1	.50
Denver Broncos	5 to 2	2.5	.29
New York Giants	3 to 1	3	.25
Cleveland Browns	9 to 2	4.5	.18
Los Angeles Rams	5 to 1	5	.17
Minnesota Vikings	6 to 1	6	.14
Buffalo Bills	8 to 1	8	.11
Pittsburgh Steelers	10 to 1	10	.09

$$.50 + .29 + .25 + .18 + .17 + .14 + .11 + .09 = 1.73 \neq 1.$$

Odds Ratios

L.A. odds of not winning 5 times San Francisco's.
Pittsburgh's 10 times larger than San Francisco's.
Ratio L.A. odds not winning to the SF odds not winning is $5/1 = 5$.
Ratio odds of Pittsburgh not winning to San Francisco not winning is $10/1 = 10$.
Pittsburgh odds of not winning are twice L.A.'s: $10/5 = 2$.
ALSO, this is odds of L.A. winning to the odds of Pittsburgh winning.
Odds L.A. doesn't win: 5 to 1
Odds L.A. wins: 1 to 5 or $1/5 = .2$.
Odds Pittsburgh wins $1/10 = .1$.
L.A. odds of winning twice those of Pittsburgh.

1.2 Independence

DEFINITION Two events are independent if knowing whether the first event occurred does not change your probability that the second event will occur.

EXAMPLE Flipping coins, tossing die, playing roulette.

EXAMPLE Eye color and intelligence?

Example Economic status (High, Low), residence (Montana, Haiti), and beverage of preference (Beer, Other).

	Beer		Other		Total
	Montana	Haiti	Montana	Haiti	
High	.021	.009	.049	.021	.1
Low	.189	.081	.441	.189	.9
Total	.210	.090	.490	.210	1.0

Factors are completely independent.

$$\begin{aligned} \Pr(\text{High}|\text{Montana, Beer}) &= .021/.210 \\ &= .1 \\ &= \Pr(\text{High})\,. \end{aligned}$$

$$\begin{aligned} \Pr(\text{High, Montana, Beer}) &= .021 = .1(.210) \\ &= \Pr(\text{High}) \Pr(\text{Montana, Beer})\,. \end{aligned}$$

$$\Pr(\text{Montana}) = .210 + .490 = .7$$

$$\Pr(\text{Beer}) = .210 + .090 = .3$$

$$\begin{aligned} \Pr(\text{Low, Montana, Beer}) &= .189 = (.9)(.7)(.3) \\ &= \Pr(\text{Low}) \Pr(\text{Montana}) \Pr(\text{Beer}) \end{aligned}$$

Example

	Democrat		Republican		Total
	Liberal	Conservative	Liberal	Conservative	
High	.12	.12	.04	.12	.4
Low	.18	.18	.06	.18	.6
Total	.30	.30	.10	.30	1.0

Status independent of philosophy and affiliation.

$$
\begin{aligned}
\Pr(\text{High}, \text{Liberal}, \text{Republican}) &= .04 \\
&= (.4)(.1) \\
&= \Pr(\text{High}) \Pr(\text{Liberal}, \text{Republican}) \, .
\end{aligned}
$$

Other divisions do not work.

$$
\begin{aligned}
\Pr(\text{High}, \text{Liberal}, \text{Republican}) &= .04 \\
&\neq (.4)(.16) \\
&= \Pr(\text{Liberal}) \Pr(\text{High}, \text{Republican}) \, .
\end{aligned}
$$

$$
\begin{aligned}
\Pr(\text{High}, \text{Liberal}, \text{Republican}) &= .04 \\
&\neq (.4)(.16) \\
&= \Pr(\text{Republican}) \Pr(\text{High}, \text{Liberal}) \, .
\end{aligned}
$$

EXAMPLE (Subject matter from real study.)

Factor	Levels
Attitude on Extramarital Coitus	Always Wrong, Not Always Wrong
Virginity	Virgin, Nonvirgin
Use of Contraceptives	Regular, Intermittent, None

	Use of Contraceptives	
	Regular	
	Virgin	Nonvirgin
Always Wrong	3/50	12/50
Not Always	3/50	12/50
	Intermittent	
	Virgin	Nonvirgin
Always Wrong	1/80	2/80
Not Always	3/80	2/80
	None	
	Virgin	Nonvirgin
Always Wrong	3/40	1/40
Not Always	6/40	2/40

Attitude and virginity independent given regular use.

$$\begin{aligned}\Pr(\text{Regular}) &= \frac{3}{50}+\frac{3}{50}+\frac{12}{50}+\frac{12}{50}\\ &= 30/50\,.\end{aligned}$$

$$\Pr(\text{Always Wrong, Virgin}|\text{Regular}) = (3/50)/(30/50) = .1$$

Conditional Probabilities Given
Regular Use of Contraceptives

	Virgin	Nonvirgin	Total
Always Wrong	.1	.4	.5
Not Always	.1	.4	.5
Total	.2	.8	1.0

Attitude and virginity independent given regular use.

Conditional Probabilities Given
No Use of Contraceptives

	Virgin	Nonvirgin	Total
Always Wrong	3/12	1/12	1/3
Not Always	6/12	2/12	2/3
Total	3/4	1/4	1

Attitude and virginity independent given no use.

Conditional Probabilities Given
Intermittent Use of Contraceptives

	Virgin	Nonvirgin	Total
Always Wrong	1/8	1/4	3/8
Not Always	3/8	1/4	5/8
Total	1/2	1/2	1

Virgins three times as likely to think not always wrong, nonvirgins evenly split.

1.3 Multinomial Sampling

	Democrat		Republican	
	Liberal	Conservative	Liberal	Conservative
High	.12	.12	.04	.12
Low	.18	.18	.06	.18

One sample of size $N = 50$ and these p_{ijk}'s.

	Democrat		Republican	
	Liberal	Conservative	Liberal	Conservative
High	5	7	4	6
Low	8	7	3	10

Expected numbers of observations for each category: Np_{ijk}.

