Chap. 7 Loglinear Models

● Logistic regression distinguishes between response variable Y and explanatory variables x_1, x_2, ...

● Loglinear models treat all variables as response variables(like correlation analysis)

ex) Survey of High school students

$Y_1 =$ used marijuana? (Yes, No)

$Y_2 =$ alcohol? (Yes, No)

$Y_3 =$ cigarettes? (Yes, No)

Any variables independent?

Strength of association?

Interaction?

Loglinear models treat cell counts as Poisson and use the log-link function.
Motivation : In $I \times J$ table, X and Y are independent?

If $P(X=i,\ Y=j) = P(X=i)P(Y=j)$ for $all\ i,j$

$\Rightarrow \qquad \pi_{ij} = \pi_{i+}\pi_{+j}$

For expected frequencies

$$\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j} \ \ \text{(indep.)}$$

$$\log\mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$$

where λ_i^X : effect of X falling in row i, λ_j^Y : effect of Y falling in column j.

This is loglinear model of independence

Treat X and Y symmetrically (differs form logistic regression, which distinguishes between $Y=$ response and $X=$ explanatory)

ex) Revisit Income and job satisfaction (See model ❹ in SAS output)

Income ($1,000)	Job satisfaction Very Dissat	Little Dissat	Moderate Satisfied	Very Satisfied	Total
< 5	2	4	13	3	22
5~15	2	6	22	4	34
15~25	0	1	15	8	24
> 25	0	3	13	8	24
Total	4	14	63	23	104

Using $x=$ income scores (3,10, 20, 30), we used SAS (PROC LOGISTIC) to fit model

$$\log\left(\frac{\pi_i}{\pi_4}\right)=\alpha_i+\beta_j x,\ \ j=1,2,3$$

(We analyzed this using multinomial logit models in ch.6)

Let's consider loglinear model (Income (I) and job satisfaction (S))

Model

$$\log(\mu_{ij})=\lambda+\lambda_i^I+\lambda_j^S$$

can be expressed as

$$\log\mu_{ij}=\lambda+\lambda_1^I Z_1+\lambda_2^I Z_2+\lambda_3^I Z_3+\lambda_1^s W_1+\lambda_2^S W_2+\lambda_3^S W_3$$

where $Z_1=\begin{cases}1, & income\ \ cat.\ 1\\ 0, & o.w.\end{cases}$ $\quad W_1=\begin{cases}1, & sat.\ \ cat.\ 1\\ 0, & o.w.\end{cases}$

$\cdots \qquad\qquad \cdots$

$Z_3=\begin{cases}1, & income\ \ cat.\ 3\\ 0, & o.w.\end{cases}$ $\quad W_3=\begin{cases}1, & sat.\ \ cat.\ 3\\ 0, & o.w.\end{cases}$

Parameter	No. non-redundant	
λ	1	
λ_i^X	$I-1$	(can set $\lambda_I^X=0$)
λ_j^Y	$J-1$	(can set $\lambda_J^Y=0$)
λ_{ij}^{XY}	$(I-1)(J-1)$	(no. of products of dummy variables)

Note:

For a Poisson loglinear model

$$df=no.\ Poisson\ count\ -\ no.\ parameters$$

$(no.\ Poisson\ counts=no.\ cells)$

ex) Independence model, $I\times J$ table

$$\log\mu_{ij}=\lambda+\lambda_i^X+\lambda_j^Y$$

$$df=IJ-\{1+(I-1)+(J-1)\}=(I-1)(J-1)$$

Test of indep. using Pearson X^2 or likelihood ratio G^2 is a goodness-of-fit test of the indep. loglinear model.

The model allowing association

$$\log\mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

has $df = 0$ (saturated), giving a perfect fit.

ex) Revisit Income and job satisfaction

- Indep. model:

$$\log(\mu_{ij}) = \lambda + \lambda_i^I + \lambda_j^S$$

has $X^2 = 11.5(p-value = 0.243)$, $G^2 = 13.5(p-value = 0.141)$, $df = 9$

- Saturated model: $X^2 = G^2 = 0$, $df = 0(All\ \widehat{\mu_{ij}} = n_{ij})$

Estimated odds ratio in saturated model using highest and lowest categories is

$$\begin{aligned}\frac{\hat{\mu}_{11}\widehat{\mu_{44}}}{\hat{\mu}_{14}\hat{\mu}_{41}} &= \exp\left[\hat{\lambda}_{11}^{IS} + \hat{\lambda}_{44}^{IS} - \hat{\lambda}_{14}^{IS} - \hat{\lambda}_{41}^{IS}\right] \\ &= \exp(24.288) = 35,294,747,720(GENMOD) \\ &= \frac{n_{11}n_{44}}{n_{14}n_{41}} = \frac{2\times 8}{3\times 0} = \infty\end{aligned}$$

since model is saturated.

(software doesn't quite get right answer when $ML\ est. = \infty$)

Income ($1,000)	Very Dissat	Little Dissat	Moderate Satisfied	Very Satisfied
	Job satisfaction			
< 5	$\boxed{2}$	4	13	$\boxed{3}$
5~15	2	6	22	4
15~25	0	1	15	8
> 25	$\boxed{0}$	3	13	$\boxed{8}$

```
data jobsatis;
 input income satis count @@;
 cards;
 3  1 2  3 2 4  3 3 13  3 4 3
 10 1 2 10 2 6 10 3 22 10 4 4
 20 1 0 20 2 1 20 3 15 20 4 8
 30 1 0 30 2 3 30 3 13 30 4 8
;

/* Independence loglinear model */
❶ proc genmod data=jobsatis;
    class income satis;
    model count = income satis / dist=poi link=log;
   run;

/* Saturated loglinear model */
❷ proc genmod data=jobsatis;
    class income satis;
    model count = income satis income*satis / dist=poi link=log;
   run;
```

❶

The GENMOD Procedure

Model Information

Data Set	WORK.JOBSATIS
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	16
Number of Observations Used	16

Class Level Information

Class	Levels	Values
income	4	3 10 20 30
satis	4	1 2 3 4

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	9	13.4673	1.4964
Scaled Deviance	9	13.4673	1.4964
Pearson Chi-Square	9	11.5242	1.2805
Scaled Pearson X2	9	11.5242	1.2805
Log Likelihood		129.0550	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	1.6692	0.2748	1.1305	2.2078	36.89	<.0001
income	3	1	-0.0870	0.2952	-0.6655	0.4915	0.09	0.7682
income	10	1	0.3483	0.2666	-0.1742	0.8708	1.71	0.1914
income	20	1	0.0000	0.2887	-0.5658	0.5658	0.00	1.0000
income	30	0	0.0000	0.0000	0.0000	0.0000	.	.
satis	1	1	-1.7492	0.5417	-2.8110	-0.6874	10.43	0.0012
satis	2	1	-0.4964	0.3390	-1.1608	0.1679	2.14	0.1431
satis	3	1	1.0076	0.2436	0.5302	1.4851	17.11	<.0001
satis	4	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

❷ The GENMOD Procedure

Model Information

Data Set	WORK.JOBSATIS
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	16
Number of Observations Used	16

Class Level Information

Class	Levels	Values
income	4	3 10 20 30
satis	4	1 2 3 4

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	0	0.0000	.
Scaled Deviance	0	0.0000	.
Pearson Chi-Square	0	0.0000	.
Scaled Pearson X2	0	0.0000	.
Log Likelihood		135.7886	

WARNING: Negative of Hessian not positive definite.

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept			1	2.0794	0.3536	1.3865	2.7724	34.59	<.0001
income	3		1	-0.9808	0.6770	-2.3077	0.3461	2.10	0.1474
income	10		1	-0.6931	0.6124	-1.8934	0.5071	1.28	0.2577
income	20		1	0.0000	0.5000	-0.9800	0.9800	0.00	1.0000
income	30		0	0.0000	0.0000	0.0000	0.0000	.	.
satis	1		1	-24.7726	0.8660	-26.4700	-23.0752	818.24	<.0001
satis	2		1	-0.9808	0.6770	-2.3077	0.3461	2.10	0.1474
satis	3		1	0.4855	0.4494	-0.3952	1.3662	1.17	0.2799
satis	4		0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	3	1	1	24.3671	1.2583	21.9009	26.8334	375.00	<.0001
income*satis	3	2	1	1.2685	1.0206	-0.7319	3.2689	1.54	0.2139
income*satis	3	3	1	0.9808	0.7824	-0.5527	2.5143	1.57	0.2100
income*satis	3	4	0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	10	1	0	24.0794	0.0000	24.0794	24.0794	.	.
income*satis	10	2	1	1.3863	0.9354	-0.4471	3.2197	2.20	0.1383
income*satis	10	3	1	1.2192	0.7053	-0.1630	2.6015	2.99	0.0838
income*satis	10	4	0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	20	1	1	-0.0000	84674.82	-165960	165959.6	0.00	1.0000
income*satis	20	2	1	-1.0986	1.2583	-3.5648	1.3676	0.76	0.3826
income*satis	20	3	1	0.1431	0.6274	-1.0865	1.3727	0.05	0.8196
income*satis	20	4	0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	30	1	0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	30	2	0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	30	3	0	0.0000	0.0000	0.0000	0.0000	.	.
income*satis	30	4	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Loglinear Models for Three-way Tables

Two-factor terms represent conditional log odds ratios, at fixed level of third variable

(X, Y, Z)

Cell probabilities: $\{\pi_{ijk}\}$

Cell counts: $\{n_{ijk}\}$

Expected counts: $\{\mu_{ijk}\}$.

Types of independence

Def. (X, Y, Z) are mutually independent if

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k} \text{ for all } i, j, k.$$

Corresponds to loglinear model of form

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

Denoted by (X, Y, Z).

Def. Y is jointly independent of X and Z if

$$\pi_{ijk} = \pi_{i+k}\pi_{+j+} \text{ for all } i, j, k.$$

Corresponding model is

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}.$$

Denoted by (XZ, Y).

Def. X and Y are conditionally independent given Z if

$$\pi_{ij|k} = \pi_{i+|k}\pi_{+j|k} \text{ for all } i, j, k.$$

Corresponding model is

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

Denoted by (XZ, YZ).

$I \times J \times K$ tables

The general loglinear model is

$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$

(X, Y, Z) ____________

(XY, Z) ________________

(XY, XZ) ______________________

(XY, XZ, YZ) _______________________

(XYZ) _________________________________

Interpretation of parameters

Local odds ratio:

$$\theta_{ij(k)} = \frac{\pi_{ijk}\,\pi_{i+1,j+1,k}}{\pi_{i+1,j,k}\,\pi_{i,j+1,k}} \quad \text{for } i=1,\dots,I;\ j=1,\dots,J;\ k=1,\dots,K.$$

● For model (XY, XZ, YZ) of no three factor interaction,

$$\begin{aligned}\log\theta_{ij(k)} &= \log\mu_{ijk} + \log\mu_{i+1,j+1,k} - \log\mu_{i+1,j,k} - \log\mu_{i,j+1,k} \\ &= \lambda_{ij}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i+1,j}^{XY} - \lambda_{i,j+1}^{XY}\end{aligned}$$

which is independent of k.

i.e., $\theta_{ij(1)} = \cdots = \theta_{ij(K)}$ for all i, j.

The conditional association between any pair of variables is identical at each category of the third. This is the homogeneous assocation model.

● Do $\{\lambda_{ij}^{XY}\}$ also describe marginal association? No.

