a. Fit loglinear models (GH, GI), (GH, HI), (GI, HI), and (GH, GI, HI). Show that models that lack the HI term fit poorly.

<table>
<thead>
<tr>
<th>Model</th>
<th>(\alpha)</th>
<th>(G_1)</th>
<th>(H_1)</th>
<th>(I_1)</th>
<th>(G_H1)</th>
<th>(G_I1)</th>
<th>(H_I1)</th>
<th>Deviance</th>
<th>DF</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GH,GI)</td>
<td>3.6420</td>
<td>-0.5749</td>
<td>-0.6054</td>
<td>1.6094</td>
<td>-0.2082</td>
<td>0.4204</td>
<td>0</td>
<td>11.6657</td>
<td>2</td>
<td>0.003</td>
</tr>
<tr>
<td>(GH,HI)</td>
<td>3.6983</td>
<td>-0.2134</td>
<td>-1.3695</td>
<td>1.5414</td>
<td>-0.2082</td>
<td>0</td>
<td>0.8724</td>
<td>4.1267</td>
<td>2</td>
<td>0.127</td>
</tr>
<tr>
<td>(GI,HI)</td>
<td>3.8682</td>
<td>-0.6436</td>
<td>-1.4572</td>
<td>1.3759</td>
<td>0</td>
<td>0.4204</td>
<td>0.8724</td>
<td>2.3831</td>
<td>2</td>
<td>0.304</td>
</tr>
<tr>
<td>(GH,GI,HI)</td>
<td>3.8521</td>
<td>-0.5976</td>
<td>-1.3750</td>
<td>1.3514</td>
<td>-0.2516</td>
<td>0.4636</td>
<td>0.8997</td>
<td>0.3007</td>
<td>1</td>
<td>0.58</td>
</tr>
</tbody>
</table>

\(G_1\) indicates the main effect of (male) compared to the baseline \(G_2\) (female)
\(H_1\) indicates the main effect of (supporting health opinion) compared to the baseline \(H_2\) (opposing)
\(I_1\) indicates the main effect of (supporting information opinion) compared to the baseline \(I_2\) (opposing)
\(G_H1\) indicates the interaction effect of (male and supporting health opinion), and \(G_H12 = G_H21 = G_H22 = 0\)
\(G_I1\) indicates the interaction effect of (male and supporting information opinion) and \(G_I12 = G_I21 = G_I22 = 0\)
\(H_I1\) indicates the interaction effect of (supporting both information and health opinions), and \(H_I12 = H_I21 = H_I22 = 0\)

H0: candidate model  
Model (GH, GI) which lacks the HI term has p-value of 0.003, so we reject H0 and the model fit poorly. Other models which include the HI term have p-value greater than 0.05, we fail to reject H0, and the model fits well.

b. For model (GH, GI, HI), show that 95% Wald confidence intervals equal (0.55, 1.10) for the GH conditional odds ratio and (0.99, 2.55) for the GI conditional odds ratio. Interpret. Is it plausible that gender has no effect on opinion for these issues?
From the equation (8.14) in the textbook, we get
\[ \log(GH_{11}) = GH_{11} + GH_{22} - GH_{12} - GH_{21} = GH_{11} - 0 - 0 = GH_{11} \]
\[ \log(GI_{11}) = GI_{11} + GI_{22} - GI_{12} - GI_{21} = GI_{11} - 0 - 0 = GI_{11} \]

Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender*health</td>
<td>1</td>
<td>-0.2516</td>
<td>0.1749</td>
<td>-0.5945</td>
<td>0.0913</td>
<td>2.07</td>
<td>0.1503</td>
</tr>
<tr>
<td>gender*info</td>
<td>1</td>
<td>0.4636</td>
<td>0.2406</td>
<td>-0.008</td>
<td>0.9352</td>
<td>3.71</td>
<td>0.0540</td>
</tr>
</tbody>
</table>

The 95% Wald confidence intervals for GH_{11} is (-0.5945, 0.0913), so the 95% Wald confidence intervals for the GH conditional odds ratio is (e^{-0.5945}, e^{0.0913}), i.e. (0.55, 1.10).

Similarly, the 95% Wald confidence intervals for GI_{11} is (-0.008, 0.9352), so the 95% Wald confidence intervals for the GI_{11} conditional odds ratio is (e^{-0.008}, e^{0.9352}), i.e. (0.99, 2.55).

Since the 95% confidence intervals for GH_{11} and GI_{11} include 1.0, we fail to reject H0 of independence; it is plausible that gender has no effect on opinion for these issues.

