HOMEWORK 5 WITH SOLUTIONS

1. (40 pts) The table below lists graduate admissions information for the six largest departments at U.C. Berkeley in the fall of 1973. This the example we saw in the lecture notes, and you can use the data files and code provided in Lesson 5.

Dept.	No. of men rejected	No. of men accepted	No. of women rejected	No. of women accepted
A	313	512	19	89
B	207	353	8	17
C	205	120	391	202
D	278	139	244	131
E	138	53	299	94
F	351	22	317	24

Let D = department, S =sex,and A = admission status (rejected or accepted).

(a) Give the marginal table $S \times A$. Analyze the marginal table for $S \times A$, reporting the X^2 test and odds ratio. Is there any evidence of gender bias in graduate admissions?

Solution:
Let D=Department, S=sex and A=admission status

```
>
> ### Create "flat" contigency tables
>
> ftable(temp1)
                      DEP   A   B   C   D   E   F
admission gender
yes       M               512 353 120 139  53  22
          F                89  17 202 131  94  24
no        M               313 207 205 278 138 351
          F                19   8 391 244 299 317
>
>
> ### Test the Mutual Independence, step by step
> ### compute the expected frequences
> E=array(NA,dim(temp1))
> for (i in 1:dim(temp1)[1]) {
+ for (j in 1:dim(temp1)[2]) {
+ for (k in 1:dim(temp1)[3]) {
+ E[i,j,k]=(margin.table(temp1,3)[k]*margin.table(temp1,2)[j]*
```

```
margin.table(temp1,1))[i]/(sum(temp1))^2
}}}

> ### compute the X^2, and G^2
> df=(length(temp1)-1)-(sum(dim(temp1))-3)
> X2=sum((temp1-E)^2/E)
> X2
[1] 1999.553
> 1-pchisq(X2,df)
[1] 0
> G2=2*sum(temp1*log(temp1/E))
> G2
[1] 2097.187
> 1-pchisq(G2,df)
[1] 0
>
```

The value of the X^2 statistic is 1999.53 with p-value=0.00 and the value of the G^2 statistic is 2097.187 with p-value 0.00 indicating that the three categorical variables Department, Sex and Admission are related with each other.

The SxA marginal table is:

	Men	Women
Accepted	1199	557
Rejected	1492	1278

R code:

```
> Gender_Admission=margin.table(temp1,c(2,1))
> Gender_Admission
      admission
gender  yes   no
     M 1199 1492
     F  557 1278
>
> result=chisq.test(Gender_Admission)
> result
```

```
Pearson's Chi-squared test with Yates' continuity correction

data:  Gender_Admission
X-squared = 92.0733, df = 1, p-value < 2.2e-16

>
> oddsSA=oddsratio(Gender_Admission,log=FALSE)
> oddsSA
[1] 1.843852
>
```

SAS Code:

```
data berkeley;
input gender $ accept $ count;
datalines;
Men Accept 1119 Men Reject 1492
Women Accept 557
Women Reject 1278
;

proc freq data=berkeley order=data;
weight count;
tables gender*accept / chisq cmh;
run;
```

Thus ignoring department gender and admission are related (p-value<0.05). The marginal odds ratio is $\hat{\theta}_{SA} = 1.84$. Thus ignoring department the odds of admission for men are 1.84 times the odds of admission for women. Men are more likely (in the odds scale) to be admitted than women, showing that there is bias gender in favor of men.

(b) Give the partial table $S \times A$ for each level of D. Examine the $S \times A$ tables within each level of D, reporting the X^2 statistics and odds ratios. Does this contradict what you found in part (a)? Why or why not? What is the Cochran-Mantel-Haenszel (CMH) statistic for our example? What does this statistic tell you about the relationship between admission and gender when adjusting for department?

Solution:

The SxA partial table at the level A of D is :

Gender	YES	NO
M	512	313
F	89	19

Conditioning on A, $X^2 = 17.25$ and $\hat{\theta_{SA(A)}} = 0.349 < 1$. Thus males are less likely to be admitted than females at the department A.

The SxA partial table at the level B of D is :

Gender	YES	NO
M	353	207
F	19	8

Conditioning on B, $X^2 = 0.25$ and $\hat{\theta_{SA(B)}} = 0.803 < 1$. Thus males are less likely to be admitted than females at the department B.