(Collapsing doesn't have same association)

ex) $2\times2\times2$ table

Let μ_{ijk} denotes expected freq. : λ_{ik}^{XZ} and λ_{jk}^{YZ} denote association parameters

$$\log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

satisfies

● $\log\theta_{XY(Z)} = 0$ (X and Y conditionally indep. given Z)

$$\theta_{XY(Z)} = \frac{\mu_{22(k)}\mu_{11(k)}}{\mu_{12(k)}\mu_{21(k)}} \Rightarrow \log\theta_{XY(k)} = \log\mu_{22(k)} + \log\mu_{11(k)} - \log\mu_{12(k)} - \log\mu_{21(k)}$$

● $\log\theta_{X(j)Z} = \begin{cases}\lambda_{11}^{XZ} + \lambda_{22}^{XZ} - \lambda_{12}^{XZ} - \lambda_{21}^{XZ} \\ 0 \text{ if } \lambda_{ik}^{XZ} = 0\end{cases}$

$$\theta_{X(j)Z} = \frac{\mu_{2(j)2}\mu_{1(j)1}}{\mu_{1(j)2}\mu_{2(j)1}} \Rightarrow \log\theta_{X(j)Z} = \log\mu_{2(j)2} + \log\mu_{1(j)1} - \log\mu_{1(j)2} - \log\mu_{2(j)1}$$

i.e., the XZ odds ratio is same at all levels Y

Denoted by $(XZ,\ YZ)$, called model of XY conditional indep. given Z.

ex)

$$\log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

called model of homogeneous association

Each pair of variables has association that is identical at all levels of third variable. Denoted by $(XY,\ XZ,\ YZ)$

ex) Revisit Berkely admission data ($2\times 2\times 6$)

Gender (M, F) $\times$ Admitted (Y, N) $\times$ Department (1,2,3,4,5,6)

Recall marginal 2×2 AG table has $\hat{\theta}=1.84$

Model(AD, DG)

A and G conditionally indep. given D

eg) for Dept. 1,

$$\begin{aligned}\hat{\theta}_{AG(1)} &= \frac{531.4\times 38.4}{293.6\times 69.6}=1.0\\ &=\hat{\theta}_{AG(2)}=\cdots=\hat{\theta}_{AG(6)}\end{aligned}$$

(see model ❷(predicted values))

But model fit poorly : $G^2=21.7,\ X^2=19.9,\ df=6(p-value<.0001)$

for H_0 : Model(AD, DG) holds.

Conclude A and G are not conditionally indep. given D

****

$$\hat{\theta}_{AG(1)}=\frac{\hat{\mu}_{22(1)}\hat{\mu}_{11(1)}}{\hat{\mu}_{12(1)}\hat{\mu}_{21(1)}}$$

where
$$\begin{aligned}\hat{\mu}_{22(1)} &= \exp\left\{\hat{\lambda}+\hat{\lambda}_2^A+\hat{\lambda}_2^G+\hat{\lambda}_{21}^{AD}+\hat{\lambda}_{21}^{GD}\right\}\\ \hat{\mu}_{11(1)} &= \exp\left\{\hat{\lambda}+\hat{\lambda}_1^A+\hat{\lambda}_1^G+\hat{\lambda}_{11}^{AD}+\hat{\lambda}_{11}^{GD}\right\}\\ \hat{\mu}_{12(1)} &= \exp\left\{\hat{\lambda}+\hat{\lambda}_1^A+\hat{\lambda}_2^G+\hat{\lambda}_{11}^{AD}+\hat{\lambda}_{21}^{GD}\right\}\\ \hat{\mu}_{21(1)} &= \exp\left\{\hat{\lambda}+\hat{\lambda}_2^A+\hat{\lambda}_1^G+\hat{\lambda}_{21}^{AD}+\hat{\lambda}_{11}^{GD}\right\}\end{aligned}$$

****

● Model($AG,\ AD,\ DG$)

Also, permits AG association with same odds ratio for each dept.

eg., for Dept.1

$$\begin{aligned}\hat{\theta}_{AG(1)} &= \frac{529.3\times 36.3}{295.7\times 71.7}=0.90 \qquad \text{(see model ❸)}\\ &=\hat{\theta}_{AG(2)}=\cdots=\hat{\theta}_{AG(6)}\\ &=\exp(\hat{\lambda}_{11}^{AG}+\hat{\lambda}_{22}^{AG}-\hat{\lambda}_{12}^{AG}-\hat{\lambda}_{21}^{AG})\\ &=\exp(-0.0999)=0.9\end{aligned}$$

Controlling for dept., estimated odds of admission for <u>males</u> equal .90 times estimated odds for <u>females.</u>

$\hat{\theta}=1.84$ ignores dept. (Simson's Paradox)

But this model also fits poorly: $G^2=20.2,\ X^2=18.8,\ df=5(p-value<0.0001)$ for H_0 : $\text{model}(AG,\ AD,\ DG)\ holds$. i.e., true AG odds ratio not identical for each dept.

● Adding 3-factor interaction term λ_{ijk}^{GAD} gives saturated mode ($|X|\times 5$ cross products of dummies).

Residual analysis

For model $(AD,\ DG)$ or $(AD,\ AG,\ DG)$, only Dept.1 has large adjusted residuals(≈ 4 in absolute value).

Dept.1 has

- fewer males accepted than expected by model.
- more femals accepted than expected by model.

If refit model $(AD,\ DG)$ to $2\times 2\times 5$ table for Dept. 2-6, $G^2=2.7$, $df=5$, good fit

```
data berkeley;
 input dept $ gender $ admit $ count @@;
 cards;
 a male   yes 512 a male   no 313
 a female yes  89 a female no  19
 b male   yes 353 b male   no 207
 b female yes  17 b female no   8
 c male   yes 120 c male   no 205
 c female yes 202 c female no 391
 d male   yes 138 d male   no 279
 d female yes 131 d female no 244
 e male   yes  53 e male   no 138
 e female yes  94 e female no 299
 f male   yes  22 f male   no 351
 f female yes  24 f female no 317
 run;

/* (A,D,G)  */
❶ proc genmod data=berkeley;
    class dept(ref=last) gender(ref=last) admit(ref=first)/param=ref;
    model count=admit gender dept/dist=poi link=log;
   run;

/* (AD,DG)  */
❷ proc genmod data=berkeley;
    class dept(ref=last) gender(ref=last) admit(ref=first)/param=ref;
    model count=admit gender dept admit*dept dept*gender/dist=poi link=log obstats;
   run;

/* (AG,AD,DG)  */
❸ proc genmod data=berkeley;
    class dept(ref=last) gender(ref=last) admit(ref=first)/param=ref;
    model count=admit gender dept admit*gender admit*dept gender*dept/dist=poi link=log obstats;
   run;

/* Delete department A  */
   data bcdef;
    set berkeley;
    if dept^='a';
   run;

/* (AD,DG) without deparment A  */
❹ proc genmod data=bcdef;
    class dept(ref=last) gender(ref=last) admit(ref=first)/param=ref;
    model count=admit gender dept admit*dept dept*gender/dist=poi link=log;
   run;

/* (AG,AD,DG) without deparment A */
❺ proc genmod data=bcdef;
    class dept(ref=last) gender(ref=last) admit(ref=first)/param=ref;
    model count=admit gender dept admit*gender admit*dept gender*dept/dist=poi link=log;
   run;
```

❶ The GENMOD Procedure

Model Information

Data Set	WORK.BERKELEY
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	24
Number of Observations Used	24

Class Level Information

Class	Value	Design Variables				
dept	a	1	0	0	0	0
	b	0	1	0	0	0
	c	0	0	1	0	0
	d	0	0	0	1	0
	e	0	0	0	0	1
	f	0	0	0	0	0
gender	female	1				
	male	0				
admit	no	0				
	yes	1				

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	16	2097.6712	131.1045
Scaled Deviance	16	2097.6712	131.1045
Pearson Chi-Square	16	2000.3281	125.0205
Scaled Pearson X2	16	2000.3281	125.0205
Log Likelihood		19464.3700	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	5.5603	0.0411	5.4797	5.6409	18281.7	<.0001
admit	yes	1	-0.4567	0.0305	-0.5165	-0.3969	224.15	<.0001
gender	female	1	-0.3829	0.0303	-0.4422	-0.3235	159.93	<.0001
dept	a	1	0.2675	0.0497	0.1701	0.3650	28.95	<.0001
dept	b	1	-0.1993	0.0558	-0.3086	-0.0900	12.77	0.0004
dept	c	1	0.2513	0.0499	0.1535	0.3491	25.37	<.0001
dept	d	1	0.1037	0.0516	0.0025	0.2048	4.04	0.0445
dept	e	1	-0.2010	0.0558	-0.3103	-0.0916	12.98	0.0003
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

❷ The GENMOD Procedure

Model Information

Data Set	WORK.BERKELEY
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	24
Number of Observations Used	24

Class Level Information

Class	Value	Design Variables				
dept	a	1	0	0	0	0
	b	0	1	0	0	0
	c	0	0	1	0	0
	d	0	0	0	1	0
	e	0	0	0	0	1
	f	0	0	0	0	0
gender	female	1				
	male	0				
admit	no	0				
	yes	1				

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	6	21.7355	3.6226
Scaled Deviance	6	21.7355	3.6226
Pearson Chi-Square	6	19.9383	3.3231
Scaled Pearson X2	6	19.9383	3.3231
Log Likelihood		20502.3379	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept			1	5.8550	0.0527	5.7517	5.9583	12342.7	<.0001
admit	yes		1	-2.6756	0.1524	-2.9744	-2.3769	308.10	<.0001
gender	female		1	-0.0897	0.0749	-0.2365	0.0572	1.43	0.2312
dept	a		1	-0.1729	0.0770	-0.3238	-0.0219	5.04	0.0248
dept	b		1	-0.5280	0.0866	-0.6978	-0.3582	37.15	<.0001
dept	c		1	-0.5031	0.0803	-0.6604	-0.3458	39.29	<.0001
dept	d		1	-0.2369	0.0763	-0.3865	-0.0873	9.63	0.0019
dept	e		1	-0.8927	0.0927	-1.0743	-0.7110	92.78	<.0001
dept*admit	a	yes	1	3.2691	0.1671	2.9417	3.5966	382.88	<.0001
dept*admit	b	yes	1	3.2185	0.1749	2.8757	3.5613	338.63	<.0001
dept*admit	c	yes	1	2.0600	0.1674	1.7319	2.3880	151.45	<.0001
dept*admit	d	yes	1	2.0108	0.1699	1.6778	2.3438	140.07	<.0001
dept*admit	e	yes	1	1.5861	0.1798	1.2337	1.9385	77.82	<.0001
dept*gender	a	female	1	-1.9436	0.1268	-2.1921	-1.6950	234.84	<.0001
dept*gender	b	female	1	-3.0194	0.2177	-3.4461	-2.5927	192.34	<.0001
dept*gender	c	female	1	0.6911	0.1019	0.4914	0.8907	46.02	<.0001
dept*gender	d	female	1	-0.0165	0.1033	-0.2190	0.1861	0.03	0.8734
dept*gender	e	female	1	0.8112	0.1157	0.5844	1.0381	49.14	<.0001
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Observation Statistics

(** Reschi: Pearson residual, Resdev: deviance residual, StResdev: standardized deviance residual, StReschi: Standardized Pearson residual)

