STAT 659 Spring 2016
Homework 10 Solution

7.3

(a) H_0 : Model fits adequately $v.s.$ H_a : Model does not fit adequately
(i)$\chi^2 = 0.4794$, $df = 1$, p-value $= 0.4887$.
(ii)$G^2 = 0.5196$, $df = 1$, p-value $= 0.4710$.
Since the p-values are greater than 0.05, so at significant level $\alpha = 0.05$, we fail to reject H_0 which means there is no evidence of lack of fit.

(b) Conditional OR for PB association: $e^{0.7211} \approx 2.0567$.
Conditional OR for PH association: $e^{1.5520} \approx 4.7209$.
Conditional OR for BH association: $e^{0.4672} \approx 1.5955$.

(c) H_0 : (PH , BH) $\Leftrightarrow$ BP are conditionally independent
The model under the null hypothesis is $\log(\mu_{i,j,k}) = \lambda + \lambda_i^P + \lambda_j^B + \lambda_k^H + \lambda_{ik}^{PH} + \lambda_{jk}^{BH}$.
$G^2 = 4.64$, $df = 1$, p-value $= 0.0313$.
At significant level $\alpha = 0.05$, we reject H_0. There is evidence of conditional association BP.

(d) 95% C.I. for BP conditional OR: $(e^{0.7211-1.96(0.3539)}, e^{0.7211+1.96(0.3539)}) \simeq (1.0278, 4.1154)$.

7.4

```
             Criteria For Assessing Goodness Of Fit
-----------------------------------------------------------------
Criterion                     DF          Value         Value/DF
Deviance                       1         0.3007           0.3007
Scaled Deviance                1         0.3007           0.3007
Pearson Chi-Square             1         0.3074           0.3074
Scaled Pearson X2              1         0.3074           0.3074

Log Likelihood 2304.2847, Full Log Likelihood -22.8414

AIC (smalleris better) 59.6827,  AICC (smaller is better) .
```

```
BIC (smaller is better) 60.2388
Algorithm converged.

            Analysis Of Maximum Likelihood Parameter Estimates
------------------------------------------------------------------------------------------
                           Likelihood Ratio
Parameter        DF     Estimate     SD        TS          Pr>ChiSq
Intercept        1       4.3426    0.1120    1502.59      <.0001
gender F         1       0.3856    0.1434     7.23        0.0072
gender M         0       0.0000    0.0000       .         .
 info  O         1      -2.7147    0.3035    80.02        <.0001
 info  S         0       0.0000    0.0000       .         .
health O         1       0.7269    0.1353    28.88        <.0001
health S         0       0.0000    0.0000       .         .
gender*info FO   1       0.4636    0.2406     3.71        0.0540
gender*info FS   0       0.0000    0.0000       .         .
gender*info MO   0       0.0000    0.0000       .         .
gender*info MS   0       0.0000    0.0000       .         .
gender*health FO 1      -0.2516    0.1749     2.07        0.1503
gender*health FS 0       0.0000    0.0000       .         .
gender*health MO 0       0.0000    0.0000       .         .
gender*health MS 0       0.0000    0.0000       .         .
info*health OO   1       0.8997    0.2852     9.95        0.0016
info*health OS   0       0.0000    0.0000       .         .
info*health SO   0       0.0000    0.0000       .         .
info*health SS   0       0.0000    0.0000       .         .

       LR Statistics For Type 3 Analysis
--------------------------------------------------
Source             DF    Chi-Square     Pr > ChiSq

gender               1      15.94         <.0001

info                 1      345.34        <.0001

health               1      68.54         <.0001

gender*info          1       3.83         0.0505

gender*health        1       2.08         0.1490

info*health          1       11.37        0.0007
```

(a) H_0 : Model fits adequately $v.s.$ H_a : Model does not fit adequately
$\chi^2 = 0.3074$, $df = 1$, p-value $= 0.579$.
At significant level $\alpha = 0.05$, we fail to reject H_0. So there is no evidence of lack of fit.

(b) The estimated conditional GI odds ratio is $\exp(0.4636) = 1.5898$. The 95% confidence interval for $OR_{\mathrm{GI(H)}}$: $(e^{-0.0009}, e^{0.9452}) \simeq (0.9991, 2.5733)$.