2. (8.2)

Table 8.17 summarizes opinions about race-based (black/white) bussing (B), would you vote for a black presidential candidate (P), and have you recently brought a black person to your home for dinner (D). The text restricts the responses to Yes or No for this problem.

8.2a) Loglinear models for survey count, in backward elimination format, starting with the saturated model (BDP) and proceeding to the main effect only model are shown in the table on the next page. The model statement in the SAS code shown on the next page was modified to produce the results for each model in the table.

Directly below is the Proc Genmod parameter estimate portion of the SAS output for fitting model (BD, BP, DP). The conditional odds ratios for the three pairs of variables BD, BP, and DP were e^{0.4002} = 1.492, e^{1.0401} = 2.829, e^{1.5254} = 4.597, respectively. Note that we fail to reject Ho \( \beta^{BD}=0 \) (p-value = 0.0960). The 95% confidence interval for the BD conditional odds ratio is (0.93, 2.30) which spans 1. Also the Likelihood ratio test for the model (BP, DP) shown on the next page indicates this simpler model omitting BD provides and adequate fit. The data seems to indicate that opinions about bussing black and white children are independent of the frequency that the respondents invited black people home for dinner.
**Loglinear modeling of the 1991 General Social Survey issues regarding Race**;

**Data**

<table>
<thead>
<tr>
<th>D</th>
<th>B</th>
<th>P</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>N</td>
<td>N</td>
<td>41</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>D</td>
<td>71</td>
</tr>
<tr>
<td>N</td>
<td>D</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>D</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>D</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>65</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>157</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>D</td>
<td>17</td>
</tr>
<tr>
<td>N</td>
<td>D</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>5</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>N</td>
<td>34</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>D</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>Y</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>N</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>D</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>Y</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>N</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>D</td>
<td>D</td>
<td>1</td>
</tr>
</tbody>
</table>

**PROC GENMOD**

```
Data = RACE;
Class B(ref='N') D(ref='N') P(ref='N') /param=ref;
Model Count = B D P B*D B*P D*P /dist=Poisson link=log WaldCI;
Run;
```

**8.2b)** Now we use the whole dataset to fit the model (BD, BP, DP) and find that the deviance is 14.825 and df is 8. Since the critical value of 8 df at 5% is 15.51, we conclude no lack of fit. The fit is adequate.
3. (8.4) Refer to Table 2.6. D = defendant’s race, V = victim’s race, and P = death penalty verdict.
Fit the loglinear model (DV, DP, PV).

a. Using the fitted values, estimate and interpret the odds ratio between D and P at each level of V.

<table>
<thead>
<tr>
<th>Observation</th>
<th>count</th>
<th>Pred</th>
<th>Xbeta</th>
<th>Std</th>
<th>HessWgt</th>
<th>Resraw</th>
<th>Reschi</th>
<th>Resdev</th>
<th>StResdev</th>
<th>StReschi</th>
<th>Reslik</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>53</td>
<td>52.817821</td>
<td>3.9668486</td>
<td>0.1373785</td>
<td>52.817821</td>
<td>0.1821794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0250674</td>
<td>0.025053</td>
<td>0.4444776</td>
<td>0.4447329</td>
<td>0.4447321</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>414</td>
<td>414.18218</td>
<td>6.0263059</td>
<td>0.0491266</td>
<td>414.18218</td>
<td>-0.182179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.008952</td>
<td>-0.008952</td>
<td>-0.444776</td>
<td>-0.4447329</td>
<td>-0.4447323</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>11.182179</td>
<td>2.413214</td>
<td>0.2967929</td>
<td>11.182179</td>
<td>-0.182179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.05448</td>
<td>-0.054629</td>
<td>-0.444749</td>
<td>-0.444733</td>
<td>-0.444751</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>36.817821</td>
<td>3.605982</td>
<td>0.1644292</td>
<td>36.817821</td>
<td>0.1821794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0300241</td>
<td>0.0299994</td>
<td>0.4440669</td>
<td>0.4447329</td>
<td>0.4447312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.1821794</td>
<td>-1.702763</td>
<td>0.6581489</td>
<td>0.1821794</td>
<td>-0.182179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.426825</td>
<td>-0.603621</td>
<td>-0.628947</td>
<td>-0.444733</td>
<td>-0.616414</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>15.817821</td>
<td>2.7611372</td>
<td>0.2500983</td>
<td>15.817821</td>
<td>0.1821794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0458064</td>
<td>0.0457189</td>
<td>0.4438833</td>
<td>0.4447329</td>
<td>0.4447239</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3.8178206</td>
<td>1.3396797</td>
<td>0.5004169</td>
<td>3.8178206</td>
<td>0.1821794</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0932377</td>
<td>0.0925105</td>
<td>0.4412644</td>
<td>0.4447329</td>
<td>0.444581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>139</td>
<td>139.18218</td>
<td>4.9357837</td>
<td>0.0847123</td>
<td>139.18218</td>
<td>-0.182179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.015442</td>
<td>-0.015446</td>
<td>-0.444833</td>
<td>-0.444733</td>
<td>-0.444733</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fitted counts for V = White:

