STA 6384, Report 4.4

Carson Slater *Baylor University*

Problem: Reproduce the results of Example 4.4.1 in the notes.

First, we create the $2 \times 2 \times 2$ contingency table from the data provided in Table 4.7 of the PDF.

```
# Create the 3-way contingency table
counts <- c(911, 538, 44, 456, 3, 43, 2, 279)
data <- as.table(array(counts, dim = c(2, 2, 2),
                       dimnames = list(Alcohol = c("Yes", "No"),
                                       Cigarettes = c("Yes", "No"),
                                       Marijuana = c("Yes", "No"))))
```

We use the `loglm` function from the MASS library to fit the various log-linear models.

```
# Fit the log-linear models
# Saturated model
mod_sat <- loglm(~ Alcohol * Cigarettes * Marijuana, data = data, fitted = TRU
# Homogeneous association model
mod_homo <- loglm(
  ~ (Alcohol + Cigarettes + Marijuana)2,
  data = data,
  fitted = TRUE
)
# Other models
mod_AC_M <- loglm(
  ~ Alcohol * Cigarettes + Marijuana,
  data = data,
  fitted = TRUE
)
mod_AM_C <- loglm(
  ~ Alcohol * Marijuana + Cigarettes,
  data = data,
  fitted = TRUE
)
mod_CM_A <- loglm(
  ~ Cigarettes * Marijuana + Alcohol,
  data = data,
  fitted = TRUE
)
mod_AC_AM <- loglm(
  ~ Alcohol * Cigarettes + Alcohol * Marijuana,
```

```r
  data = data,
  fitted = TRUE
)
mod_AC_CM <- loglm(
  ~ Alcohol * Cigarettes + Cigarettes * Marijuana,
  data = data,
  fitted = TRUE
)
mod_AM_CM <- loglm(
  ~ Alcohol * Marijuana + Cigarettes * Marijuana,
  data = data,
  fitted = TRUE
)
mod_ind <- loglm(~ Alcohol + Cigarettes + Marijuana, data = data, fitted = TRU
```

To replicate Table 4.9 from the PDF, we calculate the likelihood ratio statistic (G^2), the Pearson chi-squared statistic (X^2), degrees of freedom (df), p-value, and the Akaike Information Criterion (AIC) for each model.

The results show that the homogeneous association model, `(AC, AM, CM)`, provides a good fit to the data, as its p-value is not significant ($p = 0.541$), indicating that the model is consistent with the observed data. The AIC for this model is also the lowest among the non-saturated models, further supporting its selection.

```r
gof_stats <- data.frame(
  Model = model_names,
  G2 = sapply(models, function(x) x$lrt),
  X2 = sapply(models, function(x) x$pearson),
  df = sapply(models, function(x) x$df),
  p_value = sapply(models, function(x) pchisq(x$lrt, x$df, lower.tail = FALSE
  AIC = sapply(models, function(x) x$lrt - 2 * x$df)
) |>
  arrange(desc(G2))
```

Based on the goodness-of-fit tests, the homogeneous association model is the most appropriate. We then calculate the marginal and conditional odds ratios to interpret the associations between the variables. The marginal odds ratios are calculated from the two-way marginal tables. The odds ratio formula is $\theta = \frac{n_{11}n_{22}}{n_{12}n_{21}}$. The conditional odds ratios examine the association between two variables at each level of the third variable. For the homogeneous association model, the conditional odds ratios for a pair of variables are assumed to be equal across the levels of the third variable. We calculate them from the original three-way table.

Model	G2	X2	df	p_value	AIC
(A, C, M)	1286.02	1411.39	4	0.00	1278.02
(AM, C)	939.56	824.16	3	0.00	933.56
(CM, A)	843.83	704.91	3	0.00	837.83

Model	G2	X2	df	p_value	AIC
(AC, M)	534.21	505.60	3	0.00	528.21
(AM, CM)	497.37	443.76	2	0.00	493.37
(AC, AM)	187.75	177.61	2	0.00	183.75
(AC, CM)	92.02	80.81	2	0.00	88.02
(AC, AM, CM)	0.37	0.40	1	0.54	63.40
(ACM)	0.00	0.00	0	1.00	0.00

Type	Association	Condition	Odds_Ratio
Marginal	Alcohol-Cigarettes	None	25.136
Marginal	Alcohol-Marijuana	None	61.873
Marginal	Cigarettes-Marijuana	None	17.703
Conditional	Alcohol-Cigarettes	Marijuana = Yes	17.549
Conditional	Alcohol-Cigarettes	Marijuana = No	9.733
Conditional	Alcohol-Marijuana	Cigarettes = Yes	13.803
Conditional	Alcohol-Marijuana	Cigarettes = No	7.655
Conditional	Cigarettes-Marijuana	Alcohol = Yes	24.271
Conditional	Cigarettes-Marijuana	Alcohol = No	13.461

Marginal Associations

The marginal odds ratios look at the relationship between any two substances, ignoring the third.

Alcohol & Marijuana (OR = 61.87): This is the strongest association. The odds of a student using alcohol are almost 62 times higher if they use marijuana, compared to if they don't.

Alcohol & Cigarettes (OR = 25.14): The odds of a student using alcohol are about 25 times higher if they also smoke cigarettes.

Cigarettes & Marijuana (OR = 17.70): The odds of a student smoking cigarettes are about 18 times higher if they use marijuana.

Conditional Associations (Interaction Effects)

The conditional odds ratios show how the relationship between two substances changes depending on the use of the third substance. This reveals a significant interaction effect.

Association between Cigarettes and Marijuana: This link is much stronger among students who also use alcohol. The odds ratio is 24.3 for alcohol users but only 13.5 for non-alcohol users. This means alcohol use amplifies the association between cigarette and marijuana use.

Association between Alcohol and Cigarettes: This link is stronger among students who use marijuana. The odds ratio is 17.5 for marijuana users compared to 9.7 for non-users.

Association between Alcohol and Marijuana: Similarly, this association is stronger among cigarette smokers (OR = 13.8) than non-smokers (OR = 7.7).

Conclusion

The data reveals that not only are alcohol, cigarettes, and marijuana use strongly linked, but they also have a synergistic relationship. The use of any one of these substances substantially strengthens the statistical association between the other two. For instance, the already strong link between smoking cigarettes and using marijuana becomes even stronger if a student also drinks alcohol. This suggests that these behaviors are highly intertwined and may reinforce one another.