Since 1 lies within the interval, we are 95% confident that there is no conditional association between G and I.

(c) $H_0 : \lambda^{GI} = 0$ $v.s.$ $H_a : \lambda^{GI} \neq 0$
$\chi^2 = 3.71$, $df = 1$, p-value $= 0.054$.
At significant level $\alpha = 0.05$, there is slight evidence that G and I are conditionally independent. GI does not need to be in the model.

7.6

(a) For the mutually independence model, the deviance is 135.86 with P-value close to zero. So the model is not adequate.

(b) For the homogeneous model, the SN*JP term has the largest likelihood ratio test statistic 74.63, which indicates that the estimated conditional association between SN and JP is strongest.

(c) Similarly, the likelihood ratio test statistic between EI and TF or EI and JP are the smallest with P-values greater than 0.1, thus there is no strong evidence of conditional association between these two pairs.

7.7

(a) The test statistic is $TS = 12.3687 - 10.16 = 2.2087 \sim \chi_2$. The P-value is 0.3314 which implies the model that assumes conditional independence between E/I and T/F and between E/I and J/P holds.

(b) The 95 percent CI for the conditional odds ratio between the S/N and J/P is $(e^{-1.5075}, e^{-0.9382}) = (0.221, 0.391)$. Since this interval excludes one, so we can see S/N and J/P are conditionally associated.

(c) If we use this parameterization, then the estimated conditional odds ratio is $e^{1.2202} = 3.388$ and the 95 percent likelihood-ratio confidence interval is $(e^{0.9382}, e^{1.5075}) = (2.555, 4.515)$.

7.8

(a) For the mutually independence model,$\log(\mu_{ijkl}) = \lambda + \lambda_i^{EI} + \lambda_j^{SN} + \lambda_k^{TF} + \lambda_l^{JP}$, so the number of parameters is 5; For the homogeneous association model, $\log(\mu_{ijkl}) = \lambda + \lambda_i^{EI} + \lambda_j^{SN} + \lambda_k^{TF} + \lambda_l^{JP} + \lambda_{ij}^{EI*SN} + \lambda_{ik}^{EI*TF} + \lambda_{il}^{EI*JP} + \lambda_{jk}^{SN*TF} + \lambda_{jl}^{SN*JP} + \lambda_{kl}^{TF*JP}$, so the number of parameters is 11; for the model with all the three-factor interaction terms, $\log(\mu_{ijkl}) = \lambda + \lambda_i^{EI} + \cdots + \lambda_{kl}^{TF*JP} + \lambda_{ijk}^{EI*SN*TF} + \lambda_{ijl}^{EI*SN*JP} + \lambda_{ikl}^{EI*TF*JP} + \lambda_{jkl}^{SN*TF*JP}$, so the number of parameters is $11 + 4 = 15$.

(b) The AIC scores for the above three models are $-6940.38, -7054.1, -7049.16$ respectively. Since the second model has the smallest AIC score, it is preferred.

7.9

(a) $\widehat{OR}_{\text{AG(D)}} = e^{-0.0999} \approx 0.9$.

$$\widehat{OR}_{\text{AG}} = \frac{(1198)(1278)}{(557)(1493)} \approx 1.84.$$

Men applied in greater numbers to D(1,2) which had relatively high admission rates and women applied in greater numbers to D(3,4,5,6) which had relatively low admission rates.

(b) (i) H_0 : Model fits adequately $v.s.$ H_a : Model does not fit adequately
$G^2 = 20.2043$, $df = 5$, p-value $= 0.001$.
At significant level $\alpha = 0.05$, we reject H_0. There is evidence of lack of fit.

(ii) extraordinary standardized residuals in department 1: (-4.0273, 4.0273, 4.0272, -4.0273).
The standardized residuals show lack of fit for department 1.

(c) (i) H_0 : Model fits adequately $v.s.$ H_a : Model does not fit adequately
$G^2 = 2.5564$, $df = 4$, p-value $= 0.63$.
At significant level $\alpha = 0.05$, we fail to reject H_0. There is no evidence of lack of fit.

(ii) All absolute values of standardized residuals are less than 2. So, there is good fit.

(d) $\text{logit}(\pi(\text{A=yes})) = \alpha + \beta_i^D \times D_i + \beta^G \times G$, $i = 2, 3, 4, 5$; $G = \begin{cases} 1, & \text{Male} \\ 0, & \text{Female} \end{cases}$.

