Exercise 7.4 [16 points]

a. [3 points]

(A: Age, G: Gestation, I: Infant Survival, S: Smoking.)

Model	G^2	d.f.
(AGIS)	.008$\approx$ 0	0
(AGI, AIS, AGS, GIS)	.367	1
(AG, AI, AS, GI, GS, IS)	1.727	5
(A, G, S, I)	377.789	11

According to G^2 and degrees of freedom, the no-4-way-interaction model and the no-3-way-interaction model are not significantly different from the saturated model. The subset of models that should be further studied lies in between the independence model and the no-3-way-interaction model, such as (AG, S, I), (AS, G, I)...(AG, AI, AS, GI, GS).

b. [7 points]

(Brown's test for marginal association)

Model	G^2	p-value
(A, S)	17.609	.000
(A, G)	7.023	.008
(A, I)	10.125	.002
(S, G)	.516	.473
(S, I)	2.391	.122
(G, I)	342.328	.000

The goal of Brown's test in marginal association is to test the 2-way associations that is as "the most complex parameter in a simple model". For instance, if collapsing over G&I. i.e. the model (A, S) is significantly unfit and should be rejected, then we know the term (AS) has to be included in the eventual model (Age and Smoking show a marginal association). Similarly, we find AG, AI, & GI as significant 2-way marginal associations. Data do not provide evidence of marginal associations between Smoking and Gestation (SG) and Smoking and Infant Survival (SI).

Therefore, the marginal association test suggest that the 2-way-associations involving Age, and the associations between Gestation and Infant survival should be retained in the model (i.e. AG, AI, AS, GI)

(Brown's test for partial association)

Model	Term dropped	G^2	significance?
(AG, AI, AS, GI, GS, IS)		1.727	
(AI, AS, GI, GS, IS)	AG	4.703	n
(AG, AS, GI, GS, IS)	AI	8.180	*
(AG, AI, GI, GS, IS)	AS	19.969	*
(AG, AI, AS, GS, IS)	GI	339.336	*
(AG, AI, AS, GI, IS)	GS	1.828	n
(AG, AI, AS, GI, GS)	IS	4.055	n

The goal of Brown's test in partial association is to find the associations that can be dropped from the more specified model in order to achieve parsimony. One can do so by detecting the G^2 difference to see if dropping a certain terms results in significant unfit compared with the saturated model. For two nested models, we can use the relative G^2 to detect such a difference as well, e.g. dropping AG (Age and Gestation)

$$G^2[(AI, AS, GI, GS, IS)|(AG, AI, AS, GI, GS, IS)] = 4.703 - 1.727 = 2.976 < G^2_{\alpha=.05,df=1} = 3.84$$

is not significant. Similarly, GS and IS were also found as terms that may be dropped from the model.

Therefore, the partial association test suggest that within the no-3-way-interaction model, AI, AS, and GI should be retained in the model, and AG, GS, and IS may be considered dropped. That is, the model to consider is (AI, AS, GI). However, According to Brown (1976), the term of either test in marginal or partial association is significant should be included (more conservative). Hence, the model to consider according to Brown's method is (AG, AI, AS, GI).