Observation	count	dept HessWgt Resdev	gender	admit Lower StResdev	Pred Upper StReschi	Xbeta Resraw Reslik	Std Reschi
1	512	a	male	yes	531.43085	6.2755731	0.0424759
		531.43085		488.98011	577.56693	-19.43085	-0.842885
		-0.848101		-4.178759	-4.153057	-4.154119	
2	313	a	male	no	293.56917	5.6821133	0.0561458
		293.56917		262.97775	327.71919	19.430835	1.1340606
		1.1218832		4.1084599	4.1530548	4.1497461	
3	89	a	female	yes	69.569537	4.2423268	0.0992536
		69.569537		57.270933	84.509196	19.430463	2.3295583
		2.2320798		3.9791972	4.1529753	4.0990903	
4	19	a	female	no	38.431098	3.648867	0.1058274
		38.431098		31.232183	47.289339	-19.4311	-3.134411
		-3.477633		-4.607882	-4.153111	-4.417888	
5	353	b	male	yes	354.18803	5.8698279	0.0527164
		354.18803		319.41965	392.74091	-1.188034	-0.063127
		-0.063162		-0.50399	-0.503708	-0.503712	
6	207	b	male	no	205.81197	5.326963	0.0687566
		205.81197		179.86426	235.50296	1.1880342	0.0828121
		0.0827326		0.5032243	0.5037077	0.5036947	
7	17	b	female	yes	15.811966	2.760767	0.202468
		15.811966		10.63273	23.514024	1.1880342	0.2987693
		0.2951405		0.4975897	0.5037077	0.5015638	
8	8	b	female	no	9.1880342	2.217902	0.2072239
		9.1880342		6.1211534	13.791514	-1.188034	-0.391938
		-0.400875		-0.515193	-0.503708	-0.510693	
9	120	c	male	yes	113.99782	4.7361793	0.0713666
		113.99782		99.117242	131.11244	6.0021786	0.5621609
		0.557333		0.8606111	0.8680662	0.8649474	
10	205	c	male	no	211.00218	5.3518685	0.060543
		211.00218		187.3927	237.5862	-6.002179	-0.413205

```
                      -0.415187    -0.872231    -0.868066    -0.869012
 11          202      c       female    yes      208.00218    5.3375486     0.060849
                      208.00218    184.61759    234.34877    -6.002179    -0.416174
                        -0.4182    -0.872292    -0.868066    -0.869039
 12          391      c       female    no       384.99782    5.9532377    0.0476956
                      384.99782     350.6386    422.72392    6.0021786    0.3059002
                      0.3051105    0.8658252    0.8680662    0.8677882
 13          138      d       male      yes      141.63258    4.9532362     0.069663
                      141.63258     123.5566    162.35302    -3.632576    -0.305234
                      -0.306553    -0.548232    -0.545873    -0.546612
 14          279      d       male      no       275.36742    5.6181063    0.0552042
                      275.36742    247.12841    306.83327    3.6325758    0.2189064
                      0.2184277    0.5446796    0.5458732    0.5456815
 15          131      d       female    yes      127.36742     4.847076    0.0715647
                      127.36742    110.69866    146.54614    3.6325758    0.3218737
                      0.3203616    0.5433088    0.5458732     0.544983
 16          244      d       female    no       247.63258    5.5119461    0.0575855
                      247.63258    221.20296    277.22004    -3.632576     -0.23084
                      -0.231408    -0.547216    -0.545873    -0.546114
 17           53      e       male      yes      48.077055     3.872805    0.1016169
                      48.077055    39.395026    58.672463    4.9229452    0.7099963
                      0.6983688    0.9841487    1.0005342     0.992317
 18          138      e       male      no       142.92295    4.9623056    0.0762339
                      142.92295    123.08683    165.95577    -4.922945    -0.411788
                      -0.414187    -1.006362    -1.000534    -1.001524
 19           94      e       female    yes      98.922945    4.5943412     0.087378
                      98.922945    83.352901    117.40142    -4.922945    -0.494967
                       -0.49916     -1.00901    -1.000534    -1.002615
 20          299      e       female    no       294.07705    5.6838418    0.0558617
                      294.07705    263.57946    328.10338    4.9229452    0.2870744
                       0.286279    0.9977619    1.0005342    1.0003062
 21           22      f       male      yes      24.030812    3.1793369    0.1517219
                      24.030812    17.849355    32.352986    -2.030812    -0.414272
                      -0.420323    -0.628804    -0.619753    -0.623813
 22          351      f       male      no       348.96919    5.8549836    0.0527011
                      348.96919    314.72249    386.94245    2.0308123    0.1087117
                      0.1086065    0.6191529    0.6197526    0.6197342
 23           24      f       female    yes      21.969188    3.0896409    0.1525488
                      21.969188     16.29162    29.625366    2.0308123    0.4332742
                      0.4268436    0.6105544    0.6197526    0.6152741
 24          317      f       female    no       319.03081    5.7652877    0.0550363
                      319.03081    286.40832    355.36908    -2.030812    -0.113698
                      -0.113819    -0.620412    -0.619753    -0.619775
```

❸

```
                       The GENMOD Procedure

                         Model Information
                 Data Set                WORK.BERKELEY
                 Distribution                  Poisson
                 Link Function                     Log
                 Dependent Variable              count

               Number of Observations Read          24
               Number of Observations Used          24

                      Class Level Information
         Class       Value              Design Variables
         dept        a           1      0      0      0      0
                     b           0      1      0      0      0
                     c           0      0      1      0      0
                     d           0      0      0      1      0
                     e           0      0      0      0      1
                     f           0      0      0      0      0
         gender      female      1
                     male        0
         admit       no          0
                     yes         1
```

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5	20.2043	4.0409
Scaled Deviance	5	20.2043	4.0409
Pearson Chi-Square	5	18.8242	3.7648
Scaled Pearson X2	5	18.8242	3.7648
Log Likelihood		20503.1035	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept			1	5.8581	0.0527	5.7548	5.9613	12365.7	<.0001
admit	yes		1	-2.7244	0.1577	-3.0335	-2.4153	298.46	<.0001
gender	female		1	-0.0961	0.0751	-0.2434	0.0511	1.64	0.2006
dept	a		1	-0.1686	0.0769	-0.3194	-0.0178	4.80	0.0284
dept	b		1	-0.5284	0.0865	-0.6980	-0.3588	37.29	<.0001
dept	c		1	-0.4839	0.0814	-0.6434	-0.3245	35.38	<.0001
dept	d		1	-0.2240	0.0767	-0.3743	-0.0737	8.53	0.0035
dept	e		1	-0.8791	0.0930	-1.0613	-0.6969	89.45	<.0001
gender*admit	female	yes	1	0.0999	0.0808	-0.0586	0.2583	1.53	0.2167
dept*admit	a	yes	1	3.3065	0.1700	2.9733	3.6396	378.38	<.0001
dept*admit	b	yes	1	3.2631	0.1788	2.9127	3.6135	333.12	<.0001
dept*admit	c	yes	1	2.0439	0.1679	1.7149	2.3729	148.24	<.0001
dept*admit	d	yes	1	2.0119	0.1699	1.6788	2.3449	140.18	<.0001
dept*admit	e	yes	1	1.5672	0.1804	1.2135	1.9208	75.44	<.0001
dept*gender	a	female	1	-2.0023	0.1357	-2.2683	-1.7363	217.68	<.0001
dept*gender	b	female	1	-3.0771	0.2229	-3.5140	-2.6403	190.63	<.0001
dept*gender	c	female	1	0.6628	0.1044	0.4583	0.8674	40.34	<.0001
dept*gender	d	female	1	-0.0440	0.1057	-0.2512	0.1632	0.17	0.6774
dept*gender	e	female	1	0.7929	0.1167	0.5642	1.0215	46.19	<.0001
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Observation Statistics

(** Reschi: Pearson residual, Resdev: deviance residual, StResdev: standardized deviance residual, StReschi: Standardized Pearson residual)

Observation	count	dept / HessWgt / Resdev	gender / Lower / StResdev	admit	Pred / Upper / StReschi	Xbeta / Resraw / Reslik	Std / Reschi
1	512	a	male	yes	529.26988	6.2714985	0.0427054
		529.26988	486.77278		575.47713	-17.26988	-0.750673
		-0.754812	-4.049477		-4.027272	-4.028046	
2	313	a	male	no	295.73014	5.6894473	0.0563134
		295.73014	264.82656		330.23996	17.269864	1.0042483
		0.9947045	3.9889963		4.0272693	4.0249	
3	89	a	female	yes	71.730325	4.2729136	0.1018191
		71.730325	58.753561		87.573236	17.269675	2.039073
		1.9645109	3.879963		4.0272253	3.9899909	
4	19	a	female	no	36.270065	3.5909927	0.1165867
		36.270065	28.860876		45.581347	-17.27007	-2.867608
		-3.157703	-4.43473		-4.027316	-4.238772	
5	353	b	male	yes	353.63951	5.8682781	0.052782
		353.63951	318.88392		392.18316	-0.639511	-0.034007
6	207	b	male	no	206.36049	5.3296246	0.0687252
		206.36049	180.35475		236.11605	0.6395108	0.0445179
		0.0444949	0.2795786		0.2797229	0.2797192	
7	17	b	female	yes	16.360489	2.7948692	0.2039494
		16.360489	10.969688		24.400476	0.6395107	0.1581065
		0.1570929	0.2779296		0.2797228	0.2791512	
8	8	b	female	no	8.6395115	2.156346	0.2138246
		8.6395115	5.6817404		13.137024	-0.639512	-0.217572
		-0.220343	-0.283285		-0.279723	-0.281884	
9	120	c	male	yes	109.2453	4.6935958	0.0800879
		109.2453	93.375254		127.81261	10.754703	1.0289566
		1.0127309	1.8511689		1.8808278	1.8720004	
10	205	c	male	no	215.75471	5.3741422	0.0627097
		215.75471	190.80147		243.97136	-10.75471	-0.732181
		-0.738394	-1.896788		-1.880829	-1.883256	
11	202	c	female	yes	212.75471	5.3601399	0.0630707

```
                    212.75471    188.01536     240.7493    -10.75471    -0.737325
                    -0.743672    -1.897018    -1.880829    -1.883326
  12        391     c       female    no       380.24529    5.9408166     0.049028
                    380.24529      345.407    418.59743    10.754707    0.5515268
                    0.5489572    1.8720652    1.8808285    1.8800765
  13        138     d        male     yes      137.20741    4.9214937     0.0749396
                    137.20741    118.46468    158.91549    0.7925914    0.0676645
                    0.0675995    0.1411229    0.1412585    0.1412274
  14        279     d        male     no       279.79259    5.6340486     0.0563198
                    279.79259    250.55134    312.44653    -0.792595    -0.047384
                    -0.047407    -0.141326    -0.141259    -0.141267
  15        131     d       female    yes       131.7926    4.8812294     0.0759943
                     131.7926    113.55454    152.95987    -0.792595    -0.069041
                     -0.06911    -0.141401    -0.141259    -0.141293
  16        244     d       female    no       243.20741    5.4939146     0.0598286
                    243.20741    216.29703    273.46581    0.7925928    0.0508232
                    0.0507956    0.1411822    0.1412588    0.1412489
  17         53     e        male     yes       45.68082    3.8216785     0.1107698
                     45.68082     36.76601    56.757241    7.3191799    1.0829178
                    1.0557746    1.5925468    1.6334899    1.6156233
  18        138     e        male     no       145.31918    4.9789326      0.077011
                    145.31918    124.96001    168.99539    -7.319183    -0.607157
                    -0.612364    -1.647499    -1.633491    -1.635433
  19         94     e       female    yes      101.31918    4.6182758     0.0889611
                    101.31918     85.10751    120.61893    -7.319183    -0.727138
                    -0.736167    -1.653774    -1.633491     -1.63753
  20        299     e       female    no       291.68082    5.6756601     0.0565015
                    291.68082    261.10408    325.83825    7.3191827    0.4285571
                    0.4267832    1.6267294    1.6334905    1.6330261
  21         22     f        male     yes        22.9571    3.1336273     0.1567901
                      22.9571    16.883288    31.215985      -0.9571    -0.199756
                    -0.201168    -0.304786    -0.302645     -0.30358
  22        351     f        male     no        350.0429    5.8580557     0.0526799
                     350.0429    315.70399    388.11683    0.9570993     0.051156
                    0.0511327    0.3025072     0.302645     0.302641
  23         24     f       female    yes      23.042901    3.1373577     0.1567227
                    23.042901    16.948628    31.328511     0.957099    0.1993831
                    0.1980262    0.3005852    0.3026449    0.3017527
  24        317     f       female    no        317.9571    5.7619165      0.055192
                     317.9571    285.35734    354.28112    -0.957099    -0.053675
                    -0.053702    -0.302797    -0.302645     -0.30265
```