(i)

	$\hat{\beta}_2^D$	$\hat{\beta}_3^D$	$\hat{\beta}_4^D$	$\hat{\beta}_5^D$	$\hat{\beta}^G$	$\hat{\alpha}$
Estimates	3.2052	2.0651	2.0106	1.5921	0.0307	-2.6917

For each department, the estimated conditional odds ratio (male:female) $e^{0.0307} \approx 1.03$ between gender and admission.

(ii) In log-linear model, $\hat{\lambda}^{AG} = 0.0307 \Rightarrow e^{0.037} \approx 1.03$.
The result is the same as that of the logit model.

(iii) 95% confidence interval for $OR_{\text{AG(D)}}$: (0.87 , 1.22).
It is plausible that admissions and gender are conditionally independent for these departments.

7.10

Let B: status of seat belt in use, E: whether ejected, and I: status of injury

```
Summary Measures for Loglinear Models
---------------------------------------------------------
modnum    model_name              G2       DF     loglikhd
   1      B E I             11444.38        4   6674995.13
   2      B*E B*I E*I           2.85        1   6680715.89
   3      B*E B*I            1680.41        2   6679877.11
   4      B*E E*I            1144.64        2   6680145.00
   5      B*I E*I            7133.98        2   6677150.33
   6      B*E              775659.55        4   6292887.55
   7      B*I              739427.94        4   6311003.35
   8      E*I              120479.48        4   6620477.58
   9      B*E*I                 0.00        0   6680717.32
```

(a) The model (BE, BI, EI) is the best choice for the loglinear model except for the saturated model.

$\log(\mu_{ijk}) = \alpha + \lambda_i^B + \lambda_j^E + \lambda_k^I + \lambda_{ij}^{BE} + \lambda_{ik}^{BI} + \lambda_{jk}^{EI}$

	$\hat{\lambda}^B_{1(\text{yes})}$	$\hat{\lambda}^E_{1(\text{no})}$	$\hat{\lambda}^I_{1(\text{no})}$	$\hat{\lambda}^{BE}$	$\hat{\lambda}^{BI}$	$\hat{\lambda}^{EI}$	$\hat{\alpha}$
Estimates	-3.1564	-0.7278	2.2458	2.3996	1.7173	2.7978	6.1947

(i) the conditional odds ratio of BE is $e^{2.3996} \approx 11.02$.

(ii) the conditional odds ratio of BI is $e^{1.7173} \approx 5.57$.

(iii) the conditional odds ratio of EI is $e^{2.7978} \approx 16.41$.

(b) Define $B = \begin{cases} 1, & \text{Belt in use} \\ 0, & \text{o.w.} \end{cases}$; $E = \begin{cases} 1, & \text{Not ejected} \\ 0, & \text{o.w.} \end{cases}$.

$\text{logit}(\hat{\pi}(\text{fatal})) = -2.2455 - 1.7173B - 2.7982E.$

(i) Controlling for the ejected level, the effect on using belt is $e^{-1.7173} \approx 0.18$.

(ii) Controlling for the level of using belt, the effect on ejected is $e^{-2.7982} \approx 0.061$.

(c) $\hat{D} = 4.7675 \times 10^{-5}$ is very small. It suggests that the sample data is consistent with the model.

7.14

Let PS=S, RA=R, PV=P and BC=B

```
Summary Measures for Loglinear Models
-----------------------------------------------------------------------------------------
modnum    model_name                                  G2    DF
loglikhd  numpar   aicc
   1      PS RA BC PV                            277.085    18    2556.86      6   289.176
   2      PS*RA PS*BC PS*PV RA*BC RA*PV BC*PV      6.963     9    2691.92     15    37.491
   3      PS*RA*BC PS*RA*PV PS*BC*PV RA*BC*PV      0.451     2    2695.18     22    45.571
```