The SxA partial table at the level C of D is :

Gender	YES	NO
M	120	205
F	202	391

Conditioning on C, $X^2 = 0.75$ and $\hat{\theta_{SA(C)}} = 1.13 > 1$. Thus males are more likely to be admitted than females at the department C.

The SxA partial table at the level D is :

Gender	YES	NO
M	139	278
F	131	244

Conditioning on D, $X^2 = 0.23$ and $\hat{\theta_{SA(D)}} = 0.93 < 1$. Thus males are less likely to be admitted than females at the department D.

The SxA partial table at the level E of D is :

Gender	YES	NO
M	53	138
F	94	299

Conditioning on E, $X^2 = 1.00$ and $\hat{\theta_{SA(E)}} = 1.221$. Thus males are more likely to be admitted than females at the department E.

The SxA partial table at the level F of D is :

Gender	YES	NO
M	22	351
F	24	317

Conditioning on F, $X^2 = 0.38$ and $\hat{\theta_{SA(F)}} = 0.827$. Thus males are less likely to be admitted than females at the department F.

Yes this contradicts with the conclusion in part (a). Observe that the variables Gender and acceptance are independent given all levels of department except department A. Also, from the odds ration we can say that controlling on the department women are admitted more often than men (Simpon's paradox).

Moreover, from the output, we see that CMH=1.43 with $df = 1$ and p-value=0.2316 implying that Sex and Admission are conditionally independent of D. Therefore, after controlling for D there is not enough evidence that gender and admission are related. Controlling for D the gender bias in favor of men disappeared and this is because D is related with S and A. However, note that the conditional odds ratio of S and A do not have the same direction at all levels of D and the CMH statistic works well when the conditional odds ratio have the same direction and are comparable in size.

R CODE:

```
> SA_A=t(temp1[,,1])
> SA_A
      admission
gender yes  no
     M 512 313
     F  89  19
```

```
> chisq.test(temp1[,,1], correct=FALSE)

Pearson's Chi-squared test

data:  temp1[, , 1]
X-squared = 17.248, df = 1, p-value = 3.28e-05

> thetaSA_A=oddsratio(t(temp1[,,1]),log=FALSE)
> thetaSA_A
[1] 0.349212

> SA_B=t(temp1[,,2])
> SA_B
      admission
gender yes  no
     M 353 207
     F  17   8
> chisq.test(temp1[,,2], correct=FALSE)

Pearson's Chi-squared test

data:  temp1[, , 2]
X-squared = 0.2537, df = 1, p-value = 0.6145

> thetaSA_B=oddsratio(t(temp1[,,2]),log=FALSE)
> thetaSA_B
[1] 0.8025007


> SA_C=t(temp1[,,3])
> SA_C
      admission
gender yes  no
     M 120 205
     F 202 391
> chisq.test(temp1[,,3], correct=FALSE)

Pearson's Chi-squared test

data:  temp1[, , 3]
X-squared = 0.7535, df = 1, p-value = 0.3854
```

```
> thetaSA_C=oddsratio(t(temp1[,,3]),log=FALSE)
> thetaSA_C
[1] 1.13306

> t(temp1[,,4])
      admission
gender yes  no
     M 139 278
     F 131 244
> chisq.test(temp1[,,4], correct=FALSE)

Pearson's Chi-squared test

data:  temp1[, , 4]
X-squared = 0.225, df = 1, p-value = 0.6353

> thetaSA_D=oddsratio(t(temp1[,,4]),log=FALSE)
> thetaSA_D
[1] 0.9312977

> t(temp1[,,5])
      admission
gender yes  no
     M  53 138
     F  94 299
> chisq.test(temp1[,,5], correct=FALSE)

Pearson's Chi-squared test

data:  temp1[, , 5]
X-squared = 1.0011, df = 1, p-value = 0.3171

> thetaSA_E=oddsratio(t(temp1[,,5]),log=FALSE)
> thetaSA_E
[1] 1.221631
>

> t(temp1[,,6])
      admission
gender yes  no
     M  22 351
     F  24 317
```

```
> chisq.test(temp1[,,6], correct=FALSE)

Pearson's Chi-squared test

data:  temp1[, , 6]
X-squared = 0.3841, df = 1, p-value = 0.5354

> thetaSA_F=oddsratio(t(temp1[,,6]),log=FALSE)
> thetaSA_F
[1] 0.8278727


> mantelhaen.test(temp1,correct=FALSE)

Mantel-Haenszel chi-squared test without continuity correction

data:  temp1
Mantel-Haenszel X-squared = 1.4313, df = 1, p-value = 0.2316
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.7743773 1.0636021
sample estimates:
common odds ratio
        0.9075403
```