❹

```
                         The GENMOD Procedure
                           Model Information
                     Data Set               WORK.BCDEF
                     Distribution              Poisson
                     Link Function                 Log
                     Dependent Variable          count
                 Number of Observations Read          20
                 Number of Observations Used          20

                   Class Level Information
              Class      Value          Design Variables
              dept       b           1      0      0      0
                         c           0      1      0      0
                         d           0      0      1      0
                         e           0      0      0      1
                         f           0      0      0      0
              gender     female      1
                         male        0
              admit      no          0
                         yes         1

               Criteria For Assessing Goodness Of Fit
       Criterion                 DF           Value        Value/DF
       Deviance                   5          2.6815          0.5363
       Scaled Deviance            5          2.6815          0.5363
       Pearson Chi-Square         5          2.6904          0.5381
       Scaled Pearson X2          5          2.6904          0.5381
       Log Likelihood                    15996.8481
Algorithm converged.
```

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept			1	5.8550	0.0527	5.7517	5.9583	12342.7	<.0001
admit	yes		1	-2.6756	0.1524	-2.9744	-2.3769	308.10	<.0001
gender	female		1	-0.0897	0.0749	-0.2365	0.0572	1.43	0.2312
dept	b		1	-0.5280	0.0866	-0.6978	-0.3582	37.15	<.0001
dept	c		1	-0.5031	0.0803	-0.6604	-0.3458	39.29	<.0001
dept	d		1	-0.2369	0.0763	-0.3865	-0.0873	9.63	0.0019
dept	e		1	-0.8927	0.0927	-1.0743	-0.7110	92.78	<.0001
dept*admit	b	yes	1	3.2185	0.1749	2.8757	3.5613	338.63	<.0001
dept*admit	c	yes	1	2.0600	0.1674	1.7319	2.3880	151.44	<.0001
dept*admit	d	yes	1	2.0108	0.1699	1.6778	2.3438	140.07	<.0001
dept*admit	e	yes	1	1.5861	0.1798	1.2337	1.9385	77.82	<.0001
dept*gender	b	female	1	-3.0194	0.2177	-3.4461	-2.5926	192.34	<.0001
dept*gender	c	female	1	0.6911	0.1019	0.4914	0.8907	46.02	<.0001
dept*gender	d	female	1	-0.0165	0.1033	-0.2190	0.1861	0.03	0.8734
dept*gender	e	female	1	0.8112	0.1157	0.5844	1.0381	49.13	<.0001
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

❺ The GENMOD Procedure

Model Information

Data Set	WORK.BCDEF
Distribution	Poisson
Link Function	Log
Dependent Variable	count
Number of Observations Read	20
Number of Observations Used	20

Class Level Information

Class	Value	Design Variables			
dept	b	1	0	0	0
	c	0	1	0	0
	d	0	0	1	0
	e	0	0	0	1
	f	0	0	0	0
gender	female	1			
	male	0			
admit	no	0			
	yes	1			

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	2.5564	0.6391
Scaled Deviance	4	2.5564	0.6391
Pearson Chi-Square	4	2.5582	0.6395
Scaled Pearson X2	4	2.5582	0.6395
Log Likelihood		15996.9106	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept			1	5.8540	0.0528	5.7506	5.9575	12295.4	<.0001
admit	yes		1	-2.6611	0.1578	-2.9704	-2.3518	284.33	<.0001
gender	female		1	-0.0877	0.0751	-0.2350	0.0595	1.36	0.2430
dept	b		1	-0.5279	0.0867	-0.6978	-0.3581	37.11	<.0001
dept	c		1	-0.5092	0.0822	-0.6703	-0.3481	38.37	<.0001
dept	d		1	-0.2409	0.0773	-0.3923	-0.0895	9.72	0.0018
dept	e		1	-0.8970	0.0936	-1.0804	-0.7135	91.86	<.0001
gender*admit	female	yes	1	-0.0307	0.0868	-0.2007	0.1394	0.13	0.7235
dept*admit	b	yes	1	3.2053	0.1788	2.8548	3.5557	321.37	<.0001
dept*admit	c	yes	1	2.0652	0.1681	1.7358	2.3946	151.00	<.0001
dept*admit	d	yes	1	2.0107	0.1699	1.6777	2.3437	140.06	<.0001
dept*admit	e	yes	1	1.5922	0.1806	1.2382	1.9462	77.70	<.0001
dept*gender	b	female	1	-3.0020	0.2231	-3.4393	-2.5647	181.06	<.0001
dept*gender	c	female	1	0.6999	0.1049	0.4943	0.9055	44.52	<.0001
dept*gender	d	female	1	-0.0080	0.1061	-0.2159	0.1998	0.01	0.9397
dept*gender	e	female	1	0.8170	0.1169	0.5879	1.0461	48.85	<.0001
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Inference about conditional Association

ex) Model (AD, AG, DG)

$$\log\mu_{ijk} = \lambda + \lambda_i^G + \lambda_j^A + \lambda_k^D + \lambda_{ij}^{GA} + \lambda_{ik}^{GD} + \lambda_{jk}^{AD}$$

$H_0 : \lambda_{ij}^{GA} = 0$ (A cond. indep. of G, given D)

$$\begin{aligned} Likelihood\ ratio\ stat. &= -2(L_0 - L_1) \\ &= Deviance \text{ for } (AD,\ DG) - Deviance \text{ for } (AD,\ AG,\ DG) \\ &= 21.7 - 20.3 = 1.5\ with\ df = 6-5 = 1 (p-value = 0.21) \end{aligned}$$

H_0 is plausible, but test is shaky because model $(AD,\ AG,\ DG)$ fits poorly.

Recall $\hat{\theta}_{AG(D)} = \exp(\hat{\lambda}_{11}^{AG}) = \exp(0.0999) = 1.105$
95% C.I. for $\theta_{AG(D)}$ is

$$\exp[0.0999 \pm 1.96(0.0808)] = (0.943,\ 1.295)$$

(contains 1) Plausible that $\theta_{AG(D)} = 1$

There are equivalences between loglinear models and corresponding logit models that treat one of the variables as a response variable, others as explanatory.

ex) Model $(XY,\ XZ,\ YZ)$

Suppose Y is binary

Construct the logit

$$\begin{aligned} \log\left[\frac{P(Y=1 \mid X=i, Z=k)}{P(Y=2 \mid X=i, Z=k)}\right] &= \log\left(\frac{\pi_{i1k}}{\pi_{i2k}}\right) = \log\left(\frac{\mu_{i1k}}{\mu_{i2k}}\right) \\ &= \log\mu_{i1k} - \log\mu_{i2k} \\ &= (\lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ} \\ &\quad - (\lambda + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ}) \\ &= (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ}) \\ &= \alpha + \beta_i^X + \beta_k^Z \end{aligned}$$

Note:

- Loglinear models extend to any no. of dimensions
- Loglinear models treat all variables symmetrically; Logistic regression models treat Y as response and other variables as explanatory. Logistic regression is the more natural approach when one has a single response variable.(eg., graduate admission) See output for logit analysis of data.

ex) Auto accidents.

$G = gender(F,\ M)$, $L = location(urban,\ rural)$,

$S = seat\ belt\ use(No,\ Yes)$, $I = injury(No,\ Yes)$ I is natural response variable.

Loglinear model $(GLS,\ IG,\ IL,\ IS)$ fit quite well $(G^2 = 7.5,\ df = 4)$. Simpler to consider logit model with I as response.

$$\mathrm{logit}\hat{P}(I = Yes) = -3.34 + 0.54G + 0.76L + 0.82S$$

Controlling for other variables, estimated odds of injury are:

$e^{0.54} = 1.72$ times higher for females than males(C.I.:(1.63, 1.82))

$e^{0.76} = 2.13$ times higher in rural than urban locations(C.I.:(2.02, 2.25))

$e^{0.82} = 2.26$ times higher when not wearing seat belt(C.I.:(2.14, 2.39))

Why ever use loglinear model for contingency table?

Information about all associations, not merely effects of explanatory variables on response.

ex) Auto accident data

Loglinear model $(GI,\ GL,\ GS,\ IL,\ IS,\ LS)$ $(G^2 = 23.4,\ df = 5)$ fits almost as well as $(GLS,\ GI,\ IL,\ IS)(G^2 = 7.5,\ df = 4)$ in practical terms but n is huge (68,694). In model $(GI,\ GL,\ GS,\ IL,\ IS,\ LS)$, the estimated parameters are

Variables	$\hat{\lambda}$	odds ratio($\hat{\theta}$)	$1/\hat{\theta}$
GL	-0.21	0.81(fem. rur.)	1.23(=1/0.81)
GS	-0.46	0.63(fem. no)	1.58
GI	-0.54	0.58(fem. no)	1.72
LS	-0.08	0.92(rur. no)	1.09
LI	-0.75	0.47(rur. no)	2.13
SI	-0.81	0.44(no no)	2.26

where S: seat belt, I: injury

eg., for those not wearing seat belts, the estimated odds of being injured are 2.26 times the estimated odds or injury for those wearing seat belts, controlling for gender and location.(or interchanges S and I in terp.)

Dissimilarity Index

$$D = \sum |\ p_i - \hat{\pi}_i\ | /2 = \sum |\ n_i - \hat{\mu}_i\ | /2n$$

- A goodness-of-fit test should not be the sole criterion for selecting a model,
- $0 \le D \le 1$, with smaller values for better fit.
- $D =$ Proportion of sample cases that must move to different cells for model to

fit perfectly.

- The dissimilarity index helps indicate whether the lack of fit is important in a practical sense.

ex) Loglinear model(GLS, IG, IL, IS) has $D = 0.003$

Simpler model (GL, GS, LS, IG, IL, IS) has $G^2 = 23.4(df = 5)$ for testing fit ($p - value < 0.001$), but $D = 0.008$ (Good fit for practical purposes, and simpler to interpret GS, LS associations)

For large n, effects can be "statistically significant" without being "practically significant".