<table>
<thead>
<tr>
<th>Dependent</th>
<th>Death Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>White</td>
<td>52.817821</td>
</tr>
<tr>
<td>Black</td>
<td>11.182179</td>
</tr>
</tbody>
</table>

The odds ratio between D and P when victim is white is:

\[
\frac{52.817821 \times 36.817821}{(414.18218 \times 11.182179)} = 0.42
\]

Fitted counts for V = Black:

<table>
<thead>
<tr>
<th>Dependent</th>
<th>Death Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>White</td>
<td>0.1821794</td>
</tr>
<tr>
<td>Black</td>
<td>3.8178206</td>
</tr>
</tbody>
</table>

The odds ratio between D and P when victim is black is:

\[
\frac{0.1821794 \times 139.18218}{(15.817821 \times 3.8178206)} = 0.42
\]
This odds ratio is the same as the one when victim is black. It reflects that the model assumes the homogeneous association between D and P. The odds of death penalty for white defendant is 0.42 times that ratio for black defendant no matter the race of defendant.

b. Calculate the marginal odds ratio between D and P, (i) using the fitted values, and (ii) using the sample data. Why are they equal? Contrast the odds ratio with part (a). Explain why Simpson’s paradox occurs.

Fitted value

<table>
<thead>
<tr>
<th>Dependant</th>
<th>Death Penalty</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>White</td>
<td>52.817821+0.1821794 = 53</td>
<td>414.18218+15.817821=430</td>
</tr>
<tr>
<td>Black</td>
<td>11.182179+3.8178206 = 15</td>
<td>36.817821+139.18218 = 176</td>
</tr>
</tbody>
</table>

The marginal odds ratio between D and P is $53 \times 176 / (430 \times 15) = 1.446$

Sample data

<table>
<thead>
<tr>
<th>Dependant</th>
<th>Death Penalty</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>White</td>
<td>53+0 = 53</td>
<td>414+16 = 430</td>
</tr>
<tr>
<td>Black</td>
<td>11+4 = 15</td>
<td>37+139 = 176</td>
</tr>
</tbody>
</table>

The marginal odds ratio between D and P is $53 \times 176 / (430 \times 15) = 1.446$, same as that calculated from the fitted values. This is because the model (DV, DP, PV) is saturated if we are only interested in D and P factors.

The odds of death penalty for white defendant is 1.446 times that ratio for black defendant.

The marginal association has a different direction from the conditional associations. It is called Simpson’s paradox. This is due to ignoring the controlling factor victim’s race, which has strong association with the defendant’s race.

c. Fit the corresponding logit model, treating P as the response. Show the correspondence between parameter estimates and fit statistics.
Loglinear Model | Parameters | Fit Statistic
--- | --- | ---
α | D₁ | V₁ | P₁ | DV₁₁ | DP₁₁ | PV₁₁ | Deviance
(DV,DP,PV) | 4.9358 | -2.1746 | -1.3298 | -3.5961 | 4.5950 | -0.8678 | 2.4044 | 0.3798, DF = 1

D₂=P₂=V₂=DV₁₂=DV₂₁=DV₂₂=DP₁₂=DP₂₁=DP₂₂=PV₁₂=PV₂₁=PV₂₂=0

The corresponding logit model is (D+V).

<table>
<thead>
<tr>
<th>logit Model</th>
<th>Parameters</th>
<th>Fit Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>D</td>
<td>V</td>
</tr>
<tr>
<td>(D+V)</td>
<td>-3.5961</td>
<td>-0.8678</td>
</tr>
<tr>
<td>Correspondence to loglinear model</td>
<td>P₁,P₂=-3.5961</td>
<td>DP₁₁,DP₁₂ = -0.8678</td>
</tr>
</tbody>
</table>

*Ungrouped data, the saturated model is different from that of loglinear model.

** Grouped data, the saturated model the same as that of loglinear model.

d. Is there a simpler model that fit well? Interpret, and show the logit-loglinear connection.