c. [3 points]

```
. loglin count age smoke infant gesta, fit (age smoke infant gesta)
Variable age = A
Variable smoke = B
Variable infant = C
Variable gesta = D
Margins fit: age smoke infant gesta
Note: Regression-like constraints are assumed.  The first level of each
variable (and all interactions with it) will be dropped from estimation.

Iteration 0: Log Likelihood = -46.179688
Iteration 1: Log Likelihood = -45.988281

Poisson regression                                Number of obs    =       16
Goodness-of-fit chi2(0)     =      0.008          Model chi2(15)   =20311.062
Prob > chi2                 =          .          Prob > chi2      =   0.0000
Log Likelihood              =    -45.988          Pseudo R2        =   0.9955
```

```
------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -.198451   .2106899     -0.942   0.346      -.6113957    .2144936
    AB22 | -.6124792   .6367854     -0.962   0.336      -1.860556    .6355973
    AC22 | -.5636888   .2331685     -2.418   0.016      -1.020691   -.1066868
    AD22 | -.3405454   .3968444     -0.858   0.391      -1.118346    .4372553
  ABC222 |  .0836349   .7289596      0.115   0.909        -1.3451    1.512369
  ABD222 | -.6402806   1.297197     -0.494   0.622       -3.18274    1.902179
  ACD222 |  .1796424   .4102919      0.438   0.662      -.6245148    .9837997
ABCD2222 |   .783399   1.348972      0.581   0.561      -1.860538    3.427336
      B2 | -1.714798   .3620925     -4.736   0.000      -2.424487    -1.00511
    BC22 | -.3488946    .399106     -0.874   0.382      -1.131128    .4333388
    BD22 |  .3285041   .5826178      0.564   0.573      -.8134058    1.470414
  BCD222 | -.4328055   .6083141     -0.711   0.477      -1.625079    .7594682
      C2 |  1.840549   .1522321     12.090   0.000        1.54218    2.138919
    CD22 |  3.278441   .2551284     12.850   0.000       2.778399    3.778484
      D2 | -.7339692   .2483277     -2.956   0.003      -1.220683   -.2472558
   _cons |  3.912023   .1414214     27.662   0.000       3.634842    4.189204
------------------------------------------------------------------------------
```

Goodman's strategy for model building uses significantly non-zero standardized parameter estimates from the saturated model. Give estimates from the saturated model satisfies such conditions ($|z| > 1.96$): AC22, B2, C2, CD22, & D2, i.e., AI, S, G, GI, & I which simplifies to the model (AI, GI, S) as the starting point and build forward.

d. [3 points]

From the above, we know these strategies are consistent regarding which particular two-way associations should be retained in the model, i.e., GI & AI. In (a), we know the upper and lower limits of model building (no-3-way-interaction model and independence model, respectively). Yet there is still a large subset of models in between that need to be explored. Goodman's method gives a good starting point, though the significance of the standardized estimates at times depend on the higher order interactions. It is more reliable to supplement with Brown's method to examine both marginal and partial associations between any two variables. If either the marginal or conditional association is significant, then it should be considered in a preliminary model. We then will have ample information regarding which terms should be retained/dropped toward building a best fitting model for the data.

Exercise 7.12 [5 points]

(A: SexIQ, B: Residence, C: SES, D: Occ. Aspiration.)

Model	G^2	d.f.
(ABCD)	0	0
(ABC, ABD, ACD, BCD)	3.730	6
(AB, AC, AD, BC, BD, CD)	59.250	29
(A, B, C, D)	2626.371	40

Due to the limits of STATA, we recode the sex and IQ into one variable: SexIQ, where 1 and 2 stands more male high and male low, and 3 and 4 stands for female high and female low, correspondingly. From the above, we know the best fitting model will fall between the no-4-way-interaction model and the no-3-way-interaction model. Therefore, we use the stepwise procedure to look for the best fitting model.

Using the backward-stepwise selection, we can first break down the no-4-way-interaction into a set of 3-way-interaction models. At this stage, we find the model without the association between SexIQ, Residence, and SES is the fitting model. It tells us that the association between those three variables is not important. We then break further into subsets of models, we find that the association between SexIQ, Residence, and Occupational Aspirations is neither important. The model (AC, AB, AD, BCD) has a $\frac{\Delta G^2}{\Delta d.f} \approx 1$, so we should pay attention if there is any better fitting model (otherwise we may stop here).

Although (AC, AB, AD, BCD) seems to be the best fitting model, (AC, AD, BCD) actually fit adequately with a difference of 6 degrees of freedom. In order to answer the "best" model, these two models should both be considered. Yet by examining the standardized parameter estimates for these two models, we find most of the 3-way-interaction terms are not significantly different from zero but one of the SexIQ, Residence, and Occupational Aspiration interaction. We know there is evidence for BCD interaction, and (AC, AB, AD, BCD) is actually an adequate model that is easier to interpret.

(A: SexIQ, B: Residence, C: SES, D: Occ. Aspiration.)