4	PS*RA PS*BC PS*PV RA*BC RA*PV	32.950	11	2678.93	13	59.349
5	PS*RA PS*BC PS*PV RA*BC BC*PV	10.701	11	2690.05	13	37.100
6	PS*RA PS*BC PS*PV RA*PV BC*PV	20.725	10	2685.04	14	49.186
7	PS*RA PS*BC RA*BC RA*PV BC*PV	25.866	11	2682.47	13	52.265
8	PS*RA PS*PV RA*BC RA*PV BC*PV	64.064	10	2663.37	14	92.525
9	PS*BC PS*PV RA*BC RA*PV BC*PV	56.124	10	2667.34	14	84.585
10	PS*RA PS*BC PS*PV RA*BC	39.332	13	2675.74	11	61.621
11	PS*RA PS*BC PS*PV BC*PV	27.108	12	2681.85	12	51.449
12	PS*RA PS*BC RA*BC BC*PV	34.899	13	2677.95	11	57.188
13	PS*RA PS*PV RA*BC BC*PV	66.053	12	2662.38	12	90.394
14	PS*BC PS*PV RA*BC BC*PV	65.156	12	2662.83	12	89.498

(a) We can choose the model: (PS*RA, PS*BC, PS*PV, RA*BC, BC*PV). Additionally, we do the model checking:
H_0 : Model fits adequately $v.s.$ H_a : Model does not fit adequately
$G^2 = 10.701$, $df = 11$, p-value = 0.47.
At significant level $\alpha = 0.05$, we fail to reject H_0. There is no evidence of lack of fit on the selected model.

(b)

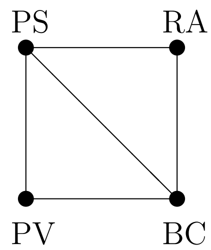
Association	$\hat{\lambda}^{PS*RA}$	$\hat{\lambda}^{PS*BC}$	$\hat{\lambda}^{PS*PV}_{(1)}$	$\hat{\lambda}^{PS*PV}_{(2)}$	$\hat{\lambda}^{RA*BC}$	$\hat{\lambda}^{BC*PV}_{(1)}$	$\hat{\lambda}^{BC*PV}_{(2)}$
Estimates	1.1898	1.13	0.8844	0.1008	0.6418	0.9729	0.6342
Conditional OR	3.29	3.1	2.42	1.11	1.9	2.65	1.89

(c) Logistic model without interactions: (RA+PV+BC) $\Longleftrightarrow$
Loglinear model: (PS*RA, PS*BC, PS*PV, RA*BC, BC*PV, RA*PV, RA*PV*BC).
H_0 : Model fits adequately $v.s.$ H_a : Model does not fit adequately
$G^2 = 5.7870$, $df = 7$, p-value=0.565.
At significant level $\alpha = 0.05$, we fail to reject H_0. There is no evidence of lack of fit.

Association	$\hat{\lambda}^{PS*RA}$	$\hat{\lambda}^{PS*BC}$	$\hat{\lambda}^{PS*PV}_{(1)}$	$\hat{\lambda}^{PS*PV}_{(2)}$
Estimates	1.1489	1.1487	0.8054	0.0774
Conditional OR	3.15	3.15	2.24	1.08

The results shown in table indicate that the effect in (b) is more significant than that in (c). Moreover, we also carry out a test:
H_0 : Simpler model in (a) is better $v.s.$ H_a : Model in (c) is better
$G^2 = 4.914$, $df = 4$, p-value = 0.30.
At significant level $\alpha = 0.05$, we fail to reject H_0. So, the model in (a) is more effective than the model in (c).

(d) The independence graph of the loglinear model selected in (a):



For each connected pair, the fitted marginal and conditional associations are not identical.

7.16

(a)

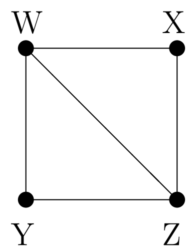
$$\begin{aligned}
\mathrm{logit}(P(I=1)) &= \log\left\{\frac{P(I=1|G=g,L=l,S=s)}{P(I=2|G=g,L=l,S=s)}\right\} \\
&= \log(\mu_{1gls}) - \log(\mu_{2gls}) \\
&= (\lambda_1^I - \lambda_2^I) + (\lambda_{g1}^{GI} - \lambda_{g2}^{GI}) + (\lambda_{1l}^{IL} - \lambda_{2l}^{IL}) + (\lambda_{1s}^{IS} - \lambda_{2s}^{IS})
\end{aligned}$$

