

STA 6384, Report 4.5

Carson Slater *Baylor University*

Problem: Agresti’s analysis of the alcohol, cigarette, and marijuana use (our Example 4.4.1) is continued on his p. 379 where he considers collapsibility issues. For the collapsibility portion of the analysis, work Agresti’s problem 10.2, p. 407. You should provide a narrative of the example, conclusions, and code. Do this using R, following Laura Thompson’s treatment (her p. 142ff).

Use Table 2 to illustrate the odds ratio collapsibility conditions.

Table 1: Alcohol, Cigarette, and Marijuana Use for High School Seniors

Alcohol Use	Cigarette Use	Marijuana Use	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

Agresti provides a three way table with high school students’ responses to whether they have ever used alcohol (A), cigarettes (C), and marijuana (M). It is possible that a teen’s use of one drug increases the probability another drug will be used. We are interested in the association between teens’ use of these drugs to identify which if any are commonly used together. We cannot establish a causal relationship from observational data, but exploring the association between drug use for different substances can provide insight.

Table 2: Observed Counts and Fitted Counts from Various Log-Linear Models

Response			Observed	Fitted Counts from Models								
Marijuana	Cigarette	Alcohol	Count	(ACM)	(AC,AM,CM)	(AM,CM)	(AC,AM)	(AC,CM)	(AC,M)	(AM,C)	(A,CM)	(A,C,M)
Yes	Yes	Yes	911	911	910.383	909.24	710.003	885.877	611.178	627.296	782.683	539.983
No	Yes	Yes	538	538	538.616	438.84	738.997	563.123	837.822	327.704	497.526	740.226
Yes	No	Yes	44	44	44.617	45.76	244.997	29.449	210.896	652.913	39.391	282.091
No	No	Yes	456	456	455.386	555.16	255.003	470.551	289.104	341.087	629.400	386.700
Yes	Yes	No	3	3	3.617	4.76	0.703	28.123	19.402	3.284	131.317	90.597
No	Yes	No	43	43	42.384	142.16	45.297	17.877	26.598	1.716	83.474	124.194
Yes	No	No	2	2	1.383	0.24	4.297	16.551	118.524	211.507	6.609	47.329
No	No	No	279	279	279.614	179.84	276.703	264.449	162.476	110.493	105.600	64.880

Collapsibility Analysis and Theory

The concept of collapsibility in three-way contingency tables is crucial for understanding when marginal associations accurately reflect conditional associations. According to Agresti’s framework, for odds ratios, collapsibility conditions relate to logistic and loglinear models - specifically, marginal and conditional odds ratios are identical if certain conditional independence conditions hold.

Table 3: Goodness-of-Fit Statistics for Log-Linear Models

Model	G2	X2	df	P_value
A.C.M	1286.02	1411.39	4	0.000
A.CM	534.21	505.60	3	0.000
AM.C	939.56	824.16	3	0.000
AC.M	843.83	704.91	3	0.000
AC.CM	92.02	80.81	2	0.000
AC.AM	497.37	443.76	2	0.000
AM.CM	187.75	177.61	2	0.000
AC.AM.CM	0.37	NA	1	0.541
AC.AM.CM	NA	0.40	1	0.527
ACM	0.00	0.00	0	1.000

Table 4: Conditional and Marginal Odds Ratios

Association	Odds Ratio
Cigarette-Marijuana	
CM A=No	17.25
CM A=Yes	17.25
CM Marginal	25.14
Alcohol-Marijuana	
AM C=No	19.81
AM C=Yes	19.81
AM Marginal	61.87
Alcohol-Cigarette	
AC M=No	7.80
AC M=Yes	7.80
AC Marginal	17.70

Key Findings from Model (AC, AM, CM)

The model (AC, AM, CM) provides excellent fit with $G^2 = 0.37$ and $p = 0.54$, indicating no evidence against this model. This model assumes no three-way interaction but allows all two-way associations.