Model	G^2	d.f.	Best model
(ABC, ABD, ACD, BCD)	3.730	6	
(ABC, ABD, ACD)	15.422	8	
(ABC, ABD, BCD)	13.977	9	
(ABC, ACD, BCD)	16.590	12	
(ABD, ACD, BCD)	5.715	12	*
(AB, ACD, BCD)	20.602	18	
(AC, ACD, BCD)	16.059	15	*
(BC, ACD, BCD)	19.691	14	
(AC, BC, CD, ABD)	29.662	17	
(AC, AB, AD, BCD)	29.324	21	*
(ABD, BCD)	153.762	18	
(AB, AD, BCD)	169.816	24	
(AC, AD, BCD)	42.617	27	*
(AB, AC, BCD)	635.484	24	
(AB, AC, AD, BC, BD)	59.250	29	

Exercise 7.13 [3 points]

```
. gen sexiq=2*sex+iq

. xi: logit i.aspire i.sexiq i.reside i.ses [freq=count]
i.aspire          Iaspir_1-2   (naturally coded; Iaspir_1 omitted)
i.sexiq           Isexiq_3-6   (naturally coded; Isexiq_3 omitted)
i.reside          Iresid_1-3   (naturally coded; Iresid_1 omitted)
i.ses             Ises_1-2     (naturally coded; Ises_1 omitted)

Iteration 0:  Log Likelihood =-3057.4304
Iteration 1:  Log Likelihood =-2238.8513
Iteration 2:  Log Likelihood =-2203.1737
Iteration 3:  Log Likelihood =-2202.3948
Iteration 4:  Log Likelihood =-2202.3942

Logit Estimates                                       Number of obs =   4511
                                                      chi2(6)       =1710.07
                                                      Prob > chi2   = 0.0000
Log Likelihood = -2202.3942                           Pseudo R2     = 0.2797

------------------------------------------------------------------------------
Iaspir_2 |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Isexiq_4 |   1.828203    .109107     16.756   0.000       1.614357    2.042049
Isexiq_5 |   .3864962   .0962878      4.014   0.000       .1977756    .5752168
Isexiq_6 |   2.118888   .1097749     19.302   0.000       1.903734    2.334043
Iresid_2 |  -.3770257   .0857494     -4.397   0.000      -.5450913    -.20896
Iresid_3 |    -.52333   .1108982     -4.719   0.000      -.7406865   -.3059734
  Ises_2 |   1.701922   .0772408     22.034   0.000       1.550533    1.853311
   _cons |  -1.105538   .0977495    -11.310   0.000      -1.297123   -.9139526
------------------------------------------------------------------------------

. loglin count sexiq reside ses aspire, fit(sexiq reside ses, sexiq aspire, reside aspire, ses aspire)
Variable sexiq = A
```

```
Variable reside = B
Variable ses = C
Variable aspire = D
Margins fit: sexiq reside ses, sexiq aspire, reside aspire, ses aspire
Note: Regression-like constraints are assumed.  The first level of each
variable (and all interactions with it) will be dropped from estimation.

Iteration 0: Log Likelihood = -162.76953
Iteration 1: Log Likelihood = -162.13477
Iteration 2: Log Likelihood = -162.13281

Poisson regression                                 Number of obs    =      48
Goodness-of-fit chi2(17)     =    37.266           Model chi2(30)   =3452.203