Model can fail good-of-fit test but still be adequate for practical purpose.

dissimilarity-index SAS program

```
data injury;
  input G L S I count @@;
cards;
0 0 0 0  7287 0 0 0 1   996 0 0 1 0 11587 0 0 1 1   759
0 1 0 0  3246 0 1 0 1   973 0 1 1 0  6134 0 1 1 1   757
1 0 0 0 10381 1 0 0 1   812 1 0 1 0 10969 1 0 1 1   380
1 1 0 0  6123 1 1 0 1  1084 1 1 1 0  6693 1 1 1 1   513
run;

ods listing close;
proc catmod data=injury;
  weight count;
  model G*I*L*S= _response_/ pred=freq;
  loglin g|i g|l g|s i|l i|s l|s;
  ods output predictedfreqs=temp1 anova=temp;
run;
quit;
data temp1 (keep=p1 functionnum);
  set temp1;
  rename predfunction=p1;
run;
proc catmod data=injury;
  weight count;
  model G*I*L*S= _response_/ pred=freq;
  loglin g|l|s g|i i|l i|s;
  ods output predictedfreqs=temp2 anova=temp3;
run;
quit;
ods output close;
ods listing;
data temp2;
  set temp2;
  rename predfunction=p2;
run;
data combo;
  merge temp1 temp2;
  by functionnum;
  Male=G+0;
  Location=L+0;
  Seat=S+0;
  Injury=I+0;
  rename obsfunction = observed;
run;
proc format;
  value male 0='Female' 1='Male';
  value location 0='Urban' 1='Rural';
  value Yesno 0='No' 1='Yes';
run;
proc sort data=combo;
  by male location seat injury;
run;
proc print data= combo;
  format male male. location location. seat yesno. injury yesno.;
  var male location seat injury observed p1 p2;
run;

proc sql;
  select  sum( abs(observed-p1) ) / (2*sum(observed)) as d1,
          sum( abs(observed-p2) ) /( 2*sum(observed) ) as d2
  from combo;
quit;
```

Obs	Male	Location	Seat	Injury	observed	p1	p2
1	Female	Urban	No	No	7287	7166.369	7273.214
2	Female	Urban	No	Yes	996	993.0169	1009.786
3	Female	Urban	Yes	No	11587	11748.31	11632.62
4	Female	Urban	Yes	Yes	759	721.3055	713.3779
5	Female	Rural	No	No	3246	3353.829	3254.662
6	Female	Rural	No	Yes	973	988.7848	964.3382
7	Female	Rural	Yes	No	6134	5985.493	6093.502
8	Female	Rural	Yes	Yes	757	781.8927	797.4979
9	Male	Urban	No	No	10381	10471.5	10358.93
10	Male	Urban	No	Yes	812	845.1187	834.0683
11	Male	Urban	Yes	No	10969	10837.83	10959.23
12	Male	Urban	Yes	Yes	380	387.5588	389.7677
13	Male	Rural	No	No	6123	6045.306	6150.192
14	Male	Rural	No	Yes	1084	1038.08	1056.808
15	Male	Rural	Yes	No	6693	6811.371	6697.644
16	Male	Rural	Yes	Yes	513	518.2429	508.3564

d1	d2
0.008219	0.002507

ex) Death Penalty for multiple homicide indictments in Florida.

Victim's Race(V)	Suspect's Race(D)	Death Penalty(P) Yes	No	Prop. Yes
White	White	53 (52.8)	414 (414.2)	0.113
	Black	11 (11.2)	37 (36.8)	0.229
Black	White	0 (0.2)	16 (15.8)	0.000
	Black	4 (3.8)	139 (139.2)	0.029

where ( )=fitted mean on $(DP,\ VP,\ VD)$

Goodness of fit

Model	G^2	df	$p-value$
$(D,\ V,\ P)$	402.8	4	0.000
$(DV,\ P)$	22.3	3	0.000
$(DV,\ PV)$	5.4	2	0.070
$(DV,\ DP)$	20.7	2	0.000
(DP, VP, DV)	0.4	1(=8-7)	0.540

Compare these two models($(DV,\ PV)$, $(DP,\ VP,\ DV)$)

$$G^2 = 5.4 - 0.4 = 5.0,\ df = 1$$

So $(DP,\ VP,\ DV)$ is better

Note:

- G^2 is monotone decreasing as model becomes more complex
- $(DP,\ VP,\ VD)$ seems to fit well
- Fitted tables can be described by odds ratios
  eg) for model $(DP,\ VP,\ DV)$

$$\hat{\theta}_{D(V)P} = \frac{52.8\times 36.8}{11.2\times 414.2}$$ table for Victim's race=w

fixed V

$$= \frac{0.2\times 139.2}{15.8\times 3.8}$$ table for Victim's race=B

$$= 0.42 = e^{-0.868}$$ $\hat{\lambda}_{11}^{SP}$

$$= e^{\hat{\lambda}_{11}^{DP}+\hat{\lambda}_{22}^{DP}-\hat{\lambda}_{12}^{DP}-\hat{\lambda}_{21}^{DP}}$$

0

difference

For Marginal DP table,

$$\hat{\theta}_{SP} = \frac{(52.8+0.2)(36.8+139.2)}{(414.2+15.8)(11.2+3.8)} = \frac{53\times 176}{430\times 15} = 1.45$$

In likelihood, suff. stat for $(DP,\ PV,\ DV)$
$n_{ij+},\ n_{+jk},\ n_{i+k}$

Estimated odds of death penalty are 45% higher for white suspect than black suspect; 42% as high for white suspect as black suspect controlling for V (Simpson's Paradox)

Note:

- Fitted marginal odds ratios=sample values, since likelihood equations for this model

$$\hat{\mu}_{dv+} = n_{dv+},\ \hat{\mu}_{d+p} = n_{d+p},\ \hat{\mu}_{+vp} = n_{+vp},\ all\ d,\ v,\ p.$$

- Fitted conditional odds ratios identical at each level of third variable ($(PD,\ PV,\ DV)$)

Corresponding Logit Models

Treat P as response variable, D, V as explanatory.

$$\log\left[\frac{P(P=1\mid D=i, V=j)}{P(P=2\mid D=i, V=j)}\right] = \log\left(\frac{\mu_{ij1}}{\mu_{ij2}}\right) = \alpha$$ corresponds to loglinear model (DV,P)

$G^2 = 22.3,\ df = 4-1=3$

$$= \alpha + \beta_j^V$$ corresponds to (PV,DV)

$G^2 = 5.4,\ df = 2$

$= \alpha + \beta_i^D$ corresponds to (PD, DV)
$G^2 = 20.7,\ df = 2$

$= \alpha + \beta_i^D + \beta_j^V$ corresponds to (PD, PV, DV)
$G^2 = 0.4,\ df = 1 = 4 - 3$

For this logit model from (PD, PV, DV),

Para.	Est.	S.E.
β_1^D	-0.868	0.367
β_1^V	2.404	0.601

$e^{2.404} = 11.1$. Controlling for defendant's race, odds of death penalty estimated to be 11.1 times higher when victims were white than when victims were black. 95% C.I. for $\theta_{(S)VP}$ is $\exp[2.404 \pm 1.96(0.601)] = (3.4,\ 35.9)$

```
data loglin;
 input d v p count;
cards;
1 1 1 53
1 1 0 414
1 0 1 0
1 0 0 16
0 1 1 11
0 1 0 37
0 0 1 4
0 0 0 139
run;

/*(DP, VP, DV)*/
proc genmod data=loglin order=data;
 class d v p;
 model count = d v p d*v d*p v*p /dist=poi link=log lrci obstats;
run;

/*(DV, PV)*/
proc genmod data=loglin order=data;
 class d v p;
 model count = d v p d*v v*p /dist=poi link=log lrci obstats;
run;
```

$d=\begin{cases}1 & white\\0 & black\end{cases},\quad v=\begin{cases}1 & white\\0 & black\end{cases}$

```
                 The GENMOD Procedure

                  Model Information
           Data Set                WORK.LOGLIN
           Distribution                Poisson
           Link Function                   Log
           Dependent Variable            count

        Number of Observations Read           8
        Number of Observations Used           8

              Class Level Information
              Class      Levels    Values
              d               2    1 0
              v               2    1 0
              p               2    1 0


               Parameter Information
        Parameter       Effect       d    v    p
        Prm1            Intercept
        Prm2            d            1
        Prm3            d            0
        Prm4            v                 1
        Prm5            v                 0
        Prm6            p                      1
        Prm7            p                      0
        Prm8            d*v          1    1
        Prm9            d*v          1    0
        Prm10           d*v          0    1
        Prm11           d*v          0    0
        Prm12           d*p          1         1
        Prm13           d*p          1         0
        Prm14           d*p          0         1
        Prm15           d*p          0         0
        Prm16           v*p               1    1
        Prm17           v*p               1    0
```

Prml8	v*p	0	1
Prml9	v*p	0	0

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1	0.3798	0.3798
Scaled Deviance	1	0.3798	0.3798
Pearson Chi-Square	1	0.1978	0.1978
Scaled Pearson X2	1	0.1978	0.1978
Log Likelihood		2926.7234	
Full Log Likelihood		-19.2098	
AIC (smaller is better)		52.4197	
AICC (smaller is better)		.	
BIC (smaller is better)		52.9758	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	4.9358	0.0847	4.7650	5.0974	3394.84	<.0001
d	1		1	-2.1746	0.2638	-2.7308	-1.6899	67.97	<.0001
d	0		0	0.0000	0.0000	0.0000	0.0000	.	.
v	1		1	-1.3298	0.1848	-1.7059	-0.9793	51.79	<.0001
v	0		0	0.0000	0.0000	0.0000	0.0000	.	.
p	1		1	-3.5961	0.5069	-4.7754	-2.7349	50.33	<.0001
p	0		0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	1	1	1	4.5950	0.3135	4.0080	5.2421	214.78	<.0001
d*v	1	0	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	0	1	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	0	0	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	1	1	1	-0.8678	0.3671	-1.5633	-0.1140	5.59	0.0181
d*p	1	0	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	0	1	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	0	0	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	1	1	1	2.4044	0.6006	1.3068	3.7175	16.03	<.0001
v*p	1	0	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	0	1	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	0	0	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Observation Statistics

```
                                                          Standard
                                                          Error of
                                                               the
                                   Predicted       Linear     Linear
Observation        count  d  v  p      Value    Predictor  Predictor     HessWgt        Lower
                                                                             Std          Std
                                         Raw      Pearson   Deviance    Deviance      Pearson
                             Upper  Residual     Residual   Residual    Residual     Residual
                        Likelihood                          DFBETA_
                          Residual  Leverage        CookD  Intercept  DFBETA_dl   DFBETA_vl
                                    DFBETA_      DFBETA_    DFBETA_   DFBETAS_    DFBETAS_
                        DFBETA_pl      dlvl         dlpl       vlpl  Intercept          dl
                         DFBETAS_  DFBETAS_     DFBETAS_   DFBETAS_   DFBETAS_
                               vl        pl         dlvl       dlpl       vlpl
          1           53  1  1  1  52.817821  3.9668486  0.1373785  52.817821  40.350013
                        69.138073  0.1821794  0.0250674   0.025053  0.4444776  0.4447329
                         0.4447321   0.996823  8.8654137  0.0013089  -0.012826  -0.006257
                        -0.049027  0.0182143  1.0605444  0.0702672  0.0154515  -0.048627
                         -0.03386  -0.096717  0.0580937  2.8891826  0.1169918
          2          414  1  1  0  414.18218  6.0263059  0.0491266  414.18218   376.1619
                        456.04532  -0.182179  -0.008952  -0.008952  -0.444765  -0.444733
                         -0.444733  0.9995949  69.713343  0.0013089  -0.012826  -0.006257
                        -0.049027  -1.067459  1.0605444  0.0702672  0.0154515  -0.048627
                         -0.03386  -0.096717  -3.404624  2.8891826  0.1169918
          3            0  1  0  1  0.1821794  -1.702763  0.6581489  0.1821794  0.0501513
                        0.6617845  -0.182179  -0.426825  -0.603621  -0.628947  -0.444733
                         -0.616414  0.0789128  0.0024207  0.0013089  -0.012826  -0.006257
                        -0.049027  0.0182143  -0.025129  0.0702672  0.0154515  -0.048627
                         -0.03386  -0.096717  0.0580937  -0.068458  0.1169918
          4           16  1  0  0  15.817821  2.7611372  0.2500983  15.817821  9.6886345
                        25.824428  0.1821794  0.0458064  0.0457189  0.4438833  0.4447329
                         0.4447239  0.9893915  2.6352105  0.0013089  1.0728473  -0.006257
                        -0.049027  -1.067459  -0.025129  0.0702672  0.0154515  4.0674093
                         -0.03386  -0.096717  -3.404624  -0.068458  0.1169918
          5           11  0  1  1  11.182179  2.4143214  0.2967929  11.182179  6.2502284
                        20.005851  -0.182179   -0.05448  -0.054629  -0.445949  -0.444733
                         -0.444751  0.9849937  1.8546433  0.0013089  -0.012826  -0.006257
                        -0.049027  0.0182143  1.0605444  -1.015406  0.0154515  -0.048627
                         -0.03386  -0.096717  0.0580937  2.8891826  -1.690608
          6           37  0  1  0  36.817821   3.605982  0.1644292  36.817821  26.674462
                        50.818342  0.1821794  0.0300241  0.0299994  0.4443669  0.4447329
```

```
                0.4447312  0.9954423  6.1712718  0.0013089  -0.012826  1.0794165
               -0.049027  -1.067459  1.0605444  -1.015406  0.0154515  -0.048627
               5.8412888  -0.096717  -3.404624  2.8891826  -1.690608
7          4   0  0  1  3.8178206  1.3396797  0.5004169  3.8178206  1.4317259
               10.180548  0.1821794  0.0932377  0.0925105  0.4412644  0.4447329
                 0.444581  0.9560474   0.614604  0.0013089  -0.012826  -0.006257
               1.0366464  0.0182143  -0.025129  -1.015406  0.0154515  -0.048627
                -0.03386  2.0450151  0.0580937  -0.068458  -1.690608
8        139   0  0  0  139.18218  4.9357837  0.0847123  139.18218  117.88985
               164.32015  -0.182179  -0.015442  -0.015446   -0.44483  -0.444733
                -0.444733  0.9987944  23.407779  -1.084365  1.0728473  1.0794165
               1.0366465  -1.067459  -0.025129  -1.015406  -12.80056  4.0674093
               5.8412888  2.0450151  -3.404624  -0.068458  -1.690608
```