	Democrat		Republican	
	Liberal	Conservative	Liberal	Conservative
High	6	6	2	6
Low	9	9	3	9

1.4 Product Multinomial Sampling

	Democrat		Republican		
	Liberal	Conservative	Liberal	Conservative	Total
High	.3	.3	.1	.3	1.0
Low	.3	.3	.1	.3	1.0

Samples of $N_1 = 30$ from high group; $N_2 = 20$ from low group

	Democrat		Republican		
	Liberal	Conservative	Liberal	Conservative	Total
High	10	10	2	8	30
Low	5	8	1	6	20

Expected counts: $N_i p_{ijk}$'s

	Democrat		Republican		
	Liberal	Conservative	Liberal	Conservative	Total
High	9	9	3	9	30
Low	6	6	2	6	20

2. Two-Way Tables

Graduate admissions at Berkeley

	Admitted	Rejected	Total
Females	557	1278	1835
Males	1198	1493	2691

OBSERVED VALUES

		Columns		
		1	2	Totals
Rows	1	n_{11}	n_{12}	$n_{1\cdot}$
	2	n_{21}	n_{22}	$n_{2\cdot}$
	Totals	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

PROBABILITIES

		Columns		
		1	2	Totals
Rows	1	p_{11}	p_{12}	$p_{1\cdot}$
	2	p_{21}	p_{22}	$p_{2\cdot}$
	Totals	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot\cdot}$

EXPECTED VALUES

		Columns		
		1	2	Totals
Rows	1	m_{11}	m_{12}	$m_{1\cdot}$
	2	m_{21}	m_{22}	$m_{2\cdot}$
	Totals	$m_{\cdot 1}$	$m_{\cdot 2}$	$m_{\cdot\cdot}$

2.1 Product Multinomial Sampling: Testing Homogeneity

Sample sizes $n_{i\cdot}$
Expected counts $m_{ij} = n_{i\cdot} p_{ij}$
H_0: rows have same probabilities $p_{1j} = \cdots p_{Ij}$, all j
With rows having the same probabilities, pool across rows to estimate probabilities

$$\hat{p}_{ij} = n_{\cdot j}/n_{\cdot\cdot}$$

$$\hat{m}_{ij}^{(0)} = n_{i\cdot}\hat{p}_{ij} = n_{i\cdot}(n_{\cdot j}/n_{\cdot\cdot})$$

Pearson chi-square statistic

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}} .$$

If $n_{ij} = 4$ and $\hat{m}_{ij}^{(0)} = 1$, the prediction is poor.
If $n_{ij} = 104$ and $\hat{m}_{ij}^{(0)} = 101$, the prediction is good.
$\hat{m}_{ij}^{(0)}$ in denominator corrects for problem.
Reject H_0 at $\alpha = .05$ if

$$X^2 \geq \chi^2(.95, 1) .$$

Graduate admissions at Berkeley

$\hat{m}_{ij}^{(0)}$	Admitted	Rejected	Total
Females	711.54	1123.46	1835
Males	1043.46	1647.54	2691

$$X^2 = 92.205$$
$$\chi^2(.999, 1) = 10.828$$

Pearson residuals

$$\tilde{r}_{ij} = \frac{n_{ij} - \hat{m}_{ij}^{(0)}}{\sqrt{\hat{m}_{ij}^{(0)}}}$$

Note $X^2 = \Sigma_{ij}\ \tilde{r}_{ij}^2$.

Graduate admissions at Berkeley

$\tilde{r}_{ij}$	Admitted	Rejected
Females	-5.79	4.61
Males	4.78	-3.81

2.2 Multinomial Sampling: Testing Independence

Example

Two factors: personality type and exercise.
Personality type: A (high stress) or B (low stress).
Exercise: regular or not

	n_{ij}	Personality A	B	Totals
Exercise	Regular	483	477	960
	Other	1101	1121	2222
	Totals	1584	1598	3182

Sample size $n_{..}$
Expected counts $m_{ij} = n_{..}p_{ij}$
H_0: Rows and columns are independent $p_{ij} = p_{i.}p_{.j}$
With rows and columns are independent $m_{ij} = n_{..}p_{i.}p_{.j}$
Independence is equivalent to $m_{ij} = m_{i.}m_{.j}/n_{..}$

$$\hat{p}_{i.} = n_{i.}/n_{..}$$

$$\hat{p}_{.j} = n_{.j}/n_{..}$$

$$\hat{m}_{ij}^{(0)} = n_{..}\hat{p}_{i.}\hat{p}_{.j} = n_{i.}n_{.j}/n_{..}$$

Pearson's chi-square

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}}$$

$$\tilde{r}_{ij} = \frac{n_{ij} - \hat{m}_{ij}^{(0)}}{\sqrt{\hat{m}_{ij}^{(0)}}}\,.$$

Personality-exercise data,

	$\hat{m}_{ij}^{(0)}$	Personality A	B	Totals
Exercise	Regular	477.9	482.1	960
	Other	1106.1	1115.9	2222
	Totals	1584	1598	3182

$$X^2 = .156$$

$$P \geq .5$$

No deviations from independence, no need to look at residuals.

EXAMPLE 52 males operated on for knee injuries
Type of injury: twisted knee, direct blow, both.
Results of the surgery: excellent (E), good (G), fair or poor (F-P).

		Result			
	n_{ij}	E	G	F-P	Totals
	Twist	21	11	4	36
Injury	Direct	3	2	2	7
	Both	7	1	1	9
	Totals	31	14	7	52

Estimated expected counts under H_0: $\hat{m}_{ij}^{(0)} = n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot}$

		Result			
	$\hat{m}_{ij}^{(0)}$	E	G	F-P	Totals
	Twist	21.5	9.7	4.8	36
Injury	Direct	4.2	1.9	.9	7
	Both	5.4	2.4	1.2	9
	Totals	31	14	7	52

$$X^2 = 3.229$$

$$df = (3-1)(3-1) = 4\,.$$

$$P = .52$$

No need to look at residuals.

2.3 Log-Linear Models for Two-Way Tables

Two-way ANOVA, unrestricted model

$$y_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} + e_{ijk}$$

$$y_{ijk} \sim N(m_{ij}, \sigma^2)$$

$$m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

Examine the m_{ij}'s: MLE's of the m_{ij}'s $\hat{m}_{ij} = \bar{y}_{ij\cdot}$

No interaction model

$$m_{ij} = u + u_{1(i)} + u_{2(j)}$$

MLE's change with model.

$$\hat{m}_{ij} = \bar{y}_{\cdots} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})$$

$I \times J$ table of counts use similar techniques. The table entries have

$$\mathrm{E}(n_{ij}) = m_{ij} \, .$$

Examine structure of m_{ij}'s: log-linear models.

Saturated (unrestricted) model

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

No interactions model

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \, .$$

Theorem: For multinomial sampling in an $I \times J$ table,

$$\log(m_{ij}) = u + u_{1(i)} + u_{2(j)} \quad \text{all } i, j$$

if and only if

$$p_{ij} = p_{i\cdot}p_{\cdot j} \quad \text{all } i, j$$

if and only if

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1 \quad \text{all } i, i', j, j'$$

if and only if

$$p_{11}p_{ij}/p_{1j}p_{i1} = 1 \quad \text{all } i \neq 1, j \neq 1$$

Theorem: For product multinomial sampling with rows as independent samples,

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} \quad \text{all } i, j$$

if and only if

$$p_{1j} = \cdots = p_{Ij}, \quad j = 1, \ldots, J$$

if and only if

$$\frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} = 1 \quad \text{all } i,\ i'\ j,\ j'$$

if and only if

$$p_{11}p_{ij}/p_{1j}p_{i1} = 1 \quad \text{all } i \neq 1, j \neq 1$$

Test of homogeneity or independence = the test of no interaction model against the more general saturated model.