Prob > chi2                  =    0.0031           Prob > chi2      =  0.0000
Log Likelihood               =  -162.133           Pseudo R2        =  0.9141

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |   -1.25947   .1404916     -8.965   0.000      -1.534829   -.9841116
      A3 |  -.0691557   .1128544     -0.613   0.540      -.2903463    .1520349
      A4 |   -1.48299   .1447751    -10.243   0.000      -1.766744   -1.199236
    AB22 |  -.2282514   .1560691     -1.463   0.144      -.5341412    .0776384
    AB23 |  -.5201494   .2014464     -2.582   0.010      -.9149772   -.1253217
    AB32 |   -.005287   .1290766     -0.041   0.967      -.2582726    .2476985
    AB33 |   .0001114   .1518004      0.001   0.999       -.297412    .2976348
    AB42 |  -.2277835     .15866     -1.436   0.151      -.5387515    .0831844
    AB43 |  -.2902763   .1966152     -1.476   0.140       -.675635    .0950824
    AC22 |   .5166361   .1671579      3.091   0.002       .1890127    .8442596
    AC32 |  -.0035608   .1625337     -0.022   0.983       -.322121    .3149993
    AC42 |   .7222694   .1655914      4.362   0.000       .3977163    1.046823
   ABC222 |   .2646456   .2084411      1.270   0.204      -.1438915    .6731826
   ABC232 |    .199704   .2911196      0.686   0.493      -.3708798    .7702879
```

```
  ABC322 |  -.2350197   .2072797     -1.134   0.257     -.6412804     .171241
  ABC332 |  -.0328167   .2715123     -0.121   0.904      -.564971    .4993376
  ABC422 |   .2261748     .20629      1.096   0.273     -.1781462    .6304958
  ABC432 |   .3755039   .2720379      1.380   0.167     -.1576807    .9086884
    AD22 |   1.828203    .109107     16.756   0.000      1.614358    2.042049
    AD32 |   .3864963   .0962878      4.014   0.000      .1977757     .575217
    AD42 |   2.118889    .109775     19.302   0.000      1.903733    2.334044
      B2 |   1.045257   .0939646     11.124   0.000      .8610896    1.229424
      B3 |   .2109859   .1101362      1.916   0.055     -.0048771    .4268488
```

```
    BC22 |  -.7242662   .1518347     -4.770   0.000     -1.021857   -.4266757
    BC32 |  -.9429092   .2009274     -4.693   0.000      -1.33672   -.5490987
    BD22 |  -.3770256   .0857494     -4.397   0.000     -.5450913   -.2089599
    BD32 |  -.5233299   .1108983     -4.719   0.000     -.7406865   -.3059732
      C2 |  -.9003063   .1250007     -7.202   0.000     -1.145303   -.6553095
    CD22 |   1.701922   .0772408     22.034   0.000      1.550533    1.853311
      D2 |  -1.105538   .0977495    -11.310   0.000     -1.297124   -.9139527
   _cons |   4.813911   .0817837     58.862   0.000      4.653618    4.974205
------------------------------------------------------------------------------
```