SAS Code:

```
data berkeley;
input dept $ gender $ accept $ count;
datalines;
A Men Reject 313
B Men Reject 207
C Men Reject 205
D Men Reject 278
E Men Reject 138
F Men Reject 351
A Men Accept 512
B Men Accept 353
C Men Accept 120
D Men Accept 139
E Men Accept 53
F Men Accept 22
A Women Reject 19
B Women Reject 8
```

```
C Women Reject 391
D Women Reject 244
E Women Reject 299
F Women Reject 317
A Women Accept 89
B Women Accept 17
C Women Accept 202
D Women Accept 131
E Women Accept 94
F Women Accept 24
;
run;

proc freq data=prob1b order=data;
weight count;
tables dept*gender*accept / chisq cmh;
run;
```

(c) Drop the department "A" from the analysis, and re-test for the conditional independence model. What is the CMH statistic now and what does it tell you about the relationship between admission and gender when adjusting for department?

Solution:

Now we drop department A and we re-test the conditional independence model.

```
> # # # #question c.  Drop the Dep A and re-run the analysis.
>
>  admission=c('yes','no')
> gender=c('M','F')
> DEP=c('B','C','D','E','F')
> table=expand.grid(admission=admission,gender=gender,DEP=DEP)
> count=c(353,207,17,8,120,205,202,391,139,278,131,244,
53,138,94,299,22,351,24,317)
> table=cbind(table,count=count)
> temp2=xtabs(count~admission+gender+DEP,table)
> ftable(temp2)
                 DEP   B   C   D   E   F
admission gender
yes       M          353 120 139  53  22
          F           17 202 131  94  24
no        M          207 205 278 138 351
          F            8 391 244 299 317
>
```

```
> X2=sum(chisq.test(temp1[,,2], correct=FALSE)$statistic+
chisq.test(temp1[,,3],correct=FALSE)$statistic
+chisq.test(temp1[,,4], correct=FALSE)$statistic,
chisq.test(temp1[,,5],
 correct=FALSE)$statistic,chisq.test(temp1[,,6],
correct=FALSE)$statistic)
> X2
[1] 2.617379
> 1-pchisq(X2,df=5)
[1] 0.7587231
> > mantelhaen.test(temp2,correct=FALSE)

Mantel-Haenszel chi-squared test without continuity correction

data:  temp2
Mantel-Haenszel X-squared = 0.1556, df = 1, p-value = 0.6932
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.8732553 1.2259454
sample estimates:
common odds ratio
          1.03468

>
```

SAS Code:

```
proc freq data=prob1b order=data;
where dept ne 'A';
weight count;
tables dept*gender*accept / chisq cmh;
run;
```

Now we observe from the analysis that the overall X^2 statistic of testing that S and A are conditionally independent given D gave a p-value of 0.758. Thus, we fail to reject the null hypothesis indicating that the conditional independence model (SD,AD) is a good fit for this data after we drop the department A. Also, CMH=0.15 which also indicates that after droping the A level of D the S and A are conditionally independent of D (i.e D is related to A but not S).

(d) In the Lesson 5 notes we saw that with the department A in the analysis, the model of homogenous associations was rejected (the Breslow-Day statistic =18.83, with df=5, p-value=0.0021). Report the Breslow-day statistics for the analysis without department A. What is your conclusion now about the this model?

Solution: The Breslow-Day statistic is 2.45 with p-value=0.653. This implies that there is homogeneous association in the model, i.e the effect of gender on admission is the same for each department. This agrees with the result found in part (c).