The GENMOD Procedure

Model Information

Data Set	WORK.LOGLIN
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	8
Number of Observations Used	8

Class Level Information

Class	Levels	Values
d	2	1 0
v	2	1 0
p	2	1 0

Parameter Information

Parameter	Effect	d	v	p
Prm1	Intercept			
Prm2	d	1		
Prm3	d	0		
Prm4	v		1	
Prm5	v		0	
Prm6	p			1
Prm7	p			0
Prm8	d*v	1	1	
Prm9	d*v	1	0	
Prm10	d*v	0	1	
Prm11	d*v	0	0	
Prm12	v*p		1	1
Prm13	v*p		1	0
Prm14	v*p		0	1
Prm15	v*p		0	0

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	5.3940	2.6970
Scaled Deviance	2	5.3940	2.6970
Pearson Chi-Square	2	5.8109	2.9054
Scaled Pearson X2	2	5.8109	2.9054
Log Likelihood		2924.2162	

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Full Log Likelihood		-21.7170	
AIC (smaller is better)		55.4339	
AICC (smaller is better)		139.4339	
BIC (smaller is better)		55.9106	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	4.9374	0.0846	4.7669	5.0987	3406.92	<.0001
d	1		1	-2.1903	0.2636	-2.7462	-1.7058	69.03	<.0001
d	0		0	0.0000	0.0000	0.0000	0.0000	.	.
v	1		1	-1.1989	0.1681	-1.5379	-0.8773	50.85	<.0001
v	0		0	0.0000	0.0000	0.0000	0.0000	.	.
p	1		1	-3.6571	0.5064	-4.8357	-2.7972	52.15	<.0001
p	0		0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	1	1	1	4.4654	0.3041	3.8966	5.0940	215.64	<.0001
d*v	1	0	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	0	1	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	0	0	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	1	1	1	1.7045	0.5237	0.7995	2.9072	10.59	0.0011
v*p	1	0	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	0	1	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	0	0	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Observation Statistics

Observation	count	d	v	p	Predicted Value	Linear Predictor	Standard Error of the Linear Predictor	HessWgt	Lower	Upper
1	53	1	1	1	58.034956	4.0610455	0.1257958	58.034956	45.353635	74.2621
2	414	1	1	0	408.96505	6.0136297	0.0491617	408.96505	371.39808	450.33191
3	0	1	0	1	0.4025249	-0.909998	0.5533579	0.4025249	0.1360735	1.1907262
4	16	1	0	0	15.597487	2.7471098	0.2503243	15.597487	9.5494464	25.47599
5	11	0	1	1	5.9650509	1.7859176	0.1857864	5.9650509	4.1445035	8.5853065
6	37	0	1	0	42.034965	3.7385018	0.1452889	42.034965	31.618446	55.883146
7	4	0	0	1	3.597566	1.2802575	0.5006975	3.597566	1.3483861	9.598498
8	139	0	0	0	139.40251	4.9373655	0.0845891	139.40251	118.10499	164.54055

Observation	Raw Residual	Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual	Leverage	CookD
1	-5.034956	-0.660923	-0.670843	-2.348113	-2.313392	-2.316245	0.9183787	10.036111
2	5.0349527	0.248973	0.2484648	2.3086673	2.31339	2.3133353	0.9884174	76.116692
3	-0.402525	-0.634449	-0.897246	-0.958241	-0.677579	-0.928245	0.1232551	0.0107572
4	0.4025129	0.1019184	0.1014847	0.6746752	0.6775585	0.6774934	0.9773738	3.3051537
5	5.0349491	2.0615221	1.8421756	2.0672431	2.3133883	2.12026	0.2058931	0.2312644
6	-5.034965	-0.776589	-0.792921	-2.362049	-2.313395	-2.318929	0.8873108	7.0233097
7	0.402434	0.2121731	0.2083911	0.6653505	0.6774257	0.6762507	0.9019027	0.7031929
8	-0.402515	-0.034092	-0.034108	-0.677888	-0.677561	-0.677562	0.9974684	30.147391

Observation	DFBETA_ Intercept	DFBETA_dl	DFBETA_vl	DFBETA_pl	DFBETA_ dlvl	DFBETA_ vlpl
1	5.35E-17	-3.02E-16	0.1197802	-3.27E-15	-0.132092	-0.963856
2	3.77E-16	7.541E-16	0.1197801	-1.1E-15	0.9308326	-0.963856
3	0.0028875	-0.028695	-0.002888	-0.114776	0.0286945	0.1147756
4	0.0028874	1.1118549	-0.002887	-0.114772	-1.111855	0.1147722
5	3.3E-17	-2.33E-16	0.11978	2.53E-16	-0.132091	0.0990686
6	-3.1E-16	1.783E-15	-0.943146	6.976E-16	0.9308348	0.0990689
7	0.0028868	-0.028688	-0.002887	1.0255753	0.0286881	-1.025575
8	-1.137666	1.1118596	1.137666	1.0257807	-1.11186	-1.025781

Observation	DFBETAS_ Intercept	DFBETAS_ dl	DFBETAS_ vl	DFBETAS_ pl	DFBETAS_ dlvl	DFBETAS_ vlpl
1	6.325E-16	-1.14E-15	0.7124703	-6.45E-15	-0.434391	-1.840385
2	4.457E-15	2.861E-15	0.7124698	-2.17E-15	3.0610949	-1.840384
3	0.0341356	-0.10885	-0.017175	-0.226648	0.0943636	0.2191523
4	0.0341346	4.2177198	-0.017175	-0.226641	-3.656397	0.2191458
5	3.901E-16	-8.84E-16	0.7124693	4.995E-16	-0.43439	0.1891614
6	-3.67E-15	6.762E-15	-5.609974	1.377E-15	3.061102	0.189162
7	0.0341279	-0.108826	-0.017171	2.0252081	0.0943423	-1.958231
8	-13.44933	4.2177375	6.7670055	2.0256137	-3.656412	-1.958623

SAS loglinear modeling of death penalty data

$$d=\begin{cases} w & (0) \\ b & (1) \end{cases},\quad p=\begin{cases} yes & (0) \\ no & (1) \end{cases},\quad v=\begin{cases} w & (0) \\ b & (1) \end{cases}$$

```
data loglin;
 input d$ v$ p$ count @@;
cards;
w w yes 53 w w no 414 w b yes 0 w b no 16
b w yes 11 b w no   37 b b yes 4 b b no 139
;
run;
  /*(DP, VP, DV)*/
❶ proc genmod data=loglin;
    class d v p;
    model count=d v p d*v d*p v*p/dist=poi link=log lrci obstats;
   run;

  /*(DV, PV)*/
❷ proc genmod data=loglin;
    class d v p;
    model count=d v p d*v d*p v*p/dist=poi link=log lrci obstats;
   run;
```

```
                    The GENMOD Procedure

                     Model Information
          Data Set                    WORK.LOGLIN
          Distribution                    Poisson
          Link Function                       Log
          Dependent Variable                count

       Number of Observations Read            8
       Number of Observations Used            8

                  Class Level Information
             Class      Levels    Values
             d               2    b w
             v               2    b w
             p               2    no yes

                   Parameter Information
      Parameter         Effect        d    v    p
      Prm1              Intercept
      Prm2              d             b
      Prm3              d             w
      Prm4              v                  b
      Prm5              v                  w
      Prm6              p                       no
      Prm7              p                       yes
      Prm8              d*v           b    b
      Prm9              d*v           b    w
      Prm10             d*v           w    b
      Prm11             d*v           w    w
      Prm12             d*p           b         no
      Prm13             d*p           b         yes
      Prm14             d*p           w         no
      Prm15             d*p           w         yes
      Prm16             v*p                b    no
      Prm17             v*p                b    yes
      Prm18             v*p                w    no
      Prm19             v*p                w    yes
```

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1	0.3798	0.3798
Scaled Deviance	1	0.3798	0.3798
Pearson Chi-Square	1	0.1978	0.1978
Scaled Pearson X2	1	0.1978	0.1978
Log Likelihood		2926.7234	
Full Log Likelihood		-19.2098	
AIC (smaller is better)		52.4197	
AICC (smaller is better)		.	
BIC (smaller is better)		52.9758	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	3.9668	0.1374	3.6850	4.2245	833.78	<.0001
d	b		1	-1.5525	0.3262	-2.2399	-0.9504	22.66	<.0001
d	w		0	0.0000	0.0000	0.0000	0.0000	.	.
v	b		1	-5.6696	0.6459	-7.0608	-4.4854	77.06	<.0001
v	w		0	0.0000	0.0000	0.0000	0.0000	.	.
p	no		1	2.0595	0.1458	1.7836	2.3565	199.40	<.0001
p	yes		0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	b	b	1	4.5950	0.3135	4.0080	5.2421	214.78	<.0001
d*v	b	w	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	w	b	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	w	w	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	b	no	1	-0.8678	0.3671	-1.5633	-0.1140	5.59	0.0181
d*p	b	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	w	no	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	w	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	b	no	1	2.4044	0.6006	1.3068	3.7175	16.03	<.0001
v*p	b	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	w	no	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	w	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Observation Statistics