We find the correspondence between the logit and loglinear models:

Logit model	Loglinear Model	Coefficient
Isexiq1	AD22	1.818
Isexiq2	AD32	.386
Isexiq3	AD42	2.119
Iresid2	BD22	-.377
Iresid3	BD32	-.523
Ises2	CD22	1.702

One can derive such a relationship from Agresti's or the lecture notes. To interpret the coefficient, e.g. holding other variables constant, people of lower SES are $e^1.702 = 5.48$ times as likely to have lower occupational aspirations (compared with those of high SES).

Exercise 8.1 [6 points]

```
. loglin count ses mental, fit(ses, mental) keep resid
Variable ses = A
Variable mental = B
Margins fit: ses, mental
Note: Regression-like constraints are assumed.  The first level of each
variable (and all interactions with it) will be dropped from estimation.

Iteration 0: Log Likelihood = -96.600586
Iteration 1: Log Likelihood = -95.795898
Iteration 2: Log Likelihood = -95.79541

Poisson regression                                Number of obs   =        24
Goodness-of-fit chi2(15)    =     47.418          Model chi2(8)   =   169.982
Prob > chi2                 =     0.0000          Prob > chi2     =    0.0000
Log Likelihood              =    -95.795          Pseudo R2       =    0.4701

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -.0670861   .0888731     -0.755   0.450     -.2412742    .1071019
      A3 |   .0911379   .0854466      1.067   0.286     -.0763343    .2586101
      A4 |   .3822983   .0801309      4.771   0.000      .2252446     .539352
      A5 |   .0113854   .0871228      0.131   0.896     -.1593721    .1821429
      A6 |   -.188447   .0917883     -2.053   0.040     -.3683487   -.0085452
      B2 |   .6734098   .0701317      9.602   0.000      .5359541    .8108654
      B3 |   .1647965   .0775871      2.124   0.034      .0127286    .3168645
      B4 |   .2367316   .0763415      3.101   0.002       .087105    .3863581
   _cons |   3.880619    .080447     48.238   0.000      3.722946    4.038292
------------------------------------------------------------------------------

. tabdisp ses mental, c(stdres) f(%5.2f)

----------+---------------------------
          |          mental
      ses |     1      2      3      4
----------+---------------------------
        1 |  2.23  -0.10   0.11  -1.96
        2 |  1.74   0.55   0.08  -2.30
        3 |  0.54   0.09   0.31  -0.88
        4 |  0.12   0.15  -0.74   0.42
        5 | -1.86   0.09  -0.50   2.02
        6 | -3.02  -0.87   0.97   2.83
----------+---------------------------

. gen u=ses-3.5
```

```
. gen v=mental-2.5
. gen uv=u*v

. poisson count A2-A6 B2-B4 uv

Iteration 0: Log Likelihood = -77.143066
Iteration 1: Log Likelihood = -77.03418

Poisson regression                                Number of obs   =        24
Goodness-of-fit chi2(14)    =      9.896          Model chi2(9)   =   207.505
Prob > chi2                 =     0.7698          Prob > chi2     =    0.0000
Log Likelihood              =    -77.034          Pseudo R2       =    0.5739

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -.0498916   .0890787     -0.560   0.575     -.2244827    .1246995
      A3 |   .1169211    .085985      1.360   0.174     -.0516064    .2854487
      A4 |   .4078719   .0808662      5.044   0.000       .249377    .5663668
      A5 |   .0278449   .0877785      0.317   0.751     -.1441978    .1998877
      A6 |  -.1900174   .0925081     -2.054   0.040       -.37133   -.0087048
      B2 |   .6963183   .0705891      9.864   0.000      .5579661    .8346705
      B3 |   .1895087   .0782725      2.421   0.015      .0360973    .3429201
      B4 |   .2423034   .0770897      3.143   0.002      .0912104    .3933964
      uv |   .0906866   .0150061      6.043   0.000      .0612751    .1200981
   _cons |   3.838777   .0792552     48.436   0.000       3.68344    3.994114
------------------------------------------------------------------------------
. predict linpred
. gen mhat=exp(linpred)
. gen ures=count-mhat
. gen pres=ures/sqrt(mhat)
. tabdisp ses mental, c(pres) f(%5.2f)

----------+---------------------------
          |          mental
      ses |     1      2      3      4
----------+---------------------------
        1 | -0.16  -1.02   1.11   0.59
        2 |  0.38  -0.10   0.58  -0.88
        3 |  0.15  -0.21   0.42  -0.28
        4 |  0.83   0.34  -1.01  -0.13
        5 | -0.47   0.79  -0.99   0.39
        6 | -1.21   0.26   0.27   0.27
----------+---------------------------
```

a. [2 points]

From the G^2 and residuals comparisons, we know that the uniform association model fits better than the independence model. This tells us that nominal tests for independence are usually more conservative, and not well suited to situations when ordinal variables are involved.

a. [2 points]

The estimate the $\hat{\beta}$ is .091 with ASE=.015. The positive value indicates that mental healthy status tend to go down as the level of parents' SES goes down. The estimated uniform local odds ratio $\hat{\theta} = e^{\cdot}091 = 1.01$ That is, the estimated odds ratio that mental health status is in category j+1 instead of j increases by a factor of 1.01 for each category change in parents' SES.

a. [2 points]

The difference in G^2 between the uniform association and the independence model is 47.418-9.896 $\approx$ 37.52 based on d.f.=1. Therefore, we can reject the hypothesis that the independence model is a better one. The z score on $\hat{\beta} = 6.043$ is also highly significant, and thus we can reject the hypothesis that $\hat{\beta} = 0$.