(b)

$$\begin{aligned}
&\mathrm{logit}(P(I=1|S=1)) - \mathrm{logit}(P(I=1|S=2)) \\
&= \log\left(\frac{\mu_{1gl1}\mu_{2gl2}}{\mu_{1gl2}\mu_{2gl1}}\right) \\
&= \log(\mu_{1gl1}) + \log(\mu_{2gl2}) - \log(\mu_{1gl2}) - \log(\mu_{2gl1}) \\
&= \lambda_{11}^{IS} + \lambda_{22}^{IS} - \lambda_{12}^{IS} - \lambda_{21}^{IS}
\end{aligned}$$

7.19

(a) The independence graph of the loglinear model selected in (a):

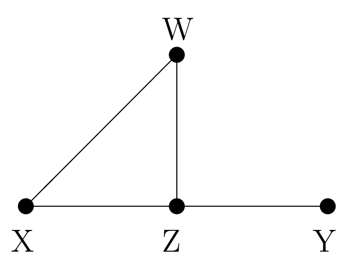


(b) (i) λ^{XY} is not in the model, so X and Y are conditionally independent.

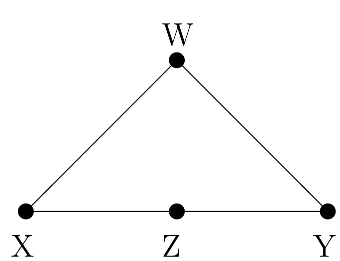
(ii) All terms in the saturated model that are not in the model (WXZ, WYZ) involve X and Y and so permit XY conditional association.

7.20

(a) Yes.

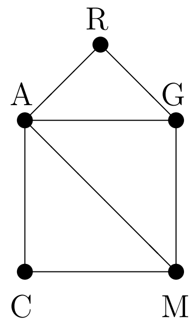


(b) No. Even if given Z, X and Y are conditional association via W.

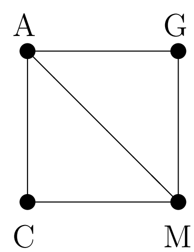


7.22

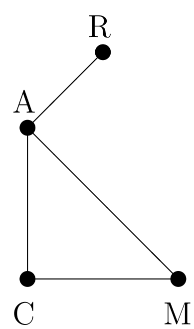
(AM) involves with G and C.



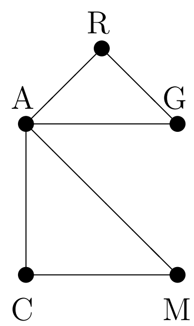
(a) (i) (AM) still involve with G and C.



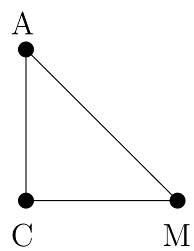
(ii) (AM) just involves with C.



(b) For given A, (RG) and (CM) are conditionally independent.



(c) (AC, AM, CM) do not involve with (GR).



7.24

```
Independence model

             Criteria For Assessing Goodness Of Fit
-----------------------------------------------------------------
Criterion                      DF           Value        Value/DF
Deviance                       24        112.5356          4.6890
Scaled Deviance                24        112.5356          4.6890
Pearson Chi-Square             24        106.1941          4.4248
Scaled Pearson X2              24        106.1941          4.4248

Log Likelihood                          2204.8616

Full Log Likelihood -144.1743

AIC (smaller is better) 312.3487

AICC (smaller is better) 325.9139

BIC (smaller is better) 331.3509
Algorithm converged.

                    Analysis Of Maximum Likelihood Parameter Estimates
-------------------------------------------------------------------------------------------
                                     Standard   Wald 95% Confidence          Wald
 Parameter          DF    Estimate     Error          Limits            Chi-Square   Pr > ChiSq
 Intercept           1      2.2811    0.1508      1.9854      2.5767        228.69       <.0001
```