Observation	count	d v p	Predicted Value	Linear Predictor	Standard Error of the Linear Predictor	HessWgt	Lower	Upper	Raw Residual
		Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual	Leverage	CookD	DFBETA_ Intercept
		DFBETA_db	DFBETA_vb	DFBETA_ pno	DFBETA_ dbvb	DFBETA_ dbpno	DFBETA_ vbpno	DFBETAS_ Intercept	DFBETAS_ db
		DFBETAS_ vb	DFBETAS_ pno	DFBETAS_ dbvb	DFBETAS_ dbpno	DFBETAS_ vbpno			
1	53	w w yes	52.817821	3.9668486	0.1373785	52.817821	40.350013	69.138073	0.1821794
		0.0250674	0.025053	0.4444776	0.4447329	0.4447321	0.996823	8.8654137	1.0822244
		-1.065932	-0.082224	-1.081785	0.0182143	1.0605444	0.0702672	7.8776834	-3.267969
		-0.127307	-7.417328	0.0580937	2.8891826	0.1169918			
2	414	w w no	414.18218	6.0263059	0.0491266	414.18218	376.1619	456.04532	-0.182179
		-0.008952	-0.008952	-0.444765	-0.444733	-0.444733	0.9995949	69.713343	-0.003449
		0.0197411	1.0034492	-1.081785	-1.067459	1.0605444	0.0702672	-0.025107	0.060523
		1.5536267	-7.417328	-3.404624	2.8891826	0.1169918			
3	0	w b yes	0.1821794	-1.702763	0.6581489	0.1821794	0.0501513	0.6617845	-0.182179
		-0.426825	-0.603621	-0.628947	-0.444733	-0.616414	0.0789128	0.0024207	-0.003449
		0.0197411	-0.082224	0.0038891	0.0182143	-0.025129	0.0702672	-0.025107	0.060523
		-0.127307	0.0266656	0.0580937	-0.068458	0.1169918			
4	16	w b no	15.817821	2.7611372	0.2500983	15.817821	9.6886345	25.824428	0.1821794
		0.0458064	0.0457189	0.4438833	0.4447329	0.4447239	0.9893915	2.6352105	-0.003449
		0.0197411	1.0034492	0.0038891	-1.067459	-0.025129	0.0702672	-0.025107	0.060523
		1.5536267	0.0266656	-3.404624	-0.068458	0.1169918			
5	11	b w yes	11.182179	2.4143214	0.2967929	11.182179	6.2502284	20.005851	-0.182179
		-0.05448	-0.054629	-0.445949	-0.444733	-0.444751	0.9849937	1.8546433	-0.003449
		-1.065932	1.0034492	0.0038891	0.0182143	1.0605444	-1.015406	-0.025107	-3.267969
		1.5536267	0.0266656	0.0580937	2.8891826	-1.690608			
6	37	b w no	36.817821	3.605982	0.1644292	36.817821	26.674462	50.818342	0.1821794
		0.0300241	0.0299994	0.4443669	0.4447329	0.4447312	0.9954423	6.1712718	-0.003449
		0.0197411	1.0034492	0.0038891	-1.067459	1.0605444	-1.015406	-0.025107	0.060523
		1.5536267	0.0266656	-3.404624	2.8891826	-1.690608			
7	4	b b yes	3.8178206	1.3396797	0.5004169	3.8178206	1.4317259	10.180548	0.1821794

		0.0932377	0.0925105	0.4412644	0.4447329	0.444581	0.9560474	0.614604	-0.003449
		0.0197411	1.0034492	0.0038891	0.0182143	-0.025129	-1.015406	-0.025107	0.060523
		1.5536267	0.0266656	0.0580937	-0.068458	-1.690608			
8	139	b b no	139.18218	4.9357837	0.0847123	139.18218	117.88985	164.32015	-0.182179
		-0.015442	-0.015446	-0.44483	-0.444733	-0.444733	0.9987944	23.407779	-0.003449
		0.0197411	1.0034492	0.0038891	-1.067459	-0.025129	-1.015406	-0.025107	0.060523
		1.5536267	0.0266656	-3.404624	-0.068458	-1.690608			

❷ The GENMOD Procedure

Model Information

Data Set	WORK.LOGLIN
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	8
Number of Observations Used	8

Class Level Information

Class	Levels	Values
d	2	b w
v	2	b w
p	2	no yes

Parameter Information

Parameter	Effect	d	v	p
Prm1	Intercept			
Prm2	d	b		
Prm3	d	w		
Prm4	v		b	
Prm5	v		w	
Prm6	p			no
Prm7	p			yes
Prm8	d*v	b	b	
Prm9	d*v	b	w	
Prm10	d*v	w	b	
Prm11	d*v	w	w	
Prm12	d*p	b		no
Prm13	d*p	b		yes
Prm14	d*p	w		no
Prm15	d*p	w		yes
Prm16	v*p		b	no
Prm17	v*p		b	yes
Prm18	v*p		w	no
Prm19	v*p		w	yes

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1	0.3798	0.3798
Scaled Deviance	1	0.3798	0.3798
Pearson Chi-Square	1	0.1978	0.1978
Scaled Pearson X2	1	0.1978	0.1978
Log Likelihood		2926.7234	
Full Log Likelihood		-19.2098	
AIC (smaller is better)		52.4197	
AICC (smaller is better)		.	
BIC (smaller is better)		52.9758	

Algorithm converged.

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

Parameter			DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	3.9668	0.1374	3.6850	4.2245	833.78	<.0001
d	b		1	-1.5525	0.3262	-2.2399	-0.9504	22.66	<.0001
d	w		0	0.0000	0.0000	0.0000	0.0000	.	.
v	b		1	-5.6696	0.6459	-7.0608	-4.4854	77.06	<.0001
v	w		0	0.0000	0.0000	0.0000	0.0000	.	.
p	no		1	2.0595	0.1458	1.7836	2.3565	199.40	<.0001
p	yes		0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	b	b	1	4.5950	0.3135	4.0080	5.2421	214.78	<.0001
d*v	b	w	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	w	b	0	0.0000	0.0000	0.0000	0.0000	.	.
d*v	w	w	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	b	no	1	-0.8678	0.3671	-1.5633	-0.1140	5.59	0.0181
d*p	b	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	w	no	0	0.0000	0.0000	0.0000	0.0000	.	.
d*p	w	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	b	no	1	2.4044	0.6006	1.3068	3.7175	16.03	<.0001
v*p	b	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	w	no	0	0.0000	0.0000	0.0000	0.0000	.	.
v*p	w	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Observation Statistics

Observation	count	d v p	Predicted Value	Linear Predictor	Standard Error of the Linear Predictor	HessWgt	Lower	Upper	Raw Residual
		Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual	Leverage	CookD	DFBETA_Intercept
		DFBETA_db	DFBETA_vb	DFBETA_pno	DFBETA_dbvb	DFBETA_dbpno	DFBETA_vbpno	DFBETAS_Intercept	DFBETAS_db
		DFBETAS_vb	DFBETAS_pno	DFBETAS_dbvb	DFBETAS_dbpno	DFBETAS_vbpno			
1	53	w w yes	52.817821	3.9668486	0.1373785	52.817821	40.350013	69.138073	0.1821794
		0.0250674	0.025053	0.4444776	0.4447329	0.4447321	0.996823	8.8654137	1.0822244
		-1.065932	-0.082224	-1.081785	0.0182143	1.0605444	0.0702672	7.8776834	-3.267969
		-0.127307	-7.417328	0.0580937	2.8891826	0.1169918			
2	414	w w no	414.18218	6.0263059	0.0491266	414.18218	376.1619	456.04532	-0.182179
		-0.008952	-0.008952	-0.444765	-0.444733	-0.444733	0.9995949	69.713343	-0.003449
		0.0197411	1.0034492	-1.081785	-1.067459	1.0605444	0.0702672	-0.025107	0.060523
		1.5536267	-7.417328	-3.404624	2.8891826	0.1169918			
3	0	w b yes	0.1821794	-1.702763	0.6581489	0.1821794	0.0501513	0.6617845	-0.182179
		-0.426825	-0.603621	-0.628947	-0.444733	-0.616414	0.0789128	0.0024207	-0.003449
		0.0197411	-0.082224	0.0038891	0.0182143	-0.025129	0.0702672	-0.025107	0.060523
		-0.127307	0.0266656	0.0580937	-0.068458	0.1169918			
4	16	w b no	15.817821	2.7611372	0.2500983	15.817821	9.6886345	25.824428	0.1821794
		0.0458064	0.0457189	0.4438833	0.4447329	0.4447239	0.9893915	2.6352105	-0.003449
		0.0197411	1.0034492	0.0038891	-1.067459	-0.025129	0.0702672	-0.025107	0.060523
		1.5536267	0.0266656	-3.404624	-0.068458	0.1169918			
5	11	b w yes	11.182179	2.4143214	0.2967929	11.182179	6.2502284	20.005851	-0.182179
		-0.05448	-0.054629	-0.445949	-0.444733	-0.444751	0.9849937	1.8546433	-0.003449
		-1.065932	1.0034492	0.0038891	0.0182143	1.0605444	-1.015406	-0.025107	-3.267969
		1.5536267	0.0266656	0.0580937	2.8891826	-1.690608			
6	37	b w no	36.817821	3.605982	0.1644292	36.817821	26.674462	50.818342	0.1821794
		0.0300241	0.0299994	0.4443669	0.4447329	0.4447312	0.9954423	6.1712718	-0.003449
		0.0197411	1.0034492	0.0038891	-1.067459	1.0605444	-1.015406	-0.025107	0.060523
		1.5536267	0.0266656	-3.404624	2.8891826	-1.690608			
7	4	b b yes	3.8178206	1.3396797	0.5004169	3.8178206	1.4317259	10.180548	0.1821794
		0.0932377	0.0925105	0.4412644	0.4447329	0.444581	0.9560474	0.614604	-0.003449
		0.0197411	1.0034492	0.0038891	0.0182143	-0.025129	-1.015406	-0.025107	0.060523
		1.5536267	0.0266656	0.0580937	-0.068458	-1.690608			
8	139	b b no	139.18218	4.9357837	0.0847123	139.18218	117.88985	164.32015	-0.182179
		-0.015442	-0.015446	-0.44483	-0.444733	-0.444733	0.9987944	23.407779	-0.003449
		0.0197411	1.0034492	0.0038891	-1.067459	-0.025129	-1.015406	-0.025107	0.060523
		1.5536267	0.0266656	-3.404624	-0.068458	-1.690608			

SAS logit modeling of death penalty data

```
data loglin;
  input d v yes n;
  cards;
1 1 53 467
1 0  0  16
0 1 11  48
0 0  4 143
;
run;
❶ proc genmod data=loglin;
    model yes/n=d v/dist=bin link=logit lrci obstats;
   run;

❷ proc genmod data=loglin;
    model yes/n=v /dist=bin link=logit lrci obstats;
   run;
```