Exercise 8.10 [10 points]

a. [7 points]

```
. loglin count race edu job, fit(race, edu, job)
Variable race = A
Variable edu = B
Variable job = C
Margins fit: race, edu, job
Note: Regression-like constraints are assumed.  The first level of each
variable (and all interactions with it) will be dropped from estimation.

Iteration 0: Log Likelihood = -49.94043
Iteration 1: Log Likelihood = -49.35791
Iteration 2: Log Likelihood = -49.357422

Poisson regression                                Number of obs    =      16
Goodness-of-fit chi2(10)    =    19.177           Model chi2(5)    =1932.440
Prob > chi2                 =    0.0381           Prob > chi2      =  0.0000
Log Likelihood              =   -49.357           Pseudo R2        =  0.9514

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -1.999614   .0884887    -22.597   0.000      -2.173049    -1.82618
      B2 |  -1.392474   .0718257    -19.387   0.000      -1.533249   -1.251698
      C2 |  -.2108948   .0627254     -3.362   0.001      -.3338343   -.0879552
      C3 |  -1.562998   .1008163    -15.503   0.000      -1.760594   -1.365402
      C4 |  -2.108015   .1274885    -16.535   0.000      -2.357888   -1.858142
   _cons |   5.993237   .0455656    131.530   0.000        5.90393    6.082544
------------------------------------------------------------------------------

. loglin count race edu job, fit(race edu, race job, edu job) keep resid
Variable race = A
Variable edu = B
Variable job = C
Margins fit: race edu, race job, edu job
Note: Regression-like constraints are assumed.  The first level of each
variable (and all interactions with it) will be dropped from estimation.

Iteration 0: Log Likelihood = -42.991211
Iteration 1: Log Likelihood = -42.569336
Iteration 2: Log Likelihood = -42.566895

Poisson regression                                Number of obs    =      16
Goodness-of-fit chi2(3)     =     5.596           Model chi2(12)   =1946.021
Prob > chi2                 =    0.1330           Prob > chi2      =  0.0000
Log Likelihood              =   -42.567           Pseudo R2        =  0.9581

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -2.157309   .1469537    -14.680   0.000      -2.445333   -1.869285
    AB22 |  -.2890383    .240586     -1.201   0.230      -.7605782    .1825017
    AC22 |   .3149542   .1975463      1.594   0.111      -.0722294    .7021378
    AC23 |   .7117516   .2753841      2.585   0.010       .1720087    1.251494
    AC24 |    .014013   .4232366      0.033   0.974      -.8155155    .8435414
      B2 |  -1.290939   .1048053    -12.318   0.000      -1.496354   -1.085525
    BC22 |  -.0201457   .1546116     -0.130   0.896      -.3231789    .2828875
    BC23 |  -.4517713   .2821149     -1.601   0.109      -1.004706    .1011638
    BC24 |  -.4573135   .3572385     -1.280   0.200      -1.157488    .2428611
      C2 |  -.2425491   .0744698     -3.257   0.001      -.3885072    -.096591
      C3 |  -1.581965   .1195029    -13.238   0.000      -1.816186   -1.347743
      C4 |  -2.028717   .1454273    -13.950   0.000       -2.31375   -1.743685
   _cons |   5.995369   .0496402    120.776   0.000       5.898076    6.092662
------------------------------------------------------------------------------

count  race  edu  job   cellhat    resid   stdres
400     1     1    1   401.565   -1.565   -0.078
319     1     1    2   315.078    3.922    0.221
 81     1     1    3    82.550   -1.550   -0.171
 52     1     1    4    52.807   -0.807   -0.111
112     1     2    1   110.435    1.565    0.149
 81     1     2    2    84.922   -3.922   -0.426
 16     1     2    3    14.450    1.550    0.408
 10     1     2    4     9.193    0.807    0.266