$\hat{m}_{ij} = n_{ij}$: MLE's in unrestricted (saturated) model.

$\hat{m}_{ij}^{(0)} = n_{i\cdot}n_{\cdot j}/n_{\cdot\cdot}$: MLE's in homogeneity or independence models = MLE's in no interaction models.

Pearson test statistic:

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(n_{ij} - \hat{m}_{ij}^{(0)}\right)^2}{\hat{m}_{ij}^{(0)}}$$

Likelihood ratio test statistic (deviance):

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log\left(\frac{n_{ij}}{\hat{m}_{ij}^{(0)}}\right)$$

$$df = (I-1)(J-1)$$

Example Knee operations: $G^2 = 3.173$, $X^2 = 3.229$, $df = 4$.

EXAMPLE

Swedish births: monthly observations (n_{ij}s)
and monthly proportions by sex

	Observations			Proportions	
Month	Female	Male	Total	Female	Male
January	3537	3743	7280	.083	.082
February	3407	3550	6957	.080	.078
March	3866	4017	7883	.091	.088
April	3711	4173	7884	.087	.091
May	3775	4117	7892	.089	.090
June	3665	3944	7609	.086	.086
July	3621	3964	7585	.085	.087
August	3596	3797	7393	.084	.083
September	3491	3712	7203	.082	.081
October	3391	3512	6903	.080	.077
November	3160	3392	6552	.074	.074
December	3371	3761	7132	.079	.082
Total	42591	45682	88273	1.000	1.000

Estimated expected Swedish births by month ($\hat{m}_{ij}$s) and pooled proportions

	Expectations			Pooled
Month	Female	Male	Total	proportions
January	3512.54	3767.46	7280	.082
February	3356.70	3600.30	6957	.079
March	3803.48	4079.52	7883	.089
April	3803.97	4080.03	7884	.089
May	3807.83	4084.17	7892	.089
June	3671.28	3937.72	7609	.086
July	3659.70	3925.30	7585	.086
August	3567.06	3825.94	7393	.084
September	3475.39	3727.61	7203	.082
October	3330.64	3572.36	6903	.078
November	3161.29	3390.71	6552	.074
December	3441.13	3690.87	7132	.081
Total	42591.00	45682.00	88273	1.000

$$G^2 = 14.988 \quad X^2 = 14.986$$

$$df = (12-1)(2-1) = 11$$

$$\chi^2(.90, 11) = 17.28$$

Pearson residuals for
Swedish birth months, ($\tilde{r}_{ij}$s)

Month	Female	Male
January	0.41271	−0.39849
February	0.86826	−0.83837
March	1.01369	−0.97880
April	−1.50731	1.45542
May	−0.53195	0.51364
June	−0.10365	0.10008
July	−0.63972	0.61770
August	0.48452	−0.46785
September	0.26481	−0.25570
October	1.04587	−1.00987
November	−0.02288	0.02209
December	−1.19554	1.15438

Example Occupations: A, professions; B, owners, managers, and officials; C, clerical and sales; and D, skilled.

Religion and occupations

	Occupation				
Religion	A	B	C	D	Total
Protestant	210	277	254	394	1135
Roman Catholic	102	140	127	279	648
Jewish	36	60	30	17	143
Total	348	477	411	690	1926

Estimated expected counts ($\hat{m}_{ij}$s)

Religion	A	B	C	D	Total
Protestant	205.08	281.10	242.20	406.62	1135
Roman Catholic	117.08	160.49	138.28	232.15	648
Jewish	25.84	35.42	30.52	51.23	143
Total	348.00	477.00	411.00	690.00	1926

$$G^2 = 64.3 \quad X^2 = 60.0$$

$$df = (3-1)(4-1) = 6$$

$$\chi^2(.995, 6) = 18.55$$

Residuals ($\tilde{r}_{ij}$s)

Religion	A	B	C	D
Protestant	0.34	−0.24	0.76	−0.63
Roman Catholic	−1.39	−1.62	−0.96	3.07
Jewish	2.00	4.13	−0.09	−4.78

Observed proportions by religion

	Occupation				
Religion	A	B	C	D	Total
Protestant	.185	.244	.224	.347	1.00
Roman Catholic	.157	.216	.196	.431	1.00
Jewish	.252	.420	.210	.119	1.00

A Lancaster–Irwin partition

Reduced table

Religion	A	B	C	D	Total
Protestant	210	277	254	394	1135
Roman Catholic	102	140	127	279	648
Total	312	417	381	673	1783

Collapsed table

Religion	A	B	C	D	Total
Prot. & R.C.	312	417	381	673	1783
Jewish	36	60	30	17	143
Total	348	477	411	690	1926

2.4 Simple Logistic Regression

O-ring Failure Data

Case	Failure	Success	Temperature
1	1	0	53
2	1	0	57
3	1	0	58
4	1	0	63
5	0	1	66
6	0	1	67
7	0	1	67
8	0	1	67
9	0	1	68
10	0	1	69
11	0	1	70
12	0	1	70
13	1	0	70
14	1	0	70
15	0	1	72
16	0	1	73
17	0	1	75
18	1	0	75
19	0	1	76
20	0	1	76
21	0	1	78
22	0	1	79
23	0	1	81

Logistic Regression Model:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \tau$$

$$p = \frac{\exp(\beta_0 + \beta_1 \tau)}{1 + \exp(\beta_0 + \beta_1 \tau)}.$$

Variable	Par.	Estimate	Std. Error	z
Intercept	β_0	15.04	7.316	2.06
Temperature	β_1	-0.2321	0.1073	-2.16

$$\hat{p} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 \tau)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 \tau)}.$$

Product Multinomial Sampling, sample sizes small, X^2 and G^2 tests unreliable.

Log-Linear Model:

$$\log(m_{ij}) = u_{1(i)} + u_{2(j)} + \eta_j \tau_i$$

$$\begin{aligned}
\log\left(\frac{p_i}{1-p_i}\right) &= \log\left(\frac{m_{i1}}{m_{i2}}\right) \\
&= \log(m_{i1}) - \log(m_{i2}) \\
&= \left[u_{1(i)} + u_{2(1)} + \eta_1 \tau_i\right] - \left[u_{1(i)} + u_{2(2)} + \eta_2 \tau_i\right] \\
&= \left[u_{2(1)} - u_{2(2)}\right] + [\eta_1 \tau_i - \eta_2 \tau_i] \\
&\equiv \beta_0 + \beta_1 \tau_i
\end{aligned}$$

where $\beta_0 \equiv \left[u_{2(1)} - u_{2(2)}\right]$ and $\beta_1 \equiv [\eta_1 - \eta_2]$.

