

Solution to Tutorial 8

1. (a) The parameter estimates for the fitted (DV, DP, PV) model are:

$$\hat{\lambda} = 4.9358, \hat{\lambda}_1^D = -2.1746, \hat{\lambda}_2^D = 0,$$

$$\hat{\lambda}_1^V = -1.3298, \hat{\lambda}_2^V = 0, \hat{\lambda}_1^P = -3.5961, \hat{\lambda}_2^P = 0.0$$

and

$$\hat{\lambda}_{11}^{DV} = 4.5950, \hat{\lambda}_{11}^{PV} = 2.4044, \hat{\lambda}_{11}^{DP} = -0.8678.$$

- (b) The fitted values are:

Victims	Defendants	Death	Penalty
Race	Race	Yes	No
White	White	52.82	414.18
	Black	11.18	36.82
Black	White	0.18	15.82
	Black	3.82	139.18

From the fitted values, estimated odds ratio between D and P when $V = \text{White}$ is 0.42 and when $V = \text{Black}$, it is 0.42. Since this is a homogeneous association model, the odds ratios are same at each level of V . Note that, this odds ratio can also be obtained as

$$\exp\{\hat{\lambda}_{11}^{DP} + \hat{\lambda}_{22}^{DP} - \hat{\lambda}_{12}^{DP} - \hat{\lambda}_{21}^{DP}\} = \exp(-0.8678) = 0.42.$$

The value of this common odds ratio is smaller than 1 indicating that the odds of getting death penalty for a white defendant is smaller than that of a black defendant controlling for the victims race.

- (c) The fitted marginal table between Death Penalty verdict and defendant's race is

Defendants	Death	Penalty
Race	Yes	No
White	53	430
Black	15	176

Note that, the marginal table from the fitted values is the same as the one obtained from the sample data. Therefore, the marginal odds ratio is 1.446. It is greater than 1 indicating that the odds of getting a death penalty for whites are greater than that of blacks if we ignore the victims race. Thus, we observe a completely different type of association in the marginal table than the conditional association controlling for victims race. This illustrates Simpsons paradox.

- (d) The goodness-of-fit test statistic for this homogeneous association model is $G^2 = 0.3798$

with $d.f. = 1$ (p-value = 0.5377). Thus, we can conclude that the homogeneous association model is a good fit.

(e) The parameter estimates of the fitted (DV, PV) model are:

$$\hat{\lambda} = 4.9374, \hat{\lambda}_1^V = -1.1989, \hat{\lambda}_2^V = 0.0, \hat{\lambda}_1^D = -2.1903, \hat{\lambda}_2^D = 0.0,$$

$$\hat{\lambda}_1^P = -3.6571, \hat{\lambda}_2^P = 0.0$$

and

$$\hat{\lambda}_{11}^{DV} = 4.4654, \hat{\lambda}_{11}^{PV} = 1.7045.$$

This is the conditional independence model between D and P controlling for V . But the estimated conditional log odds ratio between D and V controlling for P is given by $\hat{\lambda}_{11}^{DV} = 4.4654$ and the estimated conditional odds ratio between P and V controlling for D is given by $\lambda_{11}^{PV} = 1.7045$.

The likelihood ratio goodness-of-fit statistic for this model is, $G^2 = 5.3940$ with $d.f. = 2$ (p-value = 0.06741). Since the p-value is greater than 0.05, we can conclude that this model is also a good fit. The following table shows the adjusted residuals obtained from this model fit:

Victims	Defendants	Death	Penalty
Race	Race	Yes	No
White	White	-.2313	2.313
	Black	.2313	-2.313
Black	White	-0.678	0.678
	Black	0.678	-0.678

The adjusted residuals in the partial table for victims race white are little larger compared to those in the partial for victims race black. This suggests that the conditional independence model may not be a very good fit.

(f) To test for $D - P$ partial association, we need to test the null hypothesis $H_0 : \lambda_{11}^{DP} = \lambda_{12}^{DP} = \lambda_{21}^{DP} = \lambda_{22}^{DP} = 0$. The likelihood ratio test statistics for this purpose is

$$G^2[(DV, PV)|(DV, PV, DP)] = G^2(DV, PV) - G^2(DV, PV, DP)$$

$$= 5.3940 - 0.3798 = 5.0142$$

with $d.f. = 2 - 1 = 1$. It has a p-value 0.0251, which is smaller than 0.05 and thus indicates that the null hypothesis can be rejected. That is, the partial association between D and P is significant.

SAS codes:

```
data homicide;
  input V $ D $ P $ count;
datalines;
White White Yes 53
White White No 414
White Black Yes 11
White Black No 37
Black White Yes 0
```

```

Black White No 16
Black Black Yes 4
Black Black No 139
;
run;
proc catmod data=homicide order=data;
  weight count;
  model V*D*P = _response_ / zero=samp freq pred=freq;
  loglin D|V D|P P|V;
run;
quit;
proc genmod data=homicide order=data;
  class V D P;
  model count = D|V D|P P|V / dist = poi;
run;
proc genmod data=homicide order=data;
  class V D P;
  model count = D|V P|V / dist = poi;
  output out=temp STDRESCHI =adjres p=pred;
run;
proc print data=temp noobs;
  var V D P count pred adjres;
run;