 48     2     1    1    46.435    1.565    0.230
 46     2     1    2    49.922   -3.922   -0.555
 21     2     1    3    19.450    1.550    0.351
  7     2     1    4     6.193    0.807    0.324
  8     2     2    1     9.565   -1.565   -0.506
 14     2     2    2    10.078    3.922    1.235
  1     2     2    3     2.550   -1.550   -0.971
  0     2     2    4     0.807   -0.807   -0.899

. gen rj=race*job
. gen ej=edu*job
. gen rjXej=race*edu*job
. poisson count A2 B2 C2-C4 AB22 rj ej rjXej

Iteration 0: Log Likelihood = -46.285645
Iteration 1: Log Likelihood = -45.378418
Iteration 2: Log Likelihood = -45.350098

Poisson regression                                Number of obs    =      16
Goodness-of-fit chi2(6)     =    11.162           Model chi2(9)    =1940.455
Prob > chi2                 =    0.0835           Prob > chi2      =  0.0000
Log Likelihood              =   -45.350           Pseudo R2        =  0.9553

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -2.279729   .2212641    -10.303   0.000      -2.713398   -1.846059
      B2 |  -1.122783   .1739367     -6.455   0.000      -1.463692   -.7818728
      C2 |  -.3674873   .3991333     -0.921   0.357      -1.149774    .4147996
      C3 |  -1.883148   .7953456     -2.368   0.018      -3.441997   -.3242997
      C4 |   -2.59903   1.192107     -2.180   0.029      -4.935517   -.2625427
    AB22 |  -.0822611   .5481477     -0.150   0.881      -1.156611    .9920886
      rj |    .304929   .3409258      0.894   0.371      -.3632733    .9731314
      ej |  -.0186385   .3296427     -0.057   0.955      -.6647263    .6274493
   rjXej |  -.1200851   .2870627     -0.418   0.676      -.6827176    .4425474
   _cons |     5.8149    .375491     15.486   0.000       5.078951    6.550849
------------------------------------------------------------------------------

. display chiprob(6-3, 11.162-5.596)
.13474394

. poisson count A2-BC24 rjXej

Iteration 0: Log Likelihood = -42.999512
Iteration 1: Log Likelihood = -42.459473
Iteration 2: Log Likelihood = -42.445801

Poisson regression                                Number of obs    =      16
Goodness-of-fit chi2(2)     =     5.354           Model chi2(13)   =1946.263
Prob > chi2                 =    0.0688           Prob > chi2      =  0.0000
Log Likelihood              =   -42.446           Pseudo R2        =  0.9582

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      A2 |  -2.024772   .3104868     -6.521   0.000      -2.633315   -1.416229
      B2 |  -1.147698   .3133037     -3.663   0.000      -1.761762   -.5336344
      C2 |  -.0925508   .3177012     -0.291   0.771      -.7152336    .5301321
      C3 |  -1.282224   .6292847     -2.038   0.042        -2.5156   -.0488491
      C4 |  -1.575537   .9457176     -1.666   0.096      -3.429109    .2780356
    AB22 |  -.0229927   .5909228     -0.039   0.969       -1.18118    1.135195
    AC22 |   .4953336   .4219786      1.174   0.240      -.3317293    1.322396
    AC23 |   1.060361   .7676712      1.381   0.167      -.4442468    2.564969
    AC24 |   .5296205   1.135646      0.466   0.641      -1.696204    2.755445
```

```
    BC22 |   .1471041   .3782632      0.389   0.697      -.5942781    .8884863
    BC23 |  -.1077845   .7602916     -0.142   0.887      -1.597929     1.38236
    BC24 |   .0372144    1.07399      0.035   0.972      -2.067767    2.142196
   rjXej |  -.1530414   .3152663     -0.485   0.627      -.7709521    .4648692
   _cons |   6.150516   .3233923     19.019   0.000       5.516679    6.784354
```

```
------------------------------------------------------------------------------

. display chiprob(6-2, 11.162-5.354)
.21395314
```