RA	RA0	1	0.7932	0.1596	0.4804	1.1061	24.69	<.0001
RA	RA1	1	0.3390	0.1733	-0.0007	0.6787	3.82	0.0505
RA	RA2	1	0.8548	0.1581	0.5449	1.1647	29.22	<.0001
RA	RA3	1	0.5317	0.1669	0.2046	0.8588	10.15	0.0014
RA	RA4	1	0.2474	0.1768	-0.0990	0.5938	1.96	0.1616
RA	RA5	1	0.5002	0.1679	0.1712	0.8293	8.88	0.0029
RA	RA6	1	-0.1310	0.1938	-0.5108	0.2487	0.46	0.4989
RA	RA7	1	1.3276	0.1490	1.0356	1.6196	79.41	<.0001
RA	RA8	0	0.0000	0.0000	0.0000	0.0000	.	.
TBC	SA	1	0.4565	0.1014	0.2579	0.6552	20.29	<.0001
TBC	S	1	0.7118	0.0968	0.5221	0.9016	54.05	<.0001
TBC	D	1	0.1886	0.1072	-0.0216	0.3988	3.09	0.0786
TBC	SD	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	Raw Residual	Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual
1	14.846652	2.5404573	2.3834313	3.0034247	3.2012973	3.0781692
2	4.9136069	0.740028	0.726883	0.9699119	0.9874517	0.9776392
3	-7.12527	-1.394026	-1.465915	-1.771441	-1.684569	-1.744526
4	-12.63499	-2.716418	-3.079352	-3.64022	-3.211181	-3.523517
5	9.3153348	2.0004242	1.8780325	2.3013201	2.4512975	2.3524814
6	-0.991361	-0.187378	-0.188501	-0.244592	-0.243135	-0.244001
7	-5.587473	-1.371909	-1.462324	-1.718388	-1.612141	-1.68975
8	-2.736501	-0.738342	-0.765149	-0.879578	-0.848762	-0.872181
9	9.6781857	1.6058693	1.5414716	1.9522347	2.0337928	1.9833385
10	8.1144708	1.1850612	1.1531304	1.546424	1.5892453	1.5655799
11	-2.784017	-0.528171	-0.537382	-0.652654	-0.641468	-0.649073
12	-15.00864	-3.12893	-3.621385	-4.302545	-3.717462	-4.140501
13	7.7073434	1.5030986	1.4374041	1.7793472	1.8606698	1.8080147
14	3.0604752	0.5253346	0.5177219	0.6786284	0.688607	0.6828171
15	-1.112311	-0.248025	-0.250365	-0.297208	-0.294429	-0.296404
16	-9.655508	-2.365899	-2.678691	-3.110703	-2.747465	-3.021005
17	1.212743	0.2726315	0.2699154	0.3293914	0.3327059	0.3304839
18	-3.542117	-0.700865	-0.718079	-0.92792	-0.905676	-0.919061
19	-1.136069	-0.29201	-0.295782	-0.346147	-0.341733	-0.344962
20	3.4654427	0.9788231	0.938239	1.0741187	1.1205804	1.0853102
21	0.5205184	0.1031195	0.1027713	0.12699	0.1274202	0.1271386
22	3.1101512	0.5423137	0.5340867	0.698816	0.7095804	0.7033128
23	-3.490281	-0.79059	-0.816144	-0.967093	-0.936812	-0.958476
24	-0.140389	-0.034944	-0.034995	-0.040566	-0.040507	-0.040551
25	-5.552916	-1.508359	-1.634402	-1.968185	-1.816401	-1.922351
26	-1.4946	-0.357333	-0.362611	-0.462383	-0.455653	-0.459803
27	4.6328294	1.4388522	1.3477845	1.5564363	1.6616023	1.5833981
28	2.4146868	0.8241054	0.7894029	0.8917846	0.9309878	0.9004141
29	-26.27754	-3.442184	-3.766736	-5.034885	-4.601065	-4.848657
30	-10.22678	-1.179106	-1.20746	-1.709033	-1.668901	-1.689053
31	12.421166	1.8603644	1.7826853	2.2850876	2.3846584	2.324566
32	24.083153	3.9637025	3.6197802	4.5390055	4.9702651	4.7005752
33	-11.45032	-2.913057	-3.477062	-4.20399	-3.522072	-4.001131
34	-2.943845	-0.65919	-0.676494	-0.866097	-0.843943	-0.857527

35	4.1814253	1.2163031	1.1534605	1.337383	1.410246	1.3564187
36	10.212743	3.2644641	2.8566686	3.2401362	3.7026726	3.3486734

(a) (i) $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$, $i = 1, \cdots, 8$; $j = 1, 2, 3$.
The parameter estimates are obtained from the above table.