R Code:

```
>
> breslowday.test(temp2)
Breslow and Day test (with Tarone correction):
Breslow-Day X-squared         = 2.453813
Breslow-Day-Tarone X-squared  = 2.45378

Test for test of a common OR: p-value =  0.6529309
```

SAS Code:

```
proc freq;

weight count;

tables department*status*gender/chisq cmh nocol nopct;

tables department*status*gender/chisq nocol nopct;

run;
```

Breslow-Day Test for Homogeneity of the Odds Ratios	
Chi-Square	2.4538
DF	4
Pr > ChiSq	0.6529

2. (15 pts) For three-way contingency tables, when any pair of variables is conditionally independent, explain why there is homogenous association. When there is no homogeneous association, explain why no pair of variables can be conditionally independent.

Solution: Homogeneous association says that the conditional relationship between any pair of variables given the third one is the same at each level of the third one, i.e the conditional odds ratios for any pair of variables given the third one are identical. Conditional independence means that odds ratios of any pair of variables given the third one are all equal to 1 at each level of the third one.

i) If any pair of variables is conditionally independent, this means that $\theta_{XY(1)} = \theta_{XY(2)} = ... = \theta_{XY(K)} = 1$, which is the special case of $\theta_{XY(1)} = \theta_{XY(2)} = ... = \theta_{XY(K)} = c$ where the association pattern is the same for all levels of Z, i.e., the homogeneous association is present. Therefore, if there is conditional independence, there is also homogeneous association.

ii) Conversely, if there is no homogeneous association, it means at least one of the odds ratios is not equal to the other conditional odds ratios. With the same argument, if all the odds ratios were equal and equal to 1; this means that at least one off the odds ratio is not equal to one. Therefore this entails that if there is conditional independence there is also homogeneous association but homogeneous association does not necessarily mean conditional independence. In other words, if there is no homogeneous association among, say X and Y, the conditional odds-ratios of X and Y given others do not all equal. It follows that X and Y are not conditionally independent or the conditional odds-ratios should all be 1.

3. (15 pts) State if the the statement is TRUE or FALSE and give an explanation for your answer.

   Suppose that income (high, low) and gender are conditionally independent, given type of job (secretarial, construction, service, professional, etc). Then, income and gender are also independent in the 2×2 marginal table (i.e., ignoring, rather than controlling, type of job).

   Solution: This statement is FALSE. If I(Income) and G (Gender) are conditionally independent given J(Job) it does not mean that I and G are marginally independent. The example of Boys Scouts and Juvenile Delinquency and Simpson's paradox showed that ignoring S (Socioeconomic status) B (Boy scout) and D (Delinquent) were NOT independent in the $B \times D$ marginal table but B and D were conditionally independent given S.

4. (30 pts) Montana Economic Outlook This 1992 Montana poll asked a random sample of Montana residents whether their personal financial status was the worse, the same, or better than a year ago, and whether they thought the state economic outlook was better over the next year. Data and accompanying demographics about the respondents is provided

in files on ANGEL. The data file contains results for every other person included in the poll to reduce the size. The data are given in two formats: *montana.xls,* and *montana.csv*. There is also *montana.sas*, as well as *montana.R*. If you want you can read the data in any other format. There are some missing values which you can ignore for this analysis.

Here is the coding for the data: AGE = 1 under 35, 2 35-54, 3 55 and over
SEX = 0 male, 1 female
INC = yearly income: 1 for under \$20K, 2 for \$20-35K, 3 for over \$35K
POL = 1 Democrat, 2 Independent, 3 Republican
AREA = 1 Western, 2 Northeastern, 3 Southeastern Montana
FIN = Financial status 1 worse, 2 same, 3 better than a year ago
STAT = State economic outlook 0 better, 1 not better than a year ago

(a) Is there a relationship between income and party affiliation? Turn in the appropriate statistic, p-value and your conclusion. Report a measure of association such as odds ratio(s).

Solution:
H_0 : INC and POL are independent.

From the output the Pearson goodness of fit test statistic X^2 is 6.670 with pvalue=0.153. Therefore, we fail to reject the null hypothesis indicating that income and party affiliation are independent.

