Describing Stratified Multiple Responses for Sparse Data

Ivy Liu
School of Mathematical and Computing Sciences
Victoria University
Wellington, New Zealand

June 28, 2004

SUMMARY

Surveys often contain qualitative variables for which respondents may select any number of the outcome categories. For instance, for the question "What type of contraception have you used?" with possible responses (oral, condom, lubricated condom, spermicide, and diaphragm), respondents would be instructed to select as many of the outcomes that apply. This type of response is called *multiple responses.* Bilder and Loughin (2002) proposed a Cochran-Mantel-Haenszel (MH) type method to test whether the choice of type of contraception is marginally independent of an explanatory variable given a stratification variable (known as conditional multiple marginal independence, CMMI). We apply the generalized MH type estimators (Greenland, 1989) to estimate the conditional group effects among the c outcome categories and follow the bootstrap method to estimate the variances and covariances for the estimators. The method can also be used for data with dependent observations across strata. It performs well even for highly sparse data.

Key words: bootstrap method; Mantel-Haenszel estimator; multiple responses; odds ratio.

1 Introduction

In a survey, it is common that the respondents may select any number of the outcome categories. For instance, for the question "What type of contraceptives have you used?" with possible responses (oral, condom, lubricated condom, spermicide, and diaphragm), respondents would be instructed to select whichever of the outcomes apply. Categorical variables that summarize this kind of data are called *pick any/c variables* (Bilder and Loughin, 2002), where c is the number of outcome categories ($c = 5$ in this case). We can cross-classify the counts from a survey that contains a pick any/c variable along with a group variable (r levels, e.g. whether a subject had a prior history of urinary tract infection) and a stratification variable (K levels, e.g. several age groups) into an $r \times c \times K$ contingency table. In the $r \times c \times K$ table, subjects may be represented in more than one cell. Table 1 given by Bilder and Loughin (2002) shows data for the above example of 239 sexually active college women in a $2 \times 5 \times 2$ table. Also, Table 3 gives the complete information about whether a subject had a prior history of urinary tract infection (UTI) (yes or no), type of contraceptive used (oral, condom, lubricated condom, spermicide, and diaphragm), and age (≥ 24 or < 24).

Another example is a study conducted by Dr. Paul Warren in the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. For the data, 6 experts (raters) rated a bunch of non-native English utterances into 3 scales for comprehensibility (from "not easy" to "very easy" to understand) and then indicated whether there was a problem for that utterance in each of 7 items (e.g. pronunciation of consonants, vowel pronunciations, word stress, etc.). Each of the 6 raters assessed each of the 50 utterances and gave a rating to each one as well as binary choices (i.e., it was or was not a problem) on the 7 items (a-g). We want to look at the conditional relationship between the rating and the 7 items given raters. Table 2 shows 6 different 3×7 tables (i.e., $K = 6$, $r = 3$, and $c = 7$), where the cell counts are dependent across the columns for each table and also dependent across the 6 strata. The complete information about each of the 50 utterances can be displayed as in Table 3.

Both examples have data with stratified multiple responses, yet the observations are not independent across the strata in the second example. This type of data occurs frequently in health and social sciences, and language studies.

Bilder and Loughin (2002) generalized the Cochran test to determine if the group and pick any/c variable are marginally independent given a stratification variable (known as conditional multiple marginal independence, CMMI). For the UTI example, they tested whether the contraception practices of women are different based on their urinary tract infection history controlling for their age group. They used a nonparametric bootstrap method to obtain the p-value of the test. When the group and pick any/c variable are not conditionally marginally independent, it is more interesting to describe how the pick any/c variable depends on the group. Similarly, for the linguistics example, we are not interested in the differences between the raters, but we focus on describing the conditional relationship between the rating and the items given each rater. This article evaluates the conditional row effects on picking any/c variable given the stratification variable. It allows us to establish multiple comparisons among the conditional row effects and c outcome categories.

Instead of using marginal logit models for which the pick any/c variable is treated as a c-dimensional binary response discussed by Agresti and Liu (2001), this article discusses the generalized Mantel-Haenszel (MH) estimators. We extend the MH estimators for a categorical $r \times 2 \times K$ table (Greenland, 1989) and follow the bootstrap method to estimate the variances and covariances for the estimated conditional row effects among the c outcome categories. Similar to the Cochran-Mantel-Haenszel test, the generalized MH estimators are used when the conditional row effects are not expected to vary drastically among the strata. For a binary response case, it is a well known situation that the MH estimators perform much better than the ML estimators when the number of strata is large and there are few observations in each stratum (Andersen, 1980, p. 244). It is not surprising that the MH estimators should perform better than the ML estimators for the multiple responses case when the responses are considered as c different binary responses. Also, when the data are highly sparse, for instance, there is no respondent on a particular outcome category in a row, the generalized estimating equations (GEE) approach fails to estimate the conditional row effects.

