HW 2, Soc 388

Due Monday, October 15, in class

Consider the Los Angeles intermarriage dataset:

Intermarriage, LA 1990

	Wives				
Husbands:	NH Black	Mexican	Other Hisp	All Others	NH White
Non Hisp Black	4074	63	32	42	215
Mexican	25	3947	143	95	1009
Other Hispanic	16	132	239	18	304
All Others	19	78	18	1022	360
Non Hisp White	103	1156	373	492	28453

Fill in the following Table

Model #	Model Description	Terms in model	Residual df	Goodness of fit Chisquare	P	BIC	ID
1	Constant only	1	24	160,099.2	0	159843.5	73.97
2	Independence Model	9	16	43,952.7	0	43782.2	34.8
3	Independence plus single level of endogamy (same for all groups)	10	15	2,632.7	0	2472.9	5.29
4	Independence plus separate endogamy term for each group	14	11	87.9	0	-29.3	0.41
5	Same as 4, plus Black- White and Mexican- Other Hispanic interactions	16	9	20.6	.014	-75.3	0.13
6	Model 5 plus a gender symmetric Mexican- Black interaction.	17	8	6.30	0.61	-78.9	0.087
7	Another good fitting model, the Quasi-Symmetry model	19	6	1.38	0.97	-62.6	0.036
8							

1) Fill in the above table, models 1-5

2) Verify that model 1, the 'constant' model is the comparison model for the likelihood ratio chisquare that Stata lists as the second line of output for each subsequent model. How do you interpret that chisquare test?

What Stata lists as LR Chi2 is just a test between the model in question, and the Constant Only model. So, for the Independence model, Model 2, this test has 9-1=8 degrees of freedom, and a goodness of fit chisquare of 160,099- 43,953=116,146. What this means is that the independence model fits a LOT better than the constant only model (an improvement in goodness of fit of more than 100,000 using only 8 additional degrees of freedom), which of course is not surprising.

3) Does racial endogamy vary significantly between groups? What is the statistical test that answers that question?

Yes, racial endogamy varies very significantly between groups. The key test of this is the nested chisquare test between models 3 and 4. The test has 4 degrees of freedom (14-10) and a chisquare of 2544.8 (2632.7-87.9). Model 4 is a big improvement on Model 3 (reducing the goodness of fit chisquare by more than 2000 on only 4 additional degrees of freedom).

4) In model 4 which is the group with the strongest ethnic or racial endogamy? Which group has the weakest endogamy? Is the difference between the strongest and weakest statistically significant?

Blacks have the highest endogamy, with a log odds ratio of 6.38. 'Other Hispanics' have the lowest endogamy, with a log odds ratio of 1.94. Given the small standard errors of each coefficient, it seems likely that they are significantly different. In order to know for sure whether the two coefficients are significantly different you need to know not only the standard error of each, but also their covariance. Using lincom, the difference between the two coefficients is 4.44 and the standard error of this difference is 0.134, so coef/SE=Z=33.2 which definitely rejects the null hypothesis that the two coefficients are the same. Clearly black endogamy is force unto itself. In fact the difference between black endogamy and the endogamy of the second strongest group is the residual 'all others' group which for LA is nearly all Asian. Their endogamy log coefficient is 3.24; the difference between black endogamy and 'all other' endogamy is 6.38-3.24=3.14 with a SE of .120 and a Z score of 26.

5) Generate the predicted values for Model 5. Where do the predicted values and the actual values correspond exactly?