(ii) From the standardized Pearson residuals, we observe that there are many absolute values greater than 2. So, we concern the adequacy.

Linear by linear association model(1)

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	23	19.9014	0.8653
Scaled Deviance	23	19.9014	0.8653
Pearson Chi-Square	23	19.6035	0.8523
Scaled Pearson X2	23	19.6035	0.8523
Log Likelihood		2251.1787	
Full Log Likelihood		-97.8572	
AIC (smaller is better)		221.7145	
AICC(smaller is better)		238.2599	
BIC (smaller is better)		242.3002	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.4604	0.4566	-2.3554	-0.5655	10.23	0.0014
RA	RA0	1	3.0099	0.2931	2.4354	3.5843	105.46	<.0001
RA	RA1	1	2.3283	0.2841	1.7716	2.8851	67.18	<.0001
RA	RA2	1	2.6043	0.2565	2.1015	3.1071	103.07	<.0001
RA	RA3	1	2.0278	0.2427	1.5520	2.5035	69.78	<.0001
RA	RA4	1	1.4756	0.2303	1.0242	1.9270	41.06	<.0001
RA	RA5	1	1.4452	0.2039	1.0456	1.8449	50.23	<.0001
RA	RA6	1	0.5150	0.2098	0.1037	0.9263	6.02	0.0141
RA	RA7	1	1.6586	0.1548	1.3553	1.9619	114.85	<.0001
RA	RA8	0	0.0000	0.0000	0.0000	0.0000	.	.
TBC	SA	1	2.3280	0.2376	1.8624	2.7937	96.02	<.0001
TBC	S	1	2.0573	0.1900	1.6850	2.4297	117.27	<.0001
TBC	D	1	0.9090	0.1408	0.6329	1.1850	41.65	<.0001
TBC	SD	0	0.0000	0.0000	0.0000	0.0000	.	.
assoc		1	0.1215	0.0134	0.0953	0.1477	82.51	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
RA	8	260.54	<.0001
TBC	3	158.47	<.0001
assoc	1	92.63	<.0001

(b) (i) $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \beta u_i u_j$, $i = 1, \cdots, 8$; $j = 1, 2, 3$.
The parameter estimates are obtained from the above table.

(ii) $\hat{\beta} = 0.12515$ with $S.E = 0.0134$.
The positive estimate suggests that subjects having less favorable attitudes about the availability of teen birth control tend to have more times about religious attendance.
The estimated local odds ratio is $e^{0.1215} \approx 1.13$. The nonlocal odds ratio is stronger.

(c) $H_0 : \beta = 0$ $v.s.$ $H_a : \beta \neq 0$
$G^2[(X,Y)|L \times L] = 92.63$, $df = 1$, p-value < 0.0001.
At significant level $\alpha = 0.05$, we reject H_0. There is extremely strong evidence of an association.

```
Linear by linear association model(2)

              Criteria For Assessing Goodness Of Fit
-----------------------------------------------------------------
Criterion                     DF          Value        Value/DF
Deviance                      23        22.0044          0.9567
Scaled Deviance               23        22.0044          0.9567
Pearson Chi-Square            23        21.2202          0.9226
Scaled Pearson X2             23        21.2202          0.9226
Log  Likelihood                        2250.1272
Full Log Likelihood                     -98.9087
AIC (smaller is better)                 223.8175
AICC(smaller is better)                 240.3629
BIC (smaller is better)                 244.4032
Algorithm converged.

                     Analysis Of Maximum Likelihood Parameter Estimates
-------------------------------------------------------------------------------------------
                                      Standard   Wald 95% Confidence          Wald
Parameter          DF    Estimate       Error          Limits           Chi-Square   Pr > ChiSq
Intercept           1     -0.8850      0.3975    -1.6642    -0.1059         4.96       0.0260
RA          RA0     1      2.5722      0.2550     2.0724     3.0719       101.76       <.0001
RA          RA1     1      1.9435      0.2525     1.4485     2.4384        59.23       <.0001
RA          RA2     1      2.2732      0.2295     1.8234     2.7231        98.09       <.0001
RA          RA3     1      1.7511      0.2221     1.3158     2.1865        62.15       <.0001
RA          RA4     1      1.2539      0.2159     0.8308     1.6770        33.74       <.0001
RA          RA5     1      1.2787      0.1944     0.8977     1.6598        43.27       <.0001
RA          RA6     1      0.4039      0.2056     0.0009     0.8068         3.86       0.0495
RA          RA7     1      1.6029      0.1533     1.3025     1.9034       109.37       <.0001
RA          RA8     0      0.0000      0.0000     0.0000     0.0000          .            .
TBC         SA      1      2.1794      0.2238     1.7409     2.6180        94.86       <.0001
TBC         S       1      2.0739      0.1906     1.7003     2.4475       118.38       <.0001
TBC         D       1      0.6878      0.1248     0.4432     0.9325        30.37       <.0001
TBC         SD      0      0.0000      0.0000     0.0000     0.0000          .            .
assoc               1      0.0836      0.0093     0.0655     0.1018        81.65       <.0001
Scale               0      1.0000      0.0000     1.0000     1.0000
NOTE: The scale parameter was held fixed.
```

```
      LR Statistics For Type 3 Analysis
-------------------------------------------------
Source            DF     Chi-Square    Pr > ChiSq
RA                 8      253.15        <.0001
TBC                3      154.96        <.0001
assoc              1       90.53        <.0001
```