2.4.1 SAS Commands

```
options ps=60 ls=72 nodate;
data oring;
   infile 'oring.dat';
   input ID f s temp;
   n = 1;
proc genmod data = oring;
   model f/n = temp  / link=logit
                       dist=binomial;
run;
```

3. Three Dimensional Tables

		Adversity of School (k)						
		Low		Medium		High		
	Risk (j)	N	R	N	R	N	R	Total
Classroom	Nondeviant	16	7	15	34	5	3	80
Behavior (i)	Deviant	1	1	3	8	1	3	17
	Total	17	8	18	42	6	6	97

3.1 Simpson's Paradox

EXAMPLE

		Patient Sex			
		Male		Female	
Outcome		Success	Failure	Success	Failure
Treatment	1	60	20	40	80
	2	100	50	10	30

		Outcome	
		Success	Failure
Treatment	1	100	100
	2	110	80

EXAMPLE Berkeley Admissions Data.

3.2 Independence Models

3.2.1 Complete Independence

$$M^{(0)} : p_{ijk} = p_{i\cdot\cdot} p_{\cdot j \cdot} p_{\cdot\cdot k} \quad \text{or} \quad m_{ijk} = \frac{m_{i\cdot\cdot} m_{\cdot j \cdot} m_{\cdot\cdot k}}{n_{\cdots}^2}$$

MLE's

$$\hat{p}_{ijk} = \frac{n_{i\cdot\cdot}}{n_{\cdots}} \frac{n_{\cdot j \cdot}}{n_{\cdots}} \frac{n_{\cdot\cdot k}}{n_{\cdots}}$$

$$\hat{m}_{ijk}^{(0)} = \frac{n_{i\cdot\cdot} n_{\cdot j \cdot} n_{\cdot\cdot k}}{n_{\cdots}^2}$$

n_{ijk}		Diastolic Blood Pressure	
Personality	Cholesterol	Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

$\hat{m}_{ijk}^{(0)}$		Diastolic Blood Pressure	
Personality	Cholesterol	Normal	High
A	Normal	739.9	74.07
	High	193.7	19.39
B	Normal	788.2	78.90
	High	206.3	20.65

Note: $\hat{m}_{ijk}^{(0)} = \hat{m}_{i\cdot\cdot} \hat{m}_{\cdot j \cdot} \hat{m}_{\cdot\cdot k} / n_{\cdots}^2$

Note: $n_{i\cdot\cdot} = \hat{m}_{i\cdot\cdot}^{(0)}$, $n_{\cdot j \cdot} = \hat{m}_{\cdot j \cdot}^{(0)}$, and $n_{\cdot\cdot k} = \hat{m}_{\cdot\cdot k}^{(0)}$ for all i, j, and k

Type A totals:

$n_{1\cdot\cdot} = 716 + 79 + 207 + 25 = 1027$

$\hat{m}_{1\cdot\cdot}^{(0)} = 739.9 + 74.07 + 193.7 + 19.39 = 1027.06$ round-off error.

n_{ijk}		Diastolic Blood Pressure	
Personality	Cholesterol	Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

$\hat{m}^{(0)}_{ijk}$		Diastolic Blood Pressure	
Personality	Cholesterol	Normal	High
A	Normal	739.9	74.07
	High	193.7	19.39
B	Normal	788.2	78.90
	High	206.3	20.65

$$X^2 = \frac{(716 - 739.9)^2}{739.9} + \cdots + \frac{(22 - 20.65)^2}{20.65} = 8.730\,.$$

$$G^2 = 2\,[716\log(716/739.9) + \cdots + 22\log(22/20.65)] = 8.723\,.$$

$$\begin{aligned} df &= (2)(2)(2) - 2 - 2 - 2 + 2 = 4 \\ &= IJK - I - J - K + 2 \end{aligned}$$

$$\chi^2(.95, 4) = 9.49 \qquad P = .067$$

$$\tilde{r}_{ijk} = \frac{n_{ijk} - \hat{m}_{ijk}}{\sqrt{\hat{m}_{ijk}}}\,.$$

Pearson residuals

$\tilde{r}_{ijk}$		Diastolic Blood Pressure	
Personality	Cholesterol	Normal	High
A	Normal	-0.879	0.573
	High	0.956	1.274
B	Normal	1.097	-1.340
	High	-1.413	0.297

3.2.2 One factor independent of other two

Three models:
rows independent of columns and layers,

$$M^{(1)}: p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk} \ ,$$

columns independent of rows and layers,

$$M^{(2)}: p_{ijk} = p_{\cdot j\cdot}p_{i\cdot k} \ ,$$

and layers independent of rows and columns,

$$M^{(3)}: p_{ijk} = p_{\cdot\cdot k}p_{ij\cdot}.$$

Complete independence $M^{(0)}$ is a special case.

Examine $M^{(1)}$:

$$p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk} \text{ if and only if } m_{ijk} = m_{i\cdot\cdot}m_{\cdot jk}/n_{\cdot\cdot\cdot}$$

MLE's

$$\hat{p}_{ijk} = \frac{n_{i\cdot\cdot}}{n_{\cdot\cdot\cdot}}\frac{n_{\cdot jk}}{n_{\cdot\cdot\cdot}}$$

$$\hat{m}^{(1)}_{ijk} = \frac{n_{i\cdot\cdot}n_{\cdot jk}}{n_{\cdot\cdot\cdot}}$$

Example

$\hat{m}_{ijk}^{(1)}$		Adversity (k)						
		Low		Medium		High		
	Risk (j)	N	R	N	R	N	R	$\hat{m}_{i\cdot\cdot}$
Classroom	Non.	14.02	6.60	14.85	34.64	4.95	4.95	80
Behavior (i)	Dev.	2.98	1.40	3.15	7.36	1.05	1.05	17
	$\hat{m}_{\cdot jk}$	17	8	18	42	6	6	97

$$X^2 = 6.19 = \frac{(16 - 14.02)^2}{14.02} + \cdots + \frac{(3 - 1.05)^2}{1.05}\,.$$

$$G^2 = 5.56 = 2\left[16\log(16/14.02) + \cdots + 3\log(3/1.05)\right]\,.$$

$$\begin{aligned} df &= (2-1)[(2)(3)-1] = 5 \\ &= (I-1)[JK-1] \end{aligned}$$

Note: $\hat{m}_{ijk}^{(1)} = \hat{m}_{i\cdot\cdot}\hat{m}_{\cdot jk}/n_{\cdot\cdot\cdot}$

Note: $n_{i\cdot\cdot} = \hat{m}_{i\cdot\cdot}^{(0)}$, $n_{\cdot jk} = \hat{m}_{\cdot jk}^{(0)}$ for all i, j, and k