Collapsibility Assessment:

- CM Association is Collapsible:** The conditional odds ratios for cigarette-marijuana association are essentially identical across alcohol levels (≈ 2.84), and equal to the marginal odds ratio. This occurs because C is conditionally independent of A given M.
- AC and AM Associations are NOT Collapsible:** These show different conditional versus marginal odds ratios because the third variable creates confounding.

a. For model (A, C, M), all conditional odds ratios equal 1.0. Explain why all reported marginal

odds ratios equal 1.0.

In model (A, C, M) , all variables are assumed to be mutually independent. This means: - All conditional odds ratios equal 1.0 because there are no associations between any pairs of variables - Consequently, all marginal odds ratios must also equal 1.0 - If X and Y are independent conditional on all levels of Z , then X and Y are also marginally independent - This is because marginal independence follows from conditional independence when the conditioning variable has no association with either of the other variables

b. For model (AC, M) , explain why (i) all conditional odds ratios are the same as the marginal odds ratios, and (ii) all $\hat{\mu}_{ac+} = n_{ac+}$.

For model (AC, M) :

(i) All conditional odds ratios equal marginal odds ratios because:

- The model assumes M is independent of both A and C
- This means the AC association is the same regardless of marijuana use level
- Since M doesn't affect the AC relationship, collapsing over M doesn't change the association
- The AM and CM associations are both 1.0 (independence) at all conditional levels and marginally

(ii) All fitted marginal totals $\hat{\mu}_{ac+} = n_{ac+}$ because:

- The model includes the AC interaction term
- Log-linear models have the property that fitted marginal totals equal observed marginal totals for any margin included in the model
- Since (AC) is in the model, the AC marginal totals are preserved

c. For model (AM, CM) , explain why (i) the AC conditional odds ratios of 1.0 need not be the same as the AC marginal odds ratio, and (ii) the AM and CM conditional odds ratios are the same as the marginal odds ratios and all $\hat{\mu}_{a+m} = n_{a+m}$ and $\hat{\mu}_{+cm} = n_{+cm}$.

For model (AM, CM) :

(i) The AC conditional odds ratios (all 1.0) differ from the marginal AC odds ratio because:

- The model assumes A and C are conditionally independent given M
- However, both A and C are associated with M (through AM and CM terms)
- This creates a spurious marginal association between A and C due to their mutual association with M
- This is a classic example of confounding - the marginal association doesn't reflect the true conditional relationship

(ii) The AM and CM associations are collapsible because:

- For AM : C is conditionally independent of A given M , so collapsing over C doesn't affect the AM association

- For CM : A is conditionally independent of C given M , so collapsing over A doesn't affect the CM association
- The fitted marginals equal observed marginals because both AM and CM interactions are in the model

d. For model (AC, AM, CM) , explain why (i) no conditional odds ratios need be the same as the related marginal odds ratios, and (ii) the fitted marginal odds ratios must equal the sample marginal odds ratios.

For model (AC, AM, CM) :

(i) No conditional odds ratios need equal marginal odds ratios because:

- This model allows for complex confounding patterns where each variable can be associated with the other two
- The marginal association between any two variables can be influenced by their relationships with the third variable
- However, in practice, some may still be equal if certain conditional independence patterns emerge from the data
- Our analysis shows the CM association exhibits collapsibility, but AC and AM do not

(ii) All fitted marginal odds ratios equal sample marginal odds ratios because:

- The model includes all possible two-way interaction terms
- Log-linear models preserve the marginal totals for any interaction included in the model
- Since all two-way interactions are in the model, all two-way marginal totals (and hence marginal odds ratios) are preserved

Conclusions

This analysis demonstrates that collapsibility depends critically on the conditional independence structure. The cigarette-marijuana association exhibits collapsibility over alcohol, making it appropriate to examine this relationship in a collapsed two-way table. However, the alcohol-cigarette and alcohol-marijuana associations require conditional analysis to avoid misleading conclusions.

The consistent CM association across alcohol levels ($OR \approx 2.84$) suggests a particularly strong link between cigarette and marijuana use that persists regardless of alcohol consumption patterns, which has important implications for substance abuse prevention programs targeting adolescents.