

## Problem Set #3 Answer Key:

### A. Measurement Model

1. Defiance. I might exclude one of the substance abuse questions: it seems heavily weighted towards those questions. As far as what I might add: I would consider adding perhaps adding a question about being punished (either in school or by parents) and also about conflicts with parents or other authority figures.

(Note there are MANY answers to the above question!)

2. The BIC for the 2, 3, and 4 class models are 81719.7, 81085.7, and 80731.30, respectively (note:  $BIC = -2LL + s \cdot \log(N)$  where  $s$  = number of parameters and  $\log$  implies the natural log). Based on the BIC results, I would choose the 4 class model: it has the smallest BIC statistic. Also, there seems to be some scientific plausibility to all of the classes in the 4 class model.

3. Class 1 is a “non-defiant” class: it comprises about 50% of the population and most of activities are quite rare (prevalences less than 5%), while 3 of the activities (sex with more than one partner, fighting, and alcohol consumption) are relatively rare (prevalences between 10% and 20%). In Class 2, which includes about 17% of the students, individuals engage in relatively few of these activities, but tend to show a somewhat increased engagement in bringing weapons to schools (31%) and fighting (66%). They also have a about a 50% chance of alcohol consumption and more than one sex partner. But their uses of tobacco, marijuana, and drugs are rather low. Class 3 (23% of the population) shows higher incidence of substance use (tobacco, marijuana, and alcohol) than Class 2, but significantly lower incidence of fighting and bringing weapons to school. And, similar to class 2, about half have had more than one sex partner. Class 4 have high rates of all activities, with a notable increase in drug use (38%) as compared to the other three classes (<1%, 4%, and 9%). The remaining activities have prevalences of 80% and higher except for bringing a weapon to school, which is still relatively high (60%) compared to classes 1, 2, and 3. About 10% of the population falls into class 4.

4. The model shown in question 2 assumes that the way defiance should be measured in boys and girls is the same and that the items included in this model are indicative of defiance in both genders. If we have differential measurement by gender, this implies that we cannot combine boys and girls and fit one model to them both: we should fit stratified models and we will see that the class definitions (as determined by the item prevalences) differ.

5. The thresholds for classes 1, 2, and 3 in boys and girls are as follows:

	girls	boys
Class 1		
WEAPON\$1	4.757	2.637
FIGHT\$1	2.390	1.520
SEXLIFE2\$1	1.717	1.283
DRUGUSE\$1	5.471	5.337
TOBACCO\$1	3.138	2.850
ALCOH\$1	1.311	1.730
MARIJ\$1	5.192	3.924
class size:	59%	43%
Class 2		
WEAPON\$1	1.660	0.299
FIGHT\$1	-0.408	-0.758
SEXLIFE2\$1	0.239	-0.139
DRUGUSE\$1	3.038	3.550
TOBACCO\$1	1.287	1.210
ALCOH\$1	0.217	0.089
MARIJ\$1	2.172	1.679
class size:	12%	17%
Class 3		
WEAPON\$1	4.495	1.159
FIGHT\$1	1.612	0.395
SEXLIFE2\$1	0.319	-0.097
DRUGUSE\$1	2.460	2.240
TOBACCO\$1	-1.016	-0.674
ALCOH\$1	-2.365	-2.101
MARIJ\$1	0.174	-0.130
class size:	20%	28%
Class 4		
WEAPON\$1	0.601	-1.008
FIGHT\$1	-1.200	-2.461
SEXLIFE2\$1	-0.981	-2.100
DRUGUSE\$1	0.551	0.354
TOBACCO\$1	-1.412	-1.578
ALCOH\$1	-2.738	-3.264
MARIJ\$1	-1.858	-1.784
class size:	8%	12%

Based on these results, we can see that there is strong evidence of differential measurement. The definitions of classes (as can be seen by comparing the item thresholds) look very different. Even the class with low use (class 1) shows different

activity prevalences. A very striking difference in all classes is the prevalence of weapon (much more common in boys in all classes), and also in fighting (also more common in boys). Drug use does not differ drastically by class, and alcohol and marijuana differ relatively little also.

6. If conditional independence holds, then we would expect that there is no association between alcohol use and marijuana use within a class. We can test this by estimating the odds ratios and their 95% confidence intervals for marijuana use and alcohol use for each class. If these 95% confidence intervals overlap 1, then we can conclude that there does not appear to a violation of conditional independence.

Class	OR	95% CI
1	0.64	0.16, 1.80
2	1.32	0.90, 1.92
3	0.88	0.65, 1.20
4	1.20	0.29, 3.67

Based on these ORs and confidence intervals, it appears that the assumption of conditional independence is met.

#### B. Adding the regression component

1. I would think that the older kids in the sample would tend to be more likely to be in the higher classes. For example, I would expect the logOR for class 4 versus class 1 for age to be greater than 0. I would expect that kids in urban settings would be more likely to be in classes 3 and 4 than suburban kids. And, I would expect kids in rural settings to be more likely to be in classes 3 and 4 than suburban kids but I would not expect the logORs to be as large.