```

2. Since $\{\mu_{ij} = n\pi_{ij}\}$ satisfies the independence model in a $I \times J$ table, $\pi_{ij} = \pi_{i+}\pi_{+j}$. Therefore,

$$\log \mu_{ij} = \log n + \log \pi_{i+} + \log \pi_{+j}.$$

The loglinear independence model can be written as

$$\log \mu_{ij} = \log n + \log \pi_{i+} + \log \pi_{+j}.$$

(a) From the first model, we can write

$$\begin{aligned}
 \log \mu_{ia} - \log \mu_{ib} &= (\log n + \log \pi_{i+} + \log \pi_{+a}) - (\log n + \log \pi_{i+} + \log \pi_{+b}) \\
 &= \log \pi_{+a} - \log \pi_{+b} \\
 &= \log(\pi_{+a}/\pi_{+b}).
 \end{aligned}$$

Now from the second model,

$$\log \mu_{ia} - \log \mu_{ib} = (\lambda + \lambda_i^X + \lambda_a^Y) - (\lambda + \lambda_i^X + \lambda_b^Y) = \lambda_a^Y - \lambda_b^Y.$$

Thus, $\lambda_a^Y - \lambda_b^Y = \log(\pi_{+a}/\pi_{+b})$.

(b) All $\lambda_j^Y = 0$ implies that $\log(\pi_{+a}/\pi_{+b}) = 0$ or $\pi_{+a} = \pi_{+b}$ for all a and b . Let all $\pi_{+j} = \pi$, then from the relation

$$\sum_{j=1}^J \pi_{+j} = 1$$

we have $\pi = 1/J$. That is, $\pi_{+j} = 1/J$ for all j .

3. (a) We first fit the model (AD, DG) of conditional independence between gender and the admission controlling for the department:

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

The fitted values are given in the following table:

Department	Whether admitted, male		Whether admitted, female	
	Yes	No	Yes	No
1	531.43	293.57	69.57	38.43
2	354.19	205.81	15.81	9.19
3	114.00	211.00	208.00	385.00
4	141.63	275.37	127.37	247.63
5	48.08	142.92	98.92	294.08
6	24.03	348.97	21.97	319.03

The deviance for the fitted model is 21.7355 with degrees of freedom 6. Thus, it has a p-value of 0.0014, indicating that the model is a poor fit. Next we fit the homogeneous association model (AD, AG, DG):

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

The fitted values are given in the following table:

Department	Whether admitted, male		Whether admitted, female	
	Yes	No	Yes	No
1	529.27	295.73	71.73	36.27
2	353.64	206.36	16.36	8.64
3	109.25	215.75	212.75	380.25
4	137.21	279.79	131.79	243.21
5	45.68	145.32	101.32	291.68
6	22.96	350.04	23.04	317.96

The deviance for the fitted model is 20.2043 with degrees of freedom 5. Thus, it has a p-value of 0.0011, indicating that the homogeneous association model is also a poor fit.

(b) We compute the adjusted residuals for the model (AD, DG) and report them in the following table:

Department	Whether admitted, male		Whether admitted, female	
	Yes	No	Yes	No
1	-4.153	4.153	4.153	-4.153
2	-0.504	0.504	0.504	-0.504
3	0.868	-0.868	-0.868	0.868
4	-0.546	0.546	0.546	-0.546
5	1.000	-1.000	-1.000	1.000
6	-0.620	0.620	0.620	-0.620

We notice that the residuals are large for Department 1, but for other departments the residuals are small. So most likely department 1 is an outlier.

(c) We again fit the model (AD, DG) after deleting the data for Department 1. We report the fitted values in the following table:

Department	Whether admitted, male		Whether admitted, female	
	Yes	No	Yes	No
2	354.19	205.81	15.81	9.19
3	114.00	211.00	208.00	385.00
4	141.63	275.37	127.37	247.63
5	48.08	142.92	98.92	294.08
6	24.03	348.97	21.97	319.03

The deviance for the fitted model is 2.6815 with degrees of freedom 5. Thus, it has a p-value of 0.7489, indicating that the model is a very good fit and we can conclude that there is no association between gender and admission in the departments.

Some SAS codes:

```
data admit;
  input D A $ G $ count @@;
  label D = Department A=Admission G=Gender;
datalines;
1 yes male 512 1 no male 313 1 yes female 89 1 no female 19
2 yes male 353 2 no male 207 2 yes female 17 2 no female 8
3 yes male 120 3 no male 205 3 yes female 202 3 no female 391
4 yes male 138 4 no male 279 4 yes female 131 4 no female 244
5 yes male 53 5 no male 138 5 yes female 94 5 no female 299
6 yes male 22 6 no male 351 6 yes female 24 6 no female 317
;
run;
proc genmod data=admit;
  class a d g;
  model count = a|d a|g d|g / dist=p link=log;
  output out=temp STDRESCHI=adjres pred=pred;
run;
proc print data=temp noobs;
  var d g a count pred adjres;
run;
proc genmod data=admit;
  class a d g;
  model count = a|d d|g / dist=p link=log;
  output out=temp STDRESCHI=adjres pred=pred;
run;
proc print data=temp noobs;
  var d g a count pred adjres;
run;
data admit1;
  set admit;
  where d>1;
run;
proc genmod data=admit1;
  class a d g;
  model count = a|d d|g / dist=p link=log;
  output out=temp STDRESCHI=adjres pred=pred;
run;
proc print data=temp noobs;
  var d g a count pred adjres;
run;
```