In the model (DV, DP, PV), although the factor DP₁₁ is significant with p-value of 0.0181, it is less significant compared with other factors which have p-value <0.0001. Remove the factor DP₁₁, fit the model (DV, PV) and test goodness of fit.

Loglinear Model | Parameters | Fit Statistic
--- | --- | ---
α | D₁ | V₁ | P₁ | DV₁₁ | PV₁₁ | Deviance
(DV,PV) | 4.9374 | -2.1903 | -1.1989 | -3.6571 | 4.4654 | 1.7045 | 5.3940, DF = 2, p-value = 0.067

D₂=P₂=V₂=DV₁₂=DV₂₁=DV₂₂=PV₁₂=PV₂₁=PV₂₂=0

H₀: model (DV, PV) H₁: full model (DVP)

G²=5.3940, df = 2, p-value = 0.067

We fail to reject H₀ at 5%; the model (DV, PV) marginally fits well. This model indicates that the race of defendant and the result of death penalty are conditionally independent given the victim’s race.

The corresponding logit model is (V)
### logit Model

<table>
<thead>
<tr>
<th>logit Model</th>
<th>Parameters</th>
<th>Fit Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V)</td>
<td>( \alpha )</td>
<td>( V )</td>
</tr>
<tr>
<td>(-3.6571)</td>
<td>(1.7045)</td>
<td>(5.3940, \text{ DF } = 2 )**</td>
</tr>
</tbody>
</table>

** Correspondence to loglinear model

\[ P_1 - P_2 = -3.6571 \]
\[ PV_{11} - PV_{21} = 1.7045 \]

** Grouped data, the saturated model the same as that of loglinear model.

---

4. (8.13) Let’s define our notations first:

\( R_i \): religion attendance, \( i=1, 2 \) (baseline)

\( B_j \): birth control, \( j=1, 2 \) (baseline)

\( V_k \): political views, \( k=1, 2, 3 \) (baseline)

\( P_m \): premarital sex, \( m=1, 2 \) (baseline)

### Model selection

We begin with the **backward model selection**. For 3 or higher order interaction terms, we usually remove all insignificant terms together and we double check for the goodness of fit of the reduced model. When moving to 2 way interactions or main effects, we should remove one by one.

**Step1:** Run the full model (RBVP) and find two four-way interaction terms \( \text{RBVP}_{1111}, \text{RBVP}_{1121} \) are non-significant with p-value of 0.8873 and 0.5236.

**Step2:** Remove the four-way interactions, and fit the model (RBV, RBP, and BVP). The deviance is 4.0308 with df of 4, which is smaller than the critical value at 5% significance level. So we can safely remove all the 4-way interactions. In the reduced model, all three-way interactions \( \text{RBV}_{111}, \text{RBV}_{112}, \text{RBP}_{111}, \text{BVP}_{111} \) and \( \text{BVP}_{121} \) are found non-significant with p-value of 0.6988, 0.6025, 0.9922, 0.3811 and 0.2003 respectively.

**Step3:** Remove the three-way interactions, and fit the model (RB, RV, RP, BV, BP, VP). The ratio of deviance and it df is 0.7737 smaller than 1, so we can safely remove all the 3-way interactions. In the reduced model, all two-way interactions of RV (\( \text{RV}_{11} \) and \( \text{RV}_{12} \)) are non-significant (p-value of 0.0677 and 0.6302, respectively).

**Step4:** Remove the two-way interaction RV, and fit the model (RB, RP, BV, BP, PV).
We notice that the two-way interaction $PV_{12}$ are not significant with p-value of 0.5485, but $PV_{11}$ is highly significant, so we cannot remove the PV interaction from the model. However, if $PV_{11}$ is only marginally significant, you need to check the goodness of fit of the model after removing PV to make sure if it is safe to take it off.

All other terms are significant except $V_1$. Since PV remains in the model, we cannot remove the main effects of P and V. Overall, there is no more term to be removed and so this model is our final model.

**Check for the goodness of fit of the final model**

$H_0$: model (RB, RP, BV, BP, VP)  
$H_1$: full model (saturated model)

$G^2 = 10.7011$, df = 11, p-value = 0.469

We fail to reject $H_0$, so the final model fits data reasonably well.