2 Generalized Mantel-Haenszel Estimators

Let's consider each item separately. For example, we consider item "a" (consonants pronunciations) only in Table 2. The conditional association between rating and "whether there was a consonants pronunciations problem" given raters can be described using a $3 \times 2 \times 6$ table, where the column variable is "whether there was a consonants pronunciations problem" with two levels (yes, no), the row variable is rating (not easy, medium, very easy), and the stratum variable is rater. Suppose we naively treat the 3×2 tables for 6 raters as independent. If the association between rating and "whether there was a consonants pronunciations problem" doesn't change dramatically across different raters, we can use the generalized MH estimator (Greenland, 1989) to describe the conditional relationship between the row and column variables.

For a general $r \times c \times K$ table, let X_{ik}^j denote the number of utterances having a problem on item j that are rated by the kth rater (stratum) with the overall rating (row) i. The notation n_{ik} denotes the total number of utterances in the ith row and the kth stratum. Suppose that $\pi_{j|ik}$ is the probability of having a problem on item j when the utterance is at row i and stratum k. Let $N_k = n_{1k} + \ldots + n_{rk}$. Also, let the odds ratio for rows i and h be

$$\theta_{ih}^j = \frac{\pi_{j|ik}(1 - \pi_{j|hk})}{(1 - \pi_{j|ik})\pi_{j|hk}} \quad j = 1, \ldots, c, \;\; i = 1, \ldots, r, \;\; h = 1, \ldots, r, \text{ and } i \neq h,$$

for all k, which is the ratio of the odds of having a problem on item j for utterances at row i to the odds of having a problem on item j for utterances at row h, given any stratum. The generalized MH estimator (Greenland, 1989) of $\log \theta_{ih}^j$ is

$$\bar{L}_{ih}^j = (L_{i+}^j - L_{h+}^j)/r\,, \tag{1}$$

where $L_{ih}^j = \log\left(\frac{\sum_{k=1}^K X_{ik}^j (n_{hk} - X_{hk}^j)/N_k}{\sum_{k=1}^K X_{hk}^j (n_{ik} - X_{ik}^j)/N_k}\right)$ and the subscript "+" indicates summation over that subscript. When the row variable has only two levels ($r = 2$) as for the UTI example in Table 1, we can use θ_{12}^j to describe the conditional row effect on selecting item j and the generalized MH estimator of $\log \theta_{12}^j$ is simplified to the ordinary MH estimator

$$L_{12}^j = \log\left(\frac{\sum_{k=1}^K X_{1k}^j (n_{2k} - X_{2k}^j)/N_k}{\sum_{k=1}^K X_{2k}^j (n_{1k} - X_{1k}^j)/N_k}\right). \tag{2}$$

For each item, say j, Greenland (1989) proposed the *dually* consistent variance and covariance estimators for $\{\bar{L}_{ih}^j, \forall\, i \neq h\}$ for either sparse or non-sparse data. However, since the observations may not be independent across strata, these variance and covariance estimators are not consistent anymore. Also, people might be interested in comparing the conditional association across items. For instance, for the UTI example, one might be interested in comparing the UTI effects for contraceptive methods "oral" with "condom" and for the linguistics example, one might be interested in comparing the rating effects for different items. Because the generalized MH estimators are dependent across different items, it is complicated to derive the dually consistent variances and covariances estimators for them. This article uses a realistic way to find estimates applying the nonparametric bootstrap method. Therefore, we can obtain the bootstrap estimated variances and covariances of the generalized MH estimators $\{\bar{L}^j_{ih},\, j = 1, \ldots, c,\, i \neq h = 1, \ldots, r\}$.