By fitting the independence model and the partial association model, we know that the the independence model doesn't fit well ($G^2 = 19.177, d.f. = 10$), and the partial association model does ($G^2 = 5.596, d.f. = 3$). Thus, the best fitting model must fall between these two models. Since we were asked to use the model discussed in Chapter 8 and the variables are ordered, we need to test the uniform association model and the uniform interaction model. The uniform association model ($G^2 = 10.604, d.f. = 5$) and the uniform interaction model ($G^2 = 11.162, d.f. = 6$) both fit adequately well, because their fits are not significantly from the saturated model. However, by comparing with the partial association model, neither the uniform association model ($G^2 = 10.604 - 5.596, d.f. = 5 - 3 = 2$) nor the uniform interaction model ($G^2 = 11.162 - 10.604, d.f. = 6 - 3$) has a significant improvement in the model. Therefore, we should conclude the data do not give evidence to models beyond the partial association model. By examine the Pearson‘s residuals, there was no significantly large residuals in any cell. However, the coefficients on high school degree and job satisfaction (BCs) are all non-significant. We should further explore if we had included an irrelevant predictor in the study of question at hand.

b. [3 points]

```
. ologit job race edu [freq=count], table

Iteration 0:  Log Likelihood =-1354.0814
Iteration 1:  Log Likelihood =-1350.7031
Iteration 2:  Log Likelihood =-1350.7011
Iteration 3:  Log Likelihood =-1350.7011

Ordered Logit Estimates                                 Number of obs =   1216
                                                        chi2(2)       =   6.76
                                                        Prob > chi2   = 0.0340
Log Likelihood = -1350.7011                             Pseudo R2     = 0.0025

------------------------------------------------------------------------------
     job |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    race |   .3448801   .1639127      2.104   0.035        .023617    .6661432
     edu |   -.202547   .1357431     -1.492   0.136      -.4685986    .0635045
---------+--------------------------------------------------------------------
   _cut1 |   .0098889   .2543826          (Ancillary parameters)
   _cut2 |    1.84938   .2617287
   _cut3 |    2.96403   .2786193
------------------------------------------------------------------------------

     job |      Probability           Observed
---------|-------------------------------------
       1 |  Pr(      xb+u<_cut1)        0.4671
       2 |  Pr(_cut1<xb+u<_cut2)        0.3783
       3 |  Pr(_cut2<xb+u<_cut3)        0.0979
       4 |  Pr(_cut3<xb+u)              0.0567
```

```
. ologit job race edu inter [freq=count], table

Iteration 0:  Log Likelihood =-1354.0814
Iteration 1:  Log Likelihood =-1350.7011
Iteration 2:  Log Likelihood =-1350.6988
Iteration 3:  Log Likelihood =-1350.6988

Ordered Logit Estimates                                 Number of obs =   1216
                                                        chi2(3)       =   6.77
                                                        Prob > chi2   = 0.0798
Log Likelihood = -1350.6988                             Pseudo R2     = 0.0025

------------------------------------------------------------------------------
     job |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    race |   .3792014   .5285167      0.717   0.473      -.6566723    1.415075
     edu |  -.1701436   .4934271     -0.345   0.730      -1.137243    .7969557
   inter |  -.0290651   .4255571     -0.068   0.946      -.8631417    .8050115
---------+--------------------------------------------------------------------
   _cut1 |   .0482149   .6160232          (Ancillary parameters)
   _cut2 |   1.887733   .6194843
   _cut3 |   3.002446   .6276322
------------------------------------------------------------------------------

     job |      Probability           Observed
---------|-------------------------------------
       1 |  Pr(      xb+u<_cut1)        0.4671
       2 |  Pr(_cut1<xb+u<_cut2)        0.3783
       3 |  Pr(_cut2<xb+u<_cut3)        0.0979
       4 |  Pr(_cut3<xb+u)              0.0567
```

From the equation: $Logit(\pi_j) = \alpha_j + \beta_j$, we know that the output has presented cut1=α_1, cut2=α_2, cut3=α_3 and β_{race}&β_{edu}. One can derive log odds ratios, odds ratios, and estimated probability for each combination of categories, and compare with 8.10. The insignificant coefficient of education (HS degree) corresponds to the 3 insignificant BC coefficients in the no-3-way-interaction loglinear model. We also find the ologit model with race*edu interaction term does not fit well, and it corresponds to the previous unfit uniform interaction loglinear model.