❶

```
                        The GENMOD Procedure

                          Model Information
              Data Set                        WORK.LOGLIN
              Distribution                       Binomial
              Link Function                         Logit
              Response Variable (Events)              yes
              Response Variable (Trials)                n

               Number of Observations Read            4
               Number of Observations Used            4
               Number of Events                      68
               Number of Trials                     674

                          Response Profile
                   Ordered    Binary         Total
                     Value    Outcome    Frequency
                         1    Event             68
                         2    Nonevent         606

                       Parameter Information
                      Parameter       Effect
                      Prm1            Intercept
                      Prm2            d
                      Prm3            v

                   Criteria For Assessing Goodness Of Fit     Deviance of (DP, VP, DV)

         Criterion                   DF         Value        Value/DF
         Deviance                     1        0.3798          0.3798
         Scaled Deviance              1        0.3798          0.3798
         Pearson Chi-Square           1        0.1978          0.1978
         Scaled Pearson X2            1        0.1978          0.1978
         Log Likelihood                     -209.4783
         Full Log Likelihood                  -6.6499
         AIC (smaller is better)              19.2998
         AICC (smaller is better)                   .
         BIC (smaller is better)              17.4587
       Algorithm converged.


                          Analysis Of Maximum Likelihood Parameter Estimates

                             Standard     Likelihood Ratio 95%         Wald
Parameter     DF    Estimate      Error      Confidence Limits     Chi-Square     Pr > ChiSq
```

Intercept	1	-3.5961	0.5069	-4.7754	-2.7349	50.33	<.0001
d	1	-0.8678	0.3671	-1.5633	-0.1140	5.59	0.0181
v	1	2.4044	0.6006	1.3068	3.7175	16.03	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Observation Statistics

Observation	yes	n	d	v	Predicted Value	Linear Predictor	Standard Error of the Linear Predictor	HessWgt	Lower	Upper
			Raw Residual	Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual	Leverage	CookD
			DFBETA_Intercept	DFBETA_d	DFBETA_v	DFBETAS_Intercept	DFBETAS_d	DFBETAS_v		
1	53	467	1	1	0.1131003	-2.059457	0.1458456	46.844111	0.0874393	0.1450946
			0.1821794	0.0266178	0.0266044	0.4445101	0.4447329	0.4447321	0.9964178	18.338917
			-0.049027	1.0605443	0.0702672	-0.096717	2.8891823	0.1169918		
2	0	16	1	0	0.0113862	-4.463901	0.6158305	0.1801051	0.0034329	0.0370799
			-0.182179	-0.429276	-0.60535	-0.627148	-0.444733	-0.616408	0.0683043	0.0048334
			-0.049027	-0.025129	0.0702672	-0.096717	-0.068458	0.1169918		
3	11	48	0	1	0.2329621	-1.191661	0.3380944	8.5771557	0.1353672	0.3707483
			-0.182179	-0.062205	-0.062324	-0.445581	-0.444733	-0.444749	0.980436	3.3039979
			-0.049027	1.0605444	-1.015406	-0.096717	2.8891824	-1.690608		
4	4	143	0	0	0.026698	-3.596104	0.5069138	3.7158927	0.0100544	0.0689731
			0.182179	0.0945076	0.0937908	0.441359	0.4447318	0.4445801	0.9548418	1.3940224
			1.036644	-0.025129	-1.015404	2.0450103	-0.068458	-1.690604		

❷ The GENMOD Procedure

Model Information

Data Set	WORK.LOGLIN
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	yes
Response Variable (Trials)	n

Number of Observations Read	4
Number of Observations Used	4
Number of Events	68
Number of Trials	674

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	68
2	Nonevent	606

Parameter Information

Parameter	Effect
Prm1	Intercept
Prm2	v

Criteria For Assessing Goodness Of Fit

Deviance of $(DV,\ PV)$

Criterion	DF	Value	Value/DF
Deviance	2	$\boxed{5.3940}$	2.6970
Scaled Deviance	2	5.3940	2.6970
Pearson Chi-Square	2	5.8109	2.9054
Scaled Pearson X2	2	5.8109	2.9054
Log Likelihood		-211.9854	
Full Log Likelihood		-9.1570	
AIC (smaller is better)		22.3140	
AICC (smaller is better)		34.3140	
BIC (smaller is better)		21.0866	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.6571	0.5064	-4.8357	-2.7972	52.15	<.0001
v	1	1.7045	0.5237	0.7995	2.9072	10.59	0.0011
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

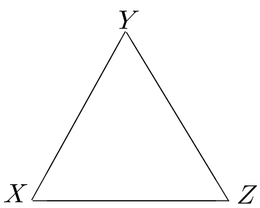
Observation Statistics

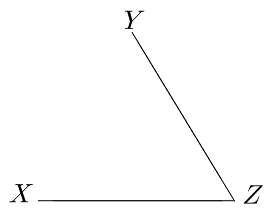
Observation	yes	n	v	Predicted Value	Linear Predictor	Standard Error of the Linear Predictor	HessWgt	Lower	Upper
			Raw Residual	Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual	Leverage
			CookD	DFBETA_ Intercept	DFBETA_v	DFBETAS_ Intercept	DFBETAS_v		
1	53	467	1	0.1242718	-1.952584	0.1335751	50.822842	0.098466	0.155673
			-5.034953	-0.706262	-0.715377	-2.343246	-2.31339	-2.316189	0.9067961
			26.034154	0	-0.963856	0	-1.840367		
2	0	16	0	0.0251573	-3.657128	0.5064099	0.3923905	0.0094741	0.0650959
			-0.402517	-0.642576	-0.902958	-0.952134	-0.677572	-0.928187	0.1006289
			0.0256841	-0.114776	0.1147756	-0.226646	0.2191502		
3	11	48	1	0.1242718	-1.952584	0.1335751	5.223761	0.098466	0.155673
			5.0349513	2.2029449	2.0055762	2.1061259	2.3133897	2.1262976	0.0932039
			0.2750375	0	0.0990687	0	0.1891598		
4	4	143	0	0.0251573	-3.657128	0.5064099	3.5069902	0.0094741	0.0650959
			0.4025066	0.2149343	0.2112023	0.6657898	0.6775546	0.67638	0.8993711
			2.0515151	1.0257807	-1.025781	2.0255938	-1.958605		

<u>Association Graphs and collapsibility</u> (Lauritzen, Graphical Models, 1996)

An association graph portrays conditional associations by edge(paths) connecting variables(nodes)

ex) $(XY,\ XZ,\ YZ)$ $\qquad\qquad\qquad\qquad$ $(XZ,\ YZ)$





- Two variables are conditionally indep., given any subset of variables that separates them in an assoc. graph.

ex) $(WX,\ XY,\ YZ)$

W and Z are conditionally indep., given X alone or Y alone or both X and Y

- For 3 way tables, XY marginal and conditional odds ratios are identical if either(**)
  - Z and X are conditionally indep. or
  - Z and Y are conditionally indep.

  (proof: exer 9.26)

ex) $(XY,\ YZ)$

$\Rightarrow$ $\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$

(X and Z conditional indep. given Y)

$\Rightarrow$ log conditional odds $\log \frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}} = (\lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY})$ does not depend on k

XY marg. association same as conditional associations(by (**)), since Z and X conditional indep.

$\rightarrow$ $P(X=i, Z=k \mid Y=j) = P(X=i \mid Y=j)P(Z=k \mid Y=j)$ for all $j = 1, \cdots, J$

Similarly YZ marginal association same as conditional associations, since X and Z conditional indep.

$\Rightarrow$ log conditional odds $\log\dfrac{\mu_{i11}\mu_{i22}}{\mu_{i12}\mu_{i21}} = (\lambda_{11}^{YZ}+\lambda_{22}^{YZ}-\lambda_{12}^{YZ}-\lambda_{21}^{YZ})$ does not depend on i

X and Z may be marginally indep. since Y and Z conditionally indep., Y and X conditionally indep.

ex) $(XY,\ Z)$
Here, X and Z are both conditionally and marginally indep.

ex) $(X,\ YZ)$
Each pair conditionally and marginally indep.

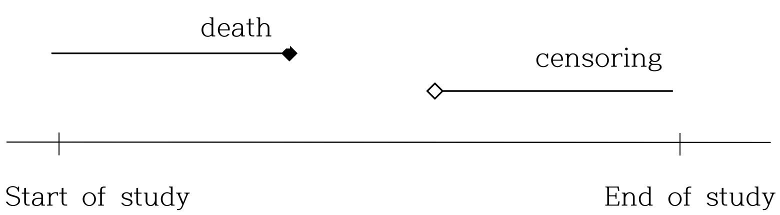
ex) $(XY,\ XZ,\ YZ)$
Each conditional association my differ from corresponding marginal association.

Loglinear Modeling of Rates

ex) Survival of patients after heart valve replacement operation.(Laird and Olivier, 1981, JASA)
109 patients observed from 3 to 97 month. 21 death, 88 censored observation.

		Type of Heart valve	
		Aortic	Mitral
Age	<55	4(1259)	1(2082)
	55+	7(1417)	9(1647)

count is no. death(n_{ij}) and ( ) is sum of times to death+times to censoring, for all subjects at that age-value combination(t_{ij})



Sampe death rate= $\dfrac{no.\ deaths}{no.\ person\ months\ of\ exposure} = \dfrac{n_{ij}}{t_{ij}}$

		Type of Heart valve	
		Aortic	Mitral
Age	<55	0.0032	0.0005
	55+	0.0049	0.0055

where $0.0032 = \frac{4}{1259}$

Let $\mu_{ij} = E(n_{ij} \mid t_{ij})$.
Model

$$\log\left(\frac{\mu_{ij}}{t_{ij}}\right) = \alpha + \beta_i^A + \beta_j^V$$

where $\{n_{ij}\}$ are indep. Poisson $\{\mu_{ij}\}$

$$\mu_{ij} = t_{ij}\exp\underbrace{(\alpha + \beta_i^A + \beta_j^V)}_{\text{rate of death(indep. of time)}}$$

This is a loglinear model with $\log\mu_{ij}$ replaced by

$$\log\mu_{ij} - \underbrace{\log t_{ij}}_{\text{offset}}$$

For ML fit,

$$G^2 = 3.2(df = 4 - 3 = 1)$$

with constraint $\hat{\beta}_2^A = \hat{\beta}_2^V = 0$

$$\hat{\beta}_1^A = -1.22(s.e = 0.51)$$

Death rate for old age group estimated to be $e^{1.22} = 3.4$ times death rate for younger group, for each valve type.

Valve type does not have significant effect.

$Z = \frac{0.3299}{0.4382}$, $Z^2 = 0.57$, $p-value = 0.45$

$LR = 3.79 - 3.22 = 0.57$, $df = 1$

Note:

- Assuming $n_{ij} \sim Pois[t_{ij}\exp(\alpha + \beta_i^A + \beta_j^V)]$ is equivalent to assuming "time to death" T has negative exponential dist. with para. $\lambda = \exp(\alpha + \beta_i^A + \beta_j^V)$
- Generalize constant rate assumption by breaking time scale into intervals and using rate λ_k in interval k.
- Permit covariates by expressing λ as function of $\underline{x}$ and para. $\underline{\beta}$

$$\lambda(\underline{x},\ \underline{\beta}) = \lambda\exp(\underline{\beta}^T\underline{x})$$

- Use $\lambda(t)\exp(\underline{\beta}^T\underline{x})$ with $\lambda(t)$ treated nonparametrically(Cox, 1972)

SAS for Poisson loglinear model with offset for modeling rates(Heart valve data)

```
data survival;
 input age valve deaths exposure;
 logexp=log(exposure);
cards;
1 1 4 1259
1 2 1 282
2 1 7 1417
2 2 9 1647
;
/* Both predictors*/
❶ proc genmod data=survival;
     class age valve;
     model deaths=age valve/dist=poi link=log offset=logexp obstats;
   run;

/* only age as predictor*/
❷ proc genmod data=survival;
     class age valve;
     model deaths=age/dist=poi link=log offset=logexp obstats;
   run;
```