(d) When comparing the result with that of (c), the results are not substantively different with these scores. However, the strength of association decrease a little bit.

(e) Column (2 , 3) $\Rightarrow A = e^{2\hat{\beta}(u_c - u_d)} \Rightarrow \log(A) = 2\hat{\beta}(u_c - u_d)$.

(i) Column (1 , 2) $\Rightarrow \log\left(e^{\hat{\beta}(u_c - u_d)}\right) = \hat{\beta}(u_c - u_d)$.

(ii) Column (3 , 4) $\Rightarrow \log\left(e^{\hat{\beta}(u_c - u_d)}\right) = \hat{\beta}(u_c - u_d)$.

So, the claim holds.

(add.)

	df	G^2	p-value	AIC	AICc
Saturated	–	–	–	247.8131	–
Independence	24	112.5356	<0.0001	312.3487	325.9139
$L \times L$	23	19.9014	**0.6479**	**221.7145**	**238.2599**
Row effects	16	10.9624	0.8118	226.7755	282.7755
Column effects	21	19.8323	0.5319	225.6454	249.6454
Row & Column	14	10.9171	0.6925	230.7302	308.5763

Based on AIC or AICc , the $L \times L$ model is better and it is adequate.

Additional Problem

I.

	df	G^2	p-value	AIC	AICc
Saturated	–	–	–	**53.2384**	–
(GC,GS,CS)	1	10.1379	0.0015	61.3763	–
(GC,GS)	2	10.1625	**0.0062**	**59.4099**	143.4009
(GC,CS)	2	25.6199	<0.0001	74.8584	158.8584
(GS,CS)	2	12.9093	0.0016	62.1477	146.1477
GC	3	26.1499	<0.0001	73.3883	103.3883
GS	3	13.4392	0.0038	60.6776	90.6776
CS	3	28.8967	<0.0001	76.1351	106.1351
(G,C,S)	4	29.4266	$<$ **0.0001**	74.6650	**87.9984**

From the above results, the saturated model may be adopted.

II. A. (WXZ, YZ)

B. (WX,WY,XZ,YZ)

C. (ACM,ARG,AMG)

Only for students having taken STAT 414,610 or 630

7.25

(a) When we take $\beta = 0$, the model reduces to $\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$.

(b) LR statistic comparing this model to (XZ , YZ).

(c) Please refer to pp.230-pp.231.

(d) The local odds ratio will become e^{β_k} for the kth level of Z. Therefore, there is a heterogeneous $L \times L$ association model.

7.26

Within row i,

$$\begin{aligned}
\log\left(\frac{\mu_{j+1}}{\mu_j}\right) &= \log(\mu_{j+1}) - \log(\mu_j) \\
&= \lambda + \lambda_i^X + \lambda_{j+1}^Y + \beta x(j+1) - (\lambda + \lambda_i^X + \lambda_{j+1}^Y + \beta x j) \\
&= (\lambda_{j+1}^Y - \lambda_j^Y) + \beta x \\
&= \alpha_j + \beta x
\end{aligned}$$