The odds of an Independent being in the middle income rather than the lower is $(22 \cdot 16)/(30 \cdot 7)$=1.67 times that of a Republican. Similarly, the odds of an Independent being in the highest income rather than in the middle one are $(22 \cdot 27)/(30 \cdot 7) = 2.83$ times that of an Republican. Similarly, the odds of an independent being in the middle income bracket rather than the lower is 0.42 times that of a Democrat.(see R code)

Table of INC by POL				
	POL(PARTY AFFILIATION)			
INC(YEARLY INCOME)	INDEPENDENT	REPUBLICAN	DEMOCRAT	Total
20-35$K	22	30	29	81
	15.762	31.962	33.276	
	27.16	37.04	35.80	
>35$K	7	27	25	59
	11.481	23.281	24.238	
	11.86	45.76	42.37	
< $20K	7	16	22	45
	8.7568	17.757	18.486	
	15.56	35.56	48.89	
Total	36	73	76	185
Frequency Missing = 24				

R Code

```
> montana=read.table("montana.csv", header=TRUE, sep=",")
> attach(montana)
> inc_pol=table(Inc,Pol, exclude=".", dnn=list("Inc","Pol"))
> inc_pol.data=as.data.frame(inc_pol)
> inc_pol.data
> temp=xtabs(Freq~Inc+Pol,inc_pol.data)
> temp
   Pol
Inc  1  2  3
  1 22  7 16
  2 29 22 30
  3 25  7 27
> result<-chisq.test(temp, correct=FALSE)
> result
        Pearson's Chi-squared test
data:  temp
X-squared = 6.6995, df = 4, p-value = 0.1526

> temp=(matrix(data=c(22,30,7,16),ncol=2,nrow=2))
> oddsratio(temp,log=FALSE)
[1] 1.67619

> temp1
```

```
     [,1] [,2]
[1,]   29   22
[2,]    7   22
> temp1=t((matrix(data=c(22,29,7,22),ncol=2,nrow=2)))
> oddsratio(temp1,log=FALSE)
[1] 2.384236
>
```

SAS Code/Output:

```
proc freq data=Montana order=data;
tables INC*POL / chisq cmh relrisk expected nocol nopct;
run;


Statistics for Table of INC by POL
Statistic DF Value Prob
 Chi-Square 4 6.6995 0.1526
Likelihood Ratio Chi-Square 4 Mantel-Haenszel Chi-Square
1
6.7040 0.1524 3.7273 0.0535
```

(b) Is there a relationship between gender and feelings about the economic outlook? Turn in the appropriate statistic, p-value and the percentage of men and of women that feel that the outlook “will be better”.

Solution:
H_0 :Gender and feelings= about economic outlook are independent.

From the output the Pearson goodness of fit test statistic X^2 is 4.204 with pvalue=0.04. Therefore, we can reject the null hypothesis indicating that gender and feelingsabout economic outlook are not independent.

58.6% of men feel like the outlook will be better and 73.2% of women feel like the outlook will be better.

R Code

```
> ### Create two-way table Sex x Stat, excluding missing data
> sex_stat=table(Sex,Stat, exclude=".", dnn=list("Sex","Stat"))
> sex_stat.data=as.data.frame(sex_stat)
> sex_stat.data
> temp=xtabs(Freq~Sex+Stat,sex_stat.data)
> temp
   Stat
Sex  0  1
```

```
  0 58 41
  1 60 22
> result<-chisq.test(temp, correct=FALSE)
> result
      Pearson's Chi-squared test
data:  temp
X-squared = 4.2045, df = 1, p-value = 0.04032
> RowSums=rowSums(temp)
> RowPercentage=100*
rbind(temp[1,]/RowSums[1],temp[2,]/RowSums[2])
> RowPercentage
            0        1
[1,] 58.58586 41.41414
[2,] 73.17073 26.82927
```

SAS Code:

```
proc freq data=Montana order=data;
tables sex*stat / chisq cmh relrisk expected
nocol nopct;
run;
```

```
    Statistics for Table of SEX by STAT

Statistic DF Value Prob  Chi-Square 1 4.2045
0.0403
 Likelihood Ratio Chi-Square 1 4.2552 0.0391
 Mantel-Haenszel Chi-Square 1 4.1812 0.0409
```

(c) What can you say about associations between gender, economic outlook and party affiliation? Does the model of complete (mutual) independence fit? State three possible conditional independence models for these three variables. Do any of these conditional independence model fit? Do your conclusions from (b) change in any way?