3.2.3 Two factors independent given a third

Rows and columns independent given layers

$$M^{(4)}: p_{ijk} = p_{i\cdot k}p_{\cdot jk}/p_{\cdot\cdot k}\,,$$

Rows and layers independent given columns

$$M^{(5)}: p_{ijk} = p_{ij\cdot}p_{\cdot jk}/p_{\cdot j\cdot}\,,$$

Columns and layers independent given rows

$$M^{(6)}: p_{ijk} = p_{ij\cdot}p_{i\cdot k}/p_{i\cdot\cdot}$$

For $M^{(4)}$

$$\begin{aligned}\hat{p}^{(4)}_{ijk} &= \hat{p}_{i\cdot k}\hat{p}_{\cdot jk}/\hat{p}_{\cdot\cdot k} \\ &= (n_{i\cdot k}/n_{\dots})(n_{\cdot jk}/n_{\dots})/(n_{\cdot\cdot k}/n_{\dots}) \\ &= n_{i\cdot k}n_{\cdot jk}/n_{\cdot\cdot k}n_{\dots}\,.\end{aligned}$$

MLE for $m_{ijk} = n_{\dots}p_{ijk}$

$$\begin{aligned}\hat{m}^{(4)}_{ijk} &= n_{\dots}\hat{p}_{i\cdot k}\hat{p}_{\cdot jk}/\hat{p}_{\cdot\cdot k} \\ &= n_{i\cdot k}n_{\cdot jk}/n_{\cdot\cdot k}\,.\end{aligned}$$

MLE's satisfy $M^{(4)}$ and
the marginal relations $\hat{m}_{i\cdot k} = n_{i\cdot k}$, $\hat{m}_{\cdot jk} = n_{\cdot jk}$, $\hat{m}_{\cdot\cdot k} = n_{\cdot\cdot k}$.

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{\left(n_{ijk} - \hat{m}^{(4)}_{ijk}\right)^2}{\hat{m}^{(4)}_{ijk}}$$

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log\left(n_{ijk}/\hat{m}^{(4)}_{ijk}\right)$$

$$df = (I-1)(J-1)K\,.$$

Pool the K separate tests for each individual layer.

EXAMPLE Examine $M^{(6)}$: cholesterol and diastolic blood pressure independent given personality type

n_{1jk}	(Personality type A) Diastolic Blood Pressure	
Cholesterol	Normal	High
Normal	716	79
High	207	25

n_{2jk}	(Personality type B) Diastolic Blood Pressure	
Cholesterol	Normal	High
Normal	819	67
High	186	22

$\hat{m}_{ijk}^{(6)}$			Diastolic Blood Pressure	
		Cholesterol	Normal	High
Personality	A	Normal	714.5	80.51
		High	208.5	23.49
	B	Normal	813.9	72.08
		High	191.1	16.92

$$X^2 = 2.188 \quad \text{and} \quad G^2 = 2.062$$

$$df = 2\,.$$

For testing complete independence $P = .067$.

3.2.4 Equal Odds Ratios Model

$$M^{(7)} : \frac{p_{111}p_{ij1}}{p_{i11}p_{1j1}} = \frac{p_{11k}p_{ijk}}{p_{i1k}p_{1jk}} \quad \text{all } i, j, k$$

$M^{(7)}$ is for layers fixed
Model unchanged if rows or columns fixed.
No simple formulae for $\hat{p}^{(7)}_{ijk}$ or $\hat{m}^{(7)}_{ijk}$
Iterative computing methods must be used.

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(n_{ijk} - \hat{m}^{(7)}_{ijk}\right)^2}{\hat{m}^{(7)}_{ijk}}$$

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk} \log\left(n_{ijk}/\hat{m}^{(7)}_{ijk}\right) .$$

$$df = (I-1)(J-1)(K-1) .$$

EXAMPLE Severity of drivers' injuries in auto accidents

n_{ijk}		Accident Type (k)			
		Collision		Rollover	
Injury (j)		Not Severe	Severe	Not Severe	Severe
Driver Ejected (i)	No	350	150	60	112
	Yes	26	23	19	80

$$I = J = K = 2, \text{so} \quad M^{(7)}: \frac{p_{111}p_{221}}{p_{121}p_{211}} = \frac{p_{112}p_{222}}{p_{122}p_{212}}.$$

$\hat{m}^{(7)}_{ijk}$		Accident Type (k)			
		Collision		Rollover	
Injury (j)		Not Severe	Severe	Not Severe	Severe
Driver Ejected (i)	No	350.5	149.5	59.51	112.5
	Yes	25.51	23.49	19.49	79.51

$$X^2 = .04323 \quad \text{and} \quad G^2 = .04334\,.$$

$$df = (2-1)(2-1)(2-1) = 1$$

$$\frac{\hat{p}_{111}\hat{p}_{221}}{\hat{p}_{121}\hat{p}_{211}} = \frac{\hat{p}_{112}\hat{p}_{222}}{\hat{p}_{122}\hat{p}_{212}}$$

$$\frac{350.5(23.49)}{25.51(149.5)} = \frac{59.51(79.51)}{19.49(112.5)} = 2.16$$

3.3 Odds Ratios for Independence Models

If $M^{(4)}$ is true, then $p_{ijk} = p_{i\cdot k}p_{\cdot jk}/p_{\cdot\cdot k}$ (Layers fixed)

$$
\begin{aligned}
\frac{p_{ijk}p_{i'j'k}}{p_{ij'k}p_{i'jk}} &= \frac{(p_{i\cdot k}p_{\cdot jk}/p_{\cdot\cdot k})(p_{i'\cdot k}p_{\cdot j'k}/p_{\cdot\cdot k})}{(p_{i\cdot k}p_{\cdot j'k}/p_{\cdot\cdot k})(p_{i'\cdot k}p_{\cdot jk}/p_{\cdot\cdot k})} \\
&= \frac{p_{i\cdot k}p_{\cdot jk}p_{i'\cdot k}p_{\cdot j'k}}{p_{i\cdot k}p_{\cdot j'k}p_{i'\cdot k}p_{\cdot jk}} \\
&= 1\,.
\end{aligned}
$$

Odds ratios with row or column fixed are constant. Row fixed:

$$
\begin{aligned}
\frac{p_{ijk}p_{ij'k'}}{p_{ijk'}p_{ij'k}} &= \frac{(p_{i\cdot k}p_{\cdot jk}/p_{\cdot\cdot k})(p_{i\cdot k'}p_{\cdot j'k'}/p_{\cdot\cdot k'})}{(p_{i\cdot k'}p_{\cdot jk'}/p_{\cdot\cdot k'})(p_{i\cdot k}p_{\cdot j'k}/p_{\cdot\cdot k})} \\
&= \frac{p_{i\cdot k}p_{\cdot jk}p_{i\cdot k'}p_{\cdot j'k'}}{p_{i\cdot k'}p_{\cdot jk'}p_{i\cdot k}p_{\cdot j'k}} \\
&= \frac{p_{\cdot jk}p_{\cdot j'k'}}{p_{\cdot jk'}p_{\cdot j'k}}\,.
\end{aligned}
$$

$M^{(5)}$, and $M^{(6)}$ are similar

$M^{(1)}$ is equivalent to

$$p_{ijk}p_{i'j'k}/p_{ij'k}p_{i'jk} = 1 \qquad \text{(layers fixed)}$$

for all i, i', j, j', and k and

$$p_{ijk}p_{i'jk'}/p_{ijk'}p_{i'jk} = 1 \qquad \text{(columns fixed)}$$

for all i, i', j, k, and k'
$M^{(2)}$, and $M^{(3)}$ are similar

$M^{(0)}$ is equivalent to

$$p_{ijk}p_{i'j'k}/p_{ij'k}p_{i'jk} = 1 \qquad \text{(layers fixed)}$$

for all i, i', j, j', and k and

$$p_{ijk}p_{i'jk'}/p_{ijk'}p_{i'jk} = 1 \qquad \text{(columns fixed)}$$

for all i, i', j, k, and k'

$$p_{ijk}p_{ij'k'}/p_{ijk'}p_{ij'k} = 1 \qquad \text{(rows fixed)}$$

for all i, j, j', k, and k'

3.4 Iterative Computation of Estimates

Fit $M^{(7)}$, compute $\hat{m}_{ijk}^{(7)}$'s.