In the marginals, of course, and also along the endogamy diagonal. But also note that the sum of the two black-white cells and the sum of the two Mexican-other Hispanic cells is exactly the same in the predicted and the actual data (the one term fixes the total for the two cells).

```
. table BW, contents (sum count sum  P_m5)
```

```
----------------------------------
       BW | sum(count)   sum(P_m5)
----------+-----------------------
        0 |      42110       42110
        1 |        318         318
----------------------------------
```

```
. table MOh, contents (sum count sum  P_m5)

----------------------------------
      MOh | sum(count)   sum(P_m5)
----------+-----------------------
        0 |      42153       42153
        1 |        275         275
----------------------------------
```

```
table husb wife, contents(sum count sum  P_m5) row col

------------------------------------------------------------------------------
           |                                wife
      husb |      black    Mexican   Oth Hisp All Others      White      Total
-----------+------------------------------------------------------------------
     black |       4074         63         32         42        215       4426
           |       4074   79.17728   25.80177   34.52584   212.4951       4426
           |
   Mexican |         25       3947        143         95       1009       5219
           |   34.56537       3947   143.2581   86.88316   1007.293       5219
           |
  Oth Hisp |         16        132        239         18        304        709
           |   10.35845   131.7419        239   26.03689   301.8628        709
           |
All Others |         19         78         18       1022        360       1497
           |   12.57129    72.4654   23.61455       1022   366.3488       1497
           |
     White |        103       1156        373        492      28453      30577
           |   105.5049   1145.615   373.3256   499.5541      28453      30577
           |
     Total |       4237       5376        805       1669      30341      42428
           |       4237       5376        805       1669      30341      42428
------------------------------------------------------------------------------
```

6) How do you interpret the coefficients for Black- White and Mexican- Other Hispanic intermarriage in Model 5?

The Black- White coefficient is -0.633. This means that the odds of marrying a White person are $e^{-.633}=0.53$, which means that for Blacks the odds of marrying a White person are about half as high as the odds of a Black person marrying someone from another group (Mexicans, Asians, Other Hispanics). As we discussed in class, the comparison group is really all of the off-diagonal cells except for Black- White and Mexican- Other Hispanic, because the model has terms that specifically fit the endogamy diagonal. So the most correct way of explaining the Black- White

interaction is as follows: the odds of Black-White intermarriage are about half as all other kinds of non-endogamous marriages (except Mexican- Other Hispanic intermarriage).

7) If you add a gender specific dimension to Black- White intermarriage in Model 5, is it significant?

No, it is not. The goodness of fit chisquare improves from 20.6 to 20.4, an improvement of 0.2 on 1 df, not a significant improvement by the LRT and a significant setback by the BIC.

8) Make a new model 7, which consists of Model 2, the independence model, plus the gender symmetric Black- White interaction term. Compare the resulting Black- White interaction term to same term from Model 5. Why is it different? Think about how the comparison group is different in the two cases.

If we add the black-white term to model 2, we get a log coefficient of -3.925, or a relative risk of $e^{-3.925}$=0.0197, whereas in in model 5 the black-white interaction was a much more modest -0.633, or a relative risk of 0.531. In Model 2 the model does not fit the endogamy diagonal. So the comparison cases for Black- White intermarriage include the highly popular cells for Black- Black and White- White endogamy. That's why the coefficient for Black- White intermarriage is so sharply negative here; it's a question of comparing it to more popular options.

9) Of models 1-5, which is the best fitting by BIC? Which fits best by the goodness of fit chisquare? Which fits best by ID?

Model 5 is the best fit by the likelihood ratio test, and by the BIC, and the ID. Note that the ID gets monotonically smaller (and better) the more terms you add into the model. If you add coefficients that are exactly zero, ID will remain the same but it cannot get worse.

10) Can you find a model that fits better than model 5 by either BIC or the goodness of fit chisquare?

Take a look at my stata log. I examined the residuals for model 5 and compared them to the actual counts, and came up with the idea to to add a single gender symmetric black-Mexican term. This term improved the goodness of fit by 14.3 on 1 degree of freedom, P=0.000156, which is highly significant by the LRT. The improvement in BIC is smaller and probably not significant. The Quasi-Symmetry model (see my stata log), which uses the maximum 10 symmetric off-diagonal interactions fits better than model 5 by LRT, but worse by BIC. Of course the saturated model fits better than model 5 by the LRT (but worse by BIC) but you won't get full credit for that one.