Solution:
H_0: the model of complete independence fit (intercept model).

The Pearsons $X^2 = 14.47$ with p-value=0.0435, therefore reject null and conclude that the model of complete independence does not fit. Hence, we move on to the conditional independence.

There are three possible conditional independence models:

(1) Sex and politics are conditionally independent given state economic outlook. The CMH statistic is 3.07 with p-value 0.0796, therefore fail to reject null model, which suggests that Sex and politics are conditionally independent given state economic outlook at 0.05 level. From the output the conditional odds ratio have the same direction so the results of CMH are trustworthy.

```
> pol=table(Sex,Stat,Pol, exclude=".",
dnn=list("Sex","Stat","Pol"))
> pol.data=as.data.frame(pol)
> pol.data
> temp=xtabs(Freq~Sex+Pol+Stat,pol.data)
> temp
> temp
> temp
, , Stat = 0

   Pol
Sex  1  2  3
  0 21 16 20
  1 27 13 16

, , Stat = 1

   Pol
Sex  1  2  3
  0 12  7 21
  1 12  1  9


> thetaGP_E=oddsratio(temp[,,1],log=FALSE)
> thetaGP_E
[1] 0.6319444
> thetaGP_E2=oddsratio(temp[,,2],log=FALSE)
> thetaGP_E2
[1] 0.1428571
```

>

SAS/OUTPUT CODE

```
PROC FREQ DATA=montana;
FORMAT AGE AGEFMT. SEX SEXFMT. INC INCFMT. POL POLFMT.
AREA AREAFMT. FIN FINFMT. STATE STATEFMT. ;
TABLES SEX*POL*STATE/CHISQ NOROW NOCOL NOPERCENT EXPECTED DEVIATI
CELLCHI2 MEASURES FISHER CMH;
RUN;
```

The SAS System

The FREQ Procedure

Summary Statistics for SEX by POL
Controlling for STATE

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	3.1139	0.0776
2	Row Mean Scores Differ	1	3.1139	0.0776
3	General Association	2	4.4149	0.1100

Effective Sample Size = 175
Frequency Missing = 34

WARNING: 16% of the data are missing.

(2) Sex and state economic outlook are conditionally independent given politics. The CMH statistic is 4.415 with p-value 0.11, therefore fail to reject null model, which suggests that Sex and state economic outlook are conditionally independent given politics at 0.05 level. Also the output below shows that the conditional odds ratio between sex and state for each politic have the same direction. So the result of the CMH statistic are trustworthy.

```
The SAS System

The FREQ Procedure

Summary Statistics for SEX by STATE
Controlling for POL

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic    Alternative Hypothesis    DF       Value      Prob
---------------------------------------------------------------
    1        Nonzero Correlation        1      3.0739    0.0796
    2        Row Mean Scores Differ     1      3.0739    0.0796
    3        General Association        1      3.0739    0.0796


Estimates of the Common Relative Risk (Row1/Row2)

Type of Study     Method              Value     95% Confidence Limits
----------------------------------------------------------------------
Case-Control      Mantel-Haenszel    0.5527      0.2858        1.0688
  (Odds Ratio)    Logit              0.5762      0.2935        1.1309

Cohort            Mantel-Haenszel    0.8226      0.6657        1.0166
  (Col1 Risk)     Logit              0.8096      0.6631        0.9885

Cohort            Mantel-Haenszel    1.4602      0.9456        2.2550
  (Col2 Risk)     Logit              1.3813      0.8970        2.1273


     Breslow-Day Test for
Homogeneity of the Odds Ratios
------------------------------
Chi-Square              1.5773
DF                           2
Pr > ChiSq              0.4544


  Effective Sample Size = 175
    Frequency Missing = 34

WARNING: 16% of the data are missing.


The FREQ Procedure

Table 2 of SEX by STATE
Controlling for POL=INDEP

SEX              STATE

Frequency       |
Expected        |
Deviation       |
Cell Chi-Square |BETTER  |NOT BETT|  Total
----------------+--------+--------+
MALE            |     16 |      7 |     23
                | 18.027 |  4.973 |
                | -2.027 |  2.027 |
                | 0.2279 | 0.8262 |
----------------+--------+--------+
FEMALE          |     13 |      1 |     14
                | 10.973 |  3.027 |
                |  2.027 | -2.027 |
                | 0.3745 | 1.3574 |
----------------+--------+--------+
Total                 29        8       37

Frequency Missing = 3
```