The nonparametric bootstrap method (Efron and Tibshirani, 1993) was conducted by randomly selecting subjects with replacement from the original data. For instance, for the UTI data, we resample N_k women with replacement from the kth stratum, where $k = 1, 2$. Similarly, for the linguistics example, we resample 50 utterances with replacement and across classify the data into the $3 \times 7 \times 6$ table. For each resample data set, the size of each stratum is the same as the original data. Denote the total number of observations as n. We take B resamples of size n and then for each resample, we calculate the generalized MH estimates $\{\bar{L}^j_{ih},\, j = 1, \ldots, c,\, i \neq h = 1, \ldots, r\}$. The bootstrap estimate of standard error of $\bar{L}^j_{ih}$ is the standard deviation of the bootstrap replicates,

$$\text{s.e. for } \bar{L}^j_{ih} = \sqrt{\frac{\sum_{b=1}^{B} \left(\bar{L}^j_{ih,b} - \sum_{b=1}^{B} \bar{L}^j_{ih,b}/B \right)^2}{B-1}},$$

where $\bar{L}^j_{ih,b}$ is the generalized MH estimate $\bar{L}^j_{ih}$ for the bth bootstrap resample. Similarly, the bootstrap estimate of covariance of $\bar{L}^j_{ih}$ and $\bar{L}^{j'}_{i'h'}$ is

$$\hat{\text{cov}}(\bar{L}^j_{ih},\, \bar{L}^{j'}_{i'h'}) = \frac{\sum_{b=1}^{B} \left(\bar{L}^j_{ih,b} - \sum_{b=1}^{B} \bar{L}^j_{ih,b}/B \right) \left(\bar{L}^{j'}_{i'h',b} - \sum_{b=1}^{B} \bar{L}^{j'}_{i'h',b}/B \right)}{B-1}.$$

3 Examples

Let's consider the UTI example in Table 1, the MH estimate of $\{\log \theta^j_{12},\, j = 1, \ldots, 5\}$ is $\{0.12,\, -0.52,\, 0.71,\, 0.64,\, 0.08\}$. The estimate of $\log \theta^1_{12}$ indicates that the odds of having used oral contraceptive for women without a prior history of UTI are estimated to be $\exp(0.12) = 1.13$ times higher than the odds for women with a prior history of UTI, given each age group. Choosing $B = 100$, the corresponding bootstrap standard error is $\{0.28,\, 0.26,\, 0.28,\, 0.32,\, 0.39\}$. Table 4 gives the bootstrap variances and covariances estimates of $\{L^j_{12},\, j = 1, \ldots, 5\}$. The conditional UTI effects are significant for contraceptives "condom", "lubricated condom", and "spermicide" at a 5% significance level. For comparing the UTI effects for contraceptives "oral" and "lubricated condom", a 95% confidence interval for $\log \theta^3_{12} - \log \theta^1_{12}$ is $(-0.39, 1.57)$. Table 5 shows all multiple comparisons of the conditional UTI effects for any two items.

In the example, since there are no women without prior history of urinary tract infection who use diaphragms, the MH estimate (2) for item 5 is undefined. In order not to smooth the data too much, we add 0.5 to each cell in the stratum with largest size. For instance, because the stratum of Age<24 contains more observations, we add 0.5 to each cell count in that stratum. The cell counts for "UTI and whether diaphragms is used" being "no and yes", "yes and yes", "no and no", and "yes and no" become 0.5, 5.5, 85.5, and 111.5 respectively.

For the Linguistics example in Table 2, by comparing rating levels 1 and 2, the MH estimate of $\{\log \theta^j_{12},\, j = 1, \ldots, 7\}$ is $\{-0.00,\, 1.19,\, 0.70,\, 0.28,\, -0.10,\, 0.88,\, -0.39\}$ with the bootstrap standard error of $\{0.81,\, 0.50,\, 0.53,\, 0.40,\, 0.47,\, 0.50,\, 1.07\}$. Comparing rating levels 1 and 3, the MH estimate of $\{\log \theta^j_{13},\, j = 1, \ldots, 7\}$ is $\{1.34,\, 1.47,\, 1.21,\, 1.49,\, 0.73,\, 1.36,\, -1.23\}$ with the bootstrap standard error of $\{0.73,\, 0.48,\, 0.58,\, 0.49,\, 0.44,\, 0.45,\, 1.17\}$. There are no significant differences between rating levels 1 and 2 for most of items, except for item b (pronunciation of vowels), given each of raters. However, the differences between rating levels 1 and 3 are significant for most of items given each of raters, except for items a, e, and g. Table 6 shows the generalized MH estimates and their bootstrap standard errors. Similarly, the bootstrap variances and covariances estimates can be obtained.

4 Conclusion

To describe stratified multiple responses, we can use the generalized MH estimators to evaluate the conditional associations between row and column variables. Like the ordinary MH estimator, the estimators can perform well even