**Brief interpretation of the fitted final model**

There are no 3 or higher order associations among these factors, which means that the association between any two factors does not depend on the other factors (homogeneous association). In addition, RV is dropped so we can conclude that religion attendance frequency is not significantly associated with political views. For other pair of factors:

1. RB: The odds of birth control being 1 (vs. 2) for people of religion attendance frequency 1 is $\exp(0.64)=1.90$ times that for people of frequency 2, when the other factors are held constant.
2. RP: The odds of opinion 1 on premarital sex for people of religion attendance frequency 1 is 3.29 ($\exp(1.19)$) times that odds for those of frequency 2, when the other factors are held constant.
3. BP: The odds of opinion 1 on premarital sex for people with opinion 1 on birth control is 3.10 ($\exp(1.13)$) times that odds for people with birth control opinion 2, when the other factors are held constant.
4. BV: The odds of opinion 1 on birth control for people with political view 1 is 2.64 \((e^{0.97})\) times that odds for people with political view 3 and 1.40 \((e^{0.97-0.63})\) times that odds for people political view 2, when the other factors are held constant.

5. PV: The odds of opinion 1 on premarital sex for people with political view 1 is 2.41 \((e^{0.88})\) times that odds for people with political view 3 and 2.18 \((e^{0.88-0.10})\) times that odds for people political view 2, when the other factors are held constant.

**SAS CODES**

```
***************prob 8.1 *******************************;
data opinion;
  input gender $ info $ health $ count;
  /* gender: male = 1, female = 2; info: support = 1, oppose = 2; health: support = 1, oppose = 2*/
datalines;
1 1 1 76
1 1 2 160
1 2 1 6
1 2 2 25
2 1 1 114
2 1 2 181
2 2 1 11
2 2 2 48
;
** model (GH, GI);
proc genmod data = opinion;
  class gender info health;
  model count = gender|health gender|info /dist=poisson link=log p r;
run;

** model (GH, HI);
proc genmod data = opinion;
  class gender info health;
  model count = gender|health health|info /dist=poisson link=log p r;
run;

** model (GI, HI);
proc genmod data = opinion;
  class gender info health;
  model count = gender|info health|info /dist=poisson link=log p r;
run;

** model (GH, GI, HI);
proc genmod data = opinion;
  class gender info health;
  model count = gender|health gender|info health|info /dist=poisson link=log p r;
run;
```
***************prob 8.4 ************************************;
data penalty;
   input victim $ defend $ death $ count;
   /* defend: white = 1, black = 2; victim: white = 1, black = 2; death: Yes= 1, No = 2*/
datalines;
1 1 1 53
1 1 2 414
1 2 1 11
1 2 2 37
2 1 1 0
2 1 2 16
2 2 1 4
2 2 2 139;
proc genmod data = penalty;
   class victim defend death;
   model count = defend|victim defend|death death|victim /dist=poisson link=log p r;
run;
** part(c);** corresponding logit model (D+V);
proc genmod data = penalty;
   class victim defend;
   freq count; ** saturated model is different from that of the loglinear model, ungrouped;**
   model death = victim defend /dist = bin link = logit p r;
run;
** grouped data, so that the saturated model is the same as that of the loglinear model;**
data penalty_logit;
   input victim $ defend $ yes total;
datalines;
1 1 53 467
1 2 11 48
2 1 0 16
2 2 4 143;
proc genmod data = penalty_logit;
   class victim defend;
   model yes/total = victim defend /dist = bin link = logit p r;
run;
** part(d);** corresponding logit model (V);
proc genmod data = penalty_logit;
```plaintext
  class victim;
  model yes/total = victim /dist = bin link = logit p r;
run;

***************prob 8.13 ****************************;

data opinion;
  input religion $ birth $ views $ premarital $ count;

class religion birth views premarital;
model count = religion|birth|views|premarital /dist=poisson link=log;
run;** step 1 : full model (RBVP) ;
proc genmod data = opinion;
  class religion birth views premarital;
  model count = religion|birth|views|premarital /dist=poisson link=log;
run;** step 2: model (RBV,RBP,BVP);
proc genmod data = opinion;
  class religion birth views premarital;
  model count = religion|birth|views religion|birth|premarital
      birth|views|premarital /dist=poisson link=log;
run;** step3: model (RB, RV, RP, BV, BP, VP);
proc genmod data = opinion;
  class religion birth views premarital;
  model count = religion|birth religion|views religion|premarital birth|views
      birth|premarital|views|premarital/dist=poisson link=log;
run;** step4: model (RB,RP, BV, BP, VP);
proc genmod data = opinion;
  class religion birth views premarital;
```
model count = religion|birth religion|premarital birth|views birth|premarital views|premarital/dist=poisson link=log;
run;

** model (RB,RP, BV, BP, VP RBV) which is correspond to the logit model in HW4;
proc genmod data = opinion;
   class religion birth views premarital;
   model count = religion|birth|views religion|birth religion|premarital birth|views birth|premarital views|premarital/dist=poisson link=log p r;
run;