2. In the estimation, class 2 as defined above appears to be the reference class. This is the class with relatively low rates of most activities except fighting, alcohol and sexlife2. The coefficients are as follows:

	logOR	Zscore
Class 1 vs. Class 2	0.44	2.6
Class 3 vs. Class 2	0.84	7.6
Class 4 vs. Class 2	0.45	3.8

These imply that class in comparing two individuals, one who is 16 and one who is 15, the 16 year old has 1.5 times ( $e^{0.44}$ ) the odds of being in class 1 versus class 2, has 1.6 times ( $e^{0.45}$ ) the odds of being in class 4 versus class 2, and has 2.3 times the odds of being in class 3 versus class 2. All of these comparisons are significant. By looking at the magnitudes, it is likely that the logOR between classes 1 and 4 is approximately 0 and not significant. But, it looks as though there is an association between age and class membership for the other classes.

3. a. No, they hardly change at all. This implies that community type is not confounded with age. This is not surprising given the nature of the way the data were collected.

b. Yes, community type is associated with class membership. It appears that there are negative associations with Class 1 vs 2, Class 3 vs. 2, and Class 4 vs 2 for urban versus suburban. It does not appear that there is a significant difference between class membership for rural vs. suburban (ORs are mostly small (1.3, 1.08, and 1.05) and they are all insignificant.

c.

	RURAL		URBAN	
	logOR	Zscore	logOR	Zscore
Class 1 vs. Class 2	0.08	0.17	-0.58	-2.5
Class 3 vs. Class 2	0.36	0.71	-0.70	-2.3
Class 4 vs. Class 2	0.05	0.09	-0.34	-1.4

Recall that the reference class is a subgroup who tends to fight and have had at least 2 sex partners. They tend to not do drugs or smoke marijuana. They have relatively low tobacco and alcohol consumption. And, they have a high rate of carrying a weapon to school (74%).

This is consistent with the odds ratios for urban. It is to be expected that urban kids would tend to take weapons to school, but may be less likely to engage in other of the activities. This is confirmed by the negative odds ratios for urban, implying that of all the classes urban kids are more likely to be in class 2.

There doesn't appear to be a strong association with rural vs suburban for class membership. There is slightly positive association for being in Class 3 versus 2 for rural versus suburban, but it is not even close to significant.

4. The thresholds are listed for both models. Ones that appear different upon inspection are highlighted in bold, and all are in class 2. This implies that there could be differential measurement (i.e. the fact that the threshold parameters are different in the two models). I would like to be able to do some additional analyses to try to figure out which of the covariates appears to be the source of the differential measurement. One way to see this is to examine the association between each of the covariates and the items within classes. This approach would be similar to that done to investigate conditional independence in part A, number 6 where we looked at the associations between items within classes. Here, we would assign individuals to classes in the same way and then, for example, estimate a logistic regression between each of the activities and age within each of the four classes (this would give us  $4 \times 7 = 28$  odds ratios for age. we would then do the same for urban and rural, for a total of  $3 \times 7 \times 4 = 84$  odds ratios).

	LCR	LCA
Class 1		
WEAPON\$1	2.44	2.637
FIGHT\$1	2.47	1.520
SEXLIFE2\$1	1.24	1.283
DRUGUSE\$1	5.33	5.337
TOBACCO\$1	2.78	2.850
ALCOH\$1	1.69	1.730
MARIJ\$1	3.70	3.924
class size:	40%	43%
Class 2		
<b>WEAPON\$1</b>	<b>0.88</b>	<b>0.299</b>
<b>FIGHT\$1</b>	<b>-2.56</b>	<b>-0.758</b>
<b>SEXLIFE2\$1</b>	<b>0.41</b>	<b>-0.139</b>
DRUGUSE\$1	3.82	3.550
TOBACCO\$1	2.08	1.210
<b>ALCOH\$1</b>	<b>0.83</b>	<b>0.089</b>
MARIJ\$1	2.54	1.679
class size:	14%	17%
Class 3		
WEAPON\$1	0.94	1.159
FIGHT\$1	0.31	0.395
SEXLIFE2\$1	-0.22	-0.097
DRUGUSE\$1	2.36	2.240
TOBACCO\$1	-0.51	-0.674
ALCOH\$1	-1.83	-2.101
MARIJ\$1	0.04	-0.130
class size:	37%	28%
Class 4		
WEAPON\$1	-1.07	-1.008
FIGHT\$1	-2.69	-2.461
SEXLIFE2\$1	-1.99	-2.100
DRUGUSE\$1	0.32	0.354
TOBACCO\$1	-1.51	-1.578
ALCOH\$1	-3.11	-3.264
MARIJ\$1	-1.79	-1.784
class size:	12%	12%