EdPsych/Psych/Soc 589
Applied Categorical Data Analysis
C.J. Anderson

Homework 2
Answer Key

1. *Problem 2.19, p 59–60 of Agresti (2007).*

   (a) H_o : party and race are independent
   H_a : party and race are dependent
   $df = 2$, $X^2 = 167.85$, $p < .001$, and $G^2 = 187.58$, $p < .001$.
   Conclusion: Reject H_o; the data support the conclusion that party and race are dependent.

   (b) Adjusted (Haberman) residuals from SAS/GENMOD:

	Political Party		
Race	democrat	independent	republican
white	11.85	−0.69	−11.77
black	−11.85	0.69	11.78

   There are large residuals for Democrats and Republicans. There are more white Republicans & black Democrats than expected if race and political party were independent. In other words, there are fewer white Democrats and black Republicans than expected under independence. The residuals for independent are fairly small.

   (c) There are 3 different ways to partition, but given the pattern in the residuals from part (b), this is the one that seems the most logical (*In the SAS program I did all three ways if you want to check you results*):

Table 1

	Political Party	
Race	Democrat	Republican
white	871	873
black	302	43

   $H_{o(1)}$: Race and major party are independent
   $H_{a(1)}$: Race and major party are dependent.
   $df = 1$, $G^2 = 187.09$, $p < .001$, reject $H_{o(1)}$.
   The odds of Democrat given white are 0.14 times the odds of Democrat given black. Alternatively, the odds of Republican given white are $1/0.14 = 7.1$ times larger than the odds of Republican given black.

Table 2

Race	Political Party: Republican or Democrat	Political Party: Independent
white	1744	444
black	345	80

$H_{o(2)}$: Major party and Independent party are independent.
$H_{a(2)}$: Major party and Independent part are dependent.
$df = 1$, $G^2 = 0.4851$, $p = .49$, retain $H_{o(2)}$.

Note: $187.09 + 0.49 = 187.58$

Conclusion: Association between race and political party membership appears to be only present for the major parties (i.e., Republican & Democrats).

2. Problem 2.27, p 62 of Agresti (2007).

   Note: To find the adjusted residuals using SAS/GENMOD, see page 333 in the text and/or lecture notes.

Family Income	Educational Aspirations: some HS	HS Grad	Some college	college grad
Low	9	44	13	10
Middle	11	52	23	22
High	9	41	12	27

   (a) H_o : Income and educational aspirations are independent.
   H_a : Income and educational aspirations are independent.
   $df = 6$, $X^2 = 8.87$, $p = .18$, or $G^2 = 8.91$, $p = .18$. Retain H_o.

   Deficiency: This test does not take into account that both variables are ordinal. It is likely that as income goes up so does educational aspiration.

   (b) Adjusted residuals:

Family Income	Educational Aspirations: some HS	HS Grad	Some college	college grad
Low	.41	1.58	−.13	−2.11
Middle	−.19	−.54	1.30	−.40
High	−.19	−.95	−1.24	2.44

   The pattern suggests that a linear association. There are positive residuals

(under predictions) along a diagonal from low ends of the variables to high and negative residuals on the "off diagonal".

(c) Test for ordinal association: $H_o : \rho = 0$ versus $H_a : \rho \neq 0$.
Using equal spacing; $r = .132$, $M^2 = 4.749$, $df = 1$, $p = .03$.
Reject H_o; there is evidence of a linear trend in the data.

3. Cross-classification of coronary heart disease (CHD) by coffee consumption. The data are from a case-control study of 66 CHD cases and 85 unmatched control cases.

(a) $H_o : \theta = 1$ (there is not relationship between CHD and coffee use.
$H_a : \theta \neq 1$ (there is a relationship between CHD and coffee use.
$df = 1$, $X^2 = 5.61$, $p = .02$ or $G^2 = 5.647$, $p = .02$.
Reject H_o; there appears to be a dependency between coffee use and CHD.
Given that $\hat{\theta} = 2.198$, the odds of CHD given heavy coffee use are 2.198 times larger than the odds of CHD given no heavy coffee use.

(b) Assuming that this is a relationship between CHD and coffee usage,

- We cannot conclude that coffee usage causes/prevents coronary heart disease because the CHD margin was fixed by design. To make such a conclusion, we would have to randomly assign individuals to coffee use and non-use groups/treatement conditions ... and wait to see who develops CHD.
- Smoking is a possible explanatory variable. Heavy smokers tend to drink a lot of coffee and it is well documented that smoking contributes to (increases the risk) of CHD.
  Another possible explanatory variable is pastry (or donut) consumption. Those who like pastries might want a cup of coffee with them and consumption of high fat food leads to increased risk of CHD.
  Others?