Two standard algorithms: *Newton-Raphson algorithm* and *iterative proportional fitting algorithm*.

Newton-Raphson does a series of weighted regression analyses. *iteratively reweighted least squares.*

Iterative proportional fitting [Deming and Stephan algorithm] for balanced ANOVA-type log-linear models

3.5 Log-Linear Models for Three-Dimensional Tables

Saturated model:

$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$.

Eight submodels:

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \tag{0}$$

$$\begin{aligned} \log(m_{ijk}) &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} &(1)\\ \log(m_{ijk}) &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} &(2)\\ \log(m_{ijk}) &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} &(3) \end{aligned}$$

$$\begin{aligned} \log(m_{ijk}) &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)} &(4)\\ \log(m_{ijk}) &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)} &(5)\\ \log(m_{ijk}) &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} &(6) \end{aligned}$$

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}\,. \tag{7}$$

Log-linear model (i) holds if and only if $M^{(i)}$ holds.

	MODEL	SHORTHAND
(0)	$\log(m_{ijk}) = u_{1(i)} + u_{2(j)} + u_{3(k)}$	[1][2][3]
(1)	$\log(m_{ijk}) = u_{1(i)} + u_{23(jk)}$	[1][23]
(2)	$\log(m_{ijk}) = u_{2(j)} + u_{13(ik)}$	[2][13]
(3)	$\log(m_{ijk}) = u_{3(k)} + u_{12(ij)}$	[3][12]
(4)	$\log(m_{ijk}) = u_{13(ik)} + u_{23(jk)}$	[13][23]
(5)	$\log(m_{ijk}) = u_{12(ij)} + u_{23(jk)}$	[12][23]
(6)	$\log(m_{ijk}) = u_{12(ij)} + u_{13(ik)}$	[12]][13]
(7)	$\log(m_{ijk}) = u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$	[12][13][23]

Test [1][23] (model (1)) against [13][23] (model (4))
use $\hat{m}_{ijk}^{(1)}$'s, $\hat{m}_{ijk}^{(4)}$'s,

Pearson chi-square

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(\hat{m}_{ijk}^{(4)} - \hat{m}_{ijk}^{(1)}\right)^2}{\hat{m}_{ijk}^{(1)}} .$$

Likelihood ratio chi-square

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \hat{m}_{ijk}^{(4)} \log\left(\hat{m}_{ijk}^{(4)} / \hat{m}_{ijk}^{(1)}\right) .$$

Test of no row-layer interactions
degrees of freedom are for row-layer interactions, $(I-1)(K-1)$

Term	Degrees of Freedom
u	1
u_1	$I-1$
u_2	$J-1$
u_3	$K-1$
u_{12}	$(I-1)(J-1)$
u_{13}	$(I-1)(K-1)$
u_{23}	$(J-1)(K-1)$
u_{123}	$(I-1)(J-1)(K-1)$

G^2 for submodel tests are obtained from original G^2s by differences.
(Doesn't work for X^2.)

Example Classroom behavior

Model	G^2	df
$M^{(0)}$: [1][2][3]	16.42	7
$M^{(1)}$: [1][23]	5.56	5
$p_{\cdot jk} = p_{\cdot j \cdot} p_{\cdot \cdot k}$	10.86	2

$$G^2 = 16.42 - 5.56 = 10.86$$

$$df = 7 - 5 = 2$$

Example Personality (1), cholesterol (2), and diastolic blood pressure (3)

Model	*df*	X^2	G^2	P
[12][13][23]	1	0.617	0.613	.434
[12][13]	2	2.188	2.062	.358
[12][23]	2	2.985	2.980	.224
[13][23]	2	4.566	4.563	.100
[1][23]	3	7.102	7.101	.067
[2][13]	3	6.189	6.184	.102
[3][12]	3	4.543	4.601	.207
[1][2][3]	4	8.730	8.723	.067

Testing [1][2][3] against [12][13]

$$\begin{aligned} G^2 &= 8.723 - 2.062 = 6.661 \\ df &= 4 - 2 = 2 \\ \chi^2(.95, 2) &= 5.99 \end{aligned}$$

Testing [3][12] against [12][13]

$$\begin{aligned} G^2 &= 4.601 - 2.062 = 2.539 \\ df &= 3 - 2 = 1 \\ \chi^2(.95, 1) &= 3.84 \end{aligned}$$

Testing [3][12] against [12][13][23]

$$\begin{aligned} G^2 &= 4.601 - 0.613 = 3.988 \\ df &= 3 - 1 = 2 \\ \chi^2(.95, 2) &= 5.99 \end{aligned}$$

3.6 Product-Multinomial and Other Sampling Plans

Opinions about legalized abortion

		Opinion (k)		
Sex (i)	Party (j)	Support	Do Not Support	Total
	Republican	79	40	119
Female	Democrat	132	71	203
	Independent	98	80	178
	Total	309	191	500
	Republican	65	94	159
Male	Democrat	141	95	236
	Independent	113	92	205
	Total	319	281	600

Estimates satisfy the constraints

$$\hat{m}_{i\cdot\cdot} = n_{i\cdot\cdot}$$

Sex (i)	Status (j)	Opinion (k) Support	Do Not Support	Total
	Low	171	79	250
Female	Not Low	138	112	250
	Total	309	191	500
	Low	152	148	300
Male	Not Low	167	133	300
	Total	319	281	600

Any model that estimates expected cell counts must also incorporate the condition that

$$\hat{m}_{ij\cdot} = n_{ij\cdot}$$

Possible Models with $m_{ij\cdot}$ Fixed by the Sampling Design
[123]
[12][13][23]
[12][13]
[12][23]
[12][3]

3.7 Model Selection: Akaike's information criterion (AIC)

Model, say, X

$$A_X = G^2(X) - [q - 2r]$$

$G^2(X)$ for testing the X model against the saturated model,
r is the *df* for the X model,
q cells in the table.

$$\begin{aligned} A_X - q &= G^2(X) - 2[q - r] \\ &= G^2(X) - 2\ (\text{test degrees of freedom})\ . \end{aligned}$$

Example Personality (1), cholesterol (2), blood pressure (3) data.