R Code

```
###Conditional Model Test (GP, EP)
> pol=table(Sex,Stat,Pol, exclude=".",
```

```
dnn=list("Sex","Stat","Pol"))
> pol.data=as.data.frame(pol)
> pol.data
> temp=xtabs(Freq~Sex+Stat+Pol,pol.data)
> temp
> temp
, , Pol = 1

   Stat
Sex  0  1
  0 21 12
  1 27 12

, , Pol = 2

   Stat
Sex  0  1
  0 16  7
  1 13  1

, , Pol = 3

   Stat
Sex  0  1
  0 20 21
  1 16  9



> thetaGE_P1=oddsratio(temp[,,1],log=FALSE)
> thetaGE_P2=oddsratio(temp[,,2],log=FALSE)
> thetaGE_P3=oddsratio(temp[,,3],log=FALSE)
> thetaGE_P1
[1] 0.7777778
> thetaGE_P2
[1] 0.1758242
> thetaGE_P3
[1] 0.5357143
>

> mantelhaen.test(temp,correct=F)
        Mantel-Haenszel chi-squared test
```

```
without continuity correction
data:  temp
Mantel-Haenszel X-squared = 3.0739, df = 1, p-value = 0.07956
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.2857584 1.0688340
sample estimates:
common odds ratio
        0.5526557
```

(3) Politics and state economic outlook are conditionally independent given sex. The CMH statistic is 5.9 with p-value 0.0523, therefore fail to reject null model, which suggests that politics and state economic outlook are conditionally independent given sex at 0.05 level.

Yes the conclusion change from 4b. In 4b gender and outlook are marginally dependent but 4c shows after controlling for party affiliation they become independent. This is another example that shows marginal independence and conditional independence are not equivalent.

The SAS System

The FREQ Procedure

Summary Statistics for POL by STATE
Controlling for SEX

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	1.4465	0.2291
2	Row Mean Scores Differ	2	5.9022	0.0523
3	General Association	2	5.9022	0.0523

Effective Sample Size = 175
Frequency Missing = 34

WARNING: 16% of the data are missing.

SAS Code:

```
proc freq data=Montana order=internal;
tables sex*stat*pol / chisq cmh;
tables pol*sex*stat / chisq cmh;
tables stat*pol*sex / chisq cmh;run;
```

```
Summary Statistics for SEX by STAT
 Controlling for POL\\

Cochran-Mantel-Haenszel Statistics
 (Based on Table Scores)\\

Statistic Alternative Hypothesis DF Value Prob

Nonzero Correlation 1  3.0739 0.0796
Row Mean Scores Differ  1  3.0739   0.0796
General Association 1 3.0739 0.0796
Effective Sample Size = 175 Frequency Missing = 34
WARNING: 16% of the data are missing.

The SAS System The FREQ Procedure

Summary Statistics for SEX by POL

Controlling for STAT

Cochran-Mantel-Haenszel Statistics
 (Based on Table Scores)

Statistic Alternative Hypothesis DF Value Prob
1 Nonzero Correlation 1  3.1139 0.0776
2 Row Mean Scores Differ 13.1139 0.0776
3 General Association 2 4.4149 0.1100
Effective Sample Size = 175 Frequency Missing = 34
WARNING: 16% of the data are missing.
```

```
The FREQ Procedure

Summary Statistics for STAT
by POL Controlling for SEX

Cochran-Mantel-Haenszel Statistics
(Based on Table Scores)
Statistic Alternative Hypothesis DF Value Prob
1 Nonzero Correlation 1  1.4465 0.2291
2 Row Mean Scores Differ 1 1.4465 0.2291
3 General Association 2 5.9022 0.0523
```