Model	*df*	G^2	$A - q$
[12][13][23]	1	0.613	−1.387
[12][13]	2	2.062	−1.938
[12][23]	2	2.980	−1.020
[13][23]	2	4.563	0.563
[1][23]	3	7.101	1.101
[2][13]	3	6.184	0.184
[3][12]	3	4.602	−1.398
[1][2][3]	4	8.723	0.723

3.8 Higher Dimensional Tables

Example

Factor	Abbreviation	Levels
Change in Muscle Tension	T	High, Low
Weight of Muscle	W	High, Low
Muscle	M	Type 1, Type 2
Drug	D	Drug 1, Drug 2

Product multinomial: total count for each muscle type fixed.

			Drug(k)	
Tension(h)	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

$$\log(m_{hijk}) = \gamma + \tau_h + \omega_i + \mu_j + \delta_k, \tag{1}$$

$$\begin{aligned}\log(m_{hijk}) &= \gamma + \tau_h + \omega_i + \mu_j + \delta_k + (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} \\ &\quad + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk},\end{aligned} \tag{2}$$

$$\begin{aligned}\log(m_{hijk}) &= \gamma + \tau_h + \omega_i + \mu_j + \delta_k + (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} \\ &\quad + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk} \\ &\quad + (\tau\omega\mu)_{hij} + (\tau\omega\delta)_{hik} + (\tau\mu\delta)_{hjk} + (\omega\mu\delta)_{ijk}.\end{aligned} \tag{3}$$

Remove some redundant parameters,

$$\log(m_{hijk}) = \tau_h + \omega_i + \mu_j + \delta_k, \tag{1}$$

$$\log(m_{hijk}) = (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk}, \tag{2}$$

$$\log(m_{hijk}) = (\tau\omega\mu)_{hij} + (\tau\omega\delta)_{hik} + (\tau\mu\delta)_{hjk} + (\omega\mu\delta)_{ijk} \tag{3}$$

Shorthand notations

$$\text{[T][W][M][D]} \tag{1}$$

$$\text{[TW][TM][WM][TD][WD][MD]} \tag{2}$$

$$\text{[TWM][TWD][TMD][WMD]} \tag{3}$$

Model	df	G^2	P
[TWM][TWD][TMD][WMD]	1	0.11	.74
[TW][TM][WM][TD][WD][MD]	5	47.67	.00
[T][W][M][D]	11	127.4	.00

Test [TWM][TWD][TMD][WMD] versus
reduced model [TW][TM][WM][TD][WD][MD]

$$G^2 = 47.67 - 0.11 = 47.56 \quad df = 5 - 1 = 4$$

3.9 Computer Commands

Muscle tension data in file 'tension.dat'
counts plus indices for tension, weight, muscle type, and drug

```
 3 1 1 1 1
21 1 1 1 2
23 1 1 2 1
11 1 1 2 2
22 1 2 1 1
32 1 2 1 2
 4 1 2 2 1
12 1 2 2 2
 3 2 1 1 1
10 2 1 1 2
41 2 1 2 1
21 2 1 2 2
45 2 2 1 1
23 2 2 1 2
 6 2 2 2 1
22 2 2 2 2
```

Fit [WMD][TWM][TWD][TMD] using SAS PROC GENMOD.

```
options ps=60 ls=72 nodate;
data tension;
   infile 'tension.dat';
   input n T W M D;
proc genmod data=tension;
   class T W M D;
   model n = W*M*T T*W*M T*W*D T*M*D / link=log
                                       dist=poisson;
run;
```

"link=log" and "dist=poisson"
GENMOD uses the Newton-Raphson algorithm.
To fit [TM][WM][MD] use T∗M W∗M M∗D
For [T][WM][D] use T W∗M D

"class" command specifies that a variable gives indices for identifying levels of an analysis of variance type factor.
(Otherwise variables treated as a predictor variable in regression)

Marginal Association:
For TWMD, test [TWM][TWD][TMD][WMD] versus [TWMD].
For TWM, test [TW][WM][TM] versus [TWM].
For WD, test [W][D] versus [WD].
Partial Association:
For TWMD, in four-factor table, same test as marginal association.
For TWM, test [TWD][TMD][WMD] versus
[TWM][TWD][TMD][WMD].
For WD, test [TW][TM][TD][WM][MD] versus
[TW][TM][TD][WM][WD][MD].

Brown's Tests for the Muscle Tension Data

Effect	Partial Association G^2	P	Marginal Association G^2	P
T	6.04	.01	—	—
W	3.55	.06	—	—
M	1.18	.28	—	—
D	0.08	.78	—	—
TW	2.35	.13	0.06	.80
TM	6.81	.01	5.27	.02
WM	63.66	.00	62.25	.00
TD	6.02	.01	6.37	.01
WD	0.65	.42	1.12	.29
MD	0.17	.68	1.40	.24
TWM	1.00	.32	2.63	.10
TWD	0.01	.93	0.04	.85
TMD	2.86	.09	6.01	.01
WMD	35.65	.00	40.49	.00
TWMD	0.14	.70	0.14	.70

Factor	Abbreviation	Levels
Race	R	White, Nonwhite
Sex	S	Male, Female
Opinion	O	Yes = Supports Legalized Abortion
		No = Opposed to Legalized Abortion
		Und = Undecided
Age	A	18-25, 26-35, 36-45, 46-55, 56-65, 66+ years

Abortion Opinion Data

			AGE					
RACE	SEX	OPINION	18-25	26-35	36-45	46-55	56-65	66+
White	Male	Yes	96	138	117	75	72	83
		No	44	64	56	48	49	60
		Und	1	2	6	5	6	8
	Female	Yes	140	171	152	101	102	111
		No	43	65	58	51	58	67
		Und	1	4	9	9	10	16
Nonwhite	Male	Yes	24	18	16	12	6	4
		No	5	7	7	6	8	10
		Und	2	1	3	4	3	4
	Female	Yes	21	25	20	17	14	13
		No	4	6	5	5	5	5
		Und	1	2	1	1	1	1

Abortion Opinions: SAS commands for [RSO][RSA][ROA][SOA]

```
options ps=60 ls=72 nodate;
data abort;
   infile 'abort.dat';
   input R S A O N;
proc genmod data=abort;
   class R S A O;
   model N = R*S*O R*S*A R*O*A S*O*A / link=log
                                       dist=poisson;
run;
```

BMDP-4F commands for [RSO][RSA][ROA][SOA],

```
/ INPUT      FILE = 'C:\LOGLIN\ABORT.DAT'.
         FORMAT = FREE.
         VARIABLES = 5.
/ VARIABLE  NAMES = R, S, A, O, N.
/ TABLE     INDICES = R, S, A, O.
            COUNT = N.
/ STAT      ALL.
/ FIT       MODEL = RSO, RSA, ROA, SOA.
/ PRINT     LINE = 79.
/ END
```

BMDP-4F uses iterative proportional fitting.
BMDP-4F is the most powerful program I am aware of for fitting analysis of variance type log-linear models.

Model	df	G^2	$A-q$
[RSO][OA]	45	24.77	−65.23
[RO][SO][OA]	48	36.60	−59.40
[R][S][OA]	52	59.78	−44.22
[RA][S][OA]	47	53.77	−40.23
[R][SO][OA]	50	50.59	−49.41
[RO][S][OA]	50	45.79	−54.21

Estimated cell counts under [RSO][OA]

			Age					
Race	Sex	Opinion	18-25	26-35	36-45	46-55	56-65	65+
		Support	105.5	132.1	114.5	76.94	72.81	79.19
	Male	Oppose	41.87	61.93	54.95	47.98	52.34	61.93
		Undec.	1.39	2.50	5.27	5.27	5.54	8.04
White								
		Support	141.0	176.7	153.1	102.9	97.38	105.9
	Female	Oppose	44.61	65.98	58.55	51.11	55.76	65.98
		Undec.	2.43	4.37	9.22	9.22	9.70	14.07
		Support	14.52	18.19	15.76	10.59	10.03	10.90
	Male	Oppose	5.61	8.30	7.36	6.43	7.01	8.30
		Undec.	0.84	1.52	3.20	3.20	3.37	4.88
NonWhite								
		Support	19.97	25.01	21.67	14.57	13.79	14.99
	Female	Oppose	3.91	5.79	5.14	4.48	4.89	5.79
		Undec.	0.35	0.62	1.32	1.32	1.39	2.01

```
 COOK'S DISTANCE
     -
     -                          *
     -
0.075+
     -
     -                     *
     -  *
     -
0.050+
     -                        *   *
     -
     -
     -             *  *
0.025+                *                             *
     -  *     *                         *
     -                           *                    *
     -    **              *
     -   ** *  *2*    * *   *       *  ** **  **    *
0.000+    *  **   2***  **** *2 *** **2  ** 2** 2*2* *2
       +---------+---------+---------+---------+---------+
       0        15        30        45        60        75
                             INDEX
```

Race, Sex, Opinion Marginal Table

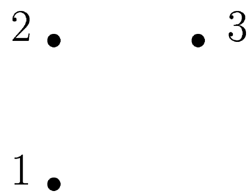
		Opinion			
Race	Sex	Support	Oppose	Undec.	Totals
White	Male	581	321	28	930
	Female	777	342	49	1168
Nonwhite	Male	80	43	17	140
	Female	110	30	7	147
	Totals	1548	736	101	2385

Opinion, Age Marginal Table

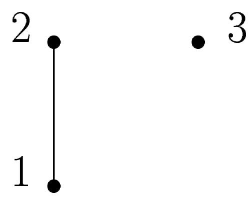
	Age						
Opinion	18-25	26-35	36-45	46-55	56-65	65+	Totals
Support	281	352	305	205	194	211	1548
Oppose	96	142	126	110	120	142	736
Undec.	5	9	19	19	20	29	101
Totals	382	503	450	334	334	382	2385

3.10 Graphical Models: Displaying models with graphs.

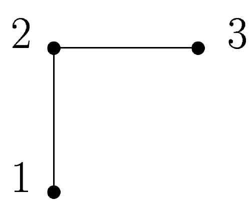
Factors 1, 2, and 3 independent: M^0: [1][2][3]



Factor 3 independent of 1 and 2: M^1: [12][3]

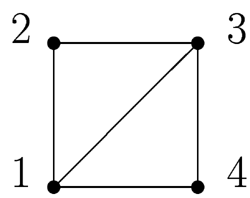


Factors 1 and 3 independent given 2: M^5: [12][23]

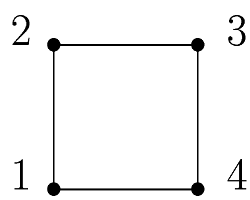


Graphical models are determined by their two-factor interactions.

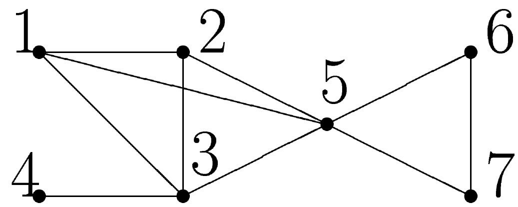
Factors 2 and 4 independent given 1 and 3: [123][134]



Factors 1 and 3 independent given 2 and 4 Factors 2 and 4 independent given 1 and 3: [12][23][34][41]



Model: [1235][34][567]



References

Christensen, Ronald (1996a). *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. London: Chapman and Hall. [Introductory and Intermediate]

Christensen, Ronald (1997). *Log-Linear Models and Logistic Regression*, Second Edition. New York: Springer-Verlag. [Intermediate and Advanced]

Christensen, Ronald (1996b). *Plane Answers to Complex Questions: The Theory of Linear Models*, Second Edition. New York: Springer-Verlag. [Advanced]

Christensen, Ronald (1990). *Linear Models for Multivariate, Time Series and Spatial Data*. New York: Springer-Verlag. [Advanced]

Everitt, B.S. (1977). *The Analysis of Contingency Tables*. London: Chapman and Hall. [Introductory]

Fienberg, Stephen E. (1980). *The Analysis of Cross-Classified Categorical Data*, Second Edition. Cambridge, Massachusetts: MIT Press. [Introductory]

Agresti, Alan (1990). *Categorical Data Analysis*. New York: John Wiley and Sons. [Intermediate]

Hosmer, David W. and Lemeshow, Stanley (1989). *Applied Logistic Regression*. New York: John Wiley and Sons. [Intermediate]

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition. London: Chapman and Hall. [Intermediate]

Anderson, E.B. (1980). *Discrete Statistical Models with Social Science Applications*. North-Holland, Amsterdam. [Advanced]

Bishop, Yvonne M.M., Fienberg, Steven E., and Holland, Paul W. (1975). *Discrete Multivariate Analysis*. Cambridge, Massachusetts: MIT Press. [Advanced]

Haberman, Shelby J. (1974). *The Analysis of Frequency Data*. Chicago: University of Chicago Press. [Advanced]

Santner, Thomas J. and Duffy, Diane E. (1989). *The Statistical Analysis of Discrete Data*. New York: Springer-Verlag. [Advanced]

Edwards, David (1995). *Introduction to Graphical Modeling*. Berlin: Springer-Verlag. [Intermediate]

Lauritzen, Steffen L. (1996). *Graphical Models*. Oxford: Oxford University Press. [Advanced]

Whittaker, Joe (1990). *Graphical Models in Applied Multivariate Statistics*. New York: John and Wiley and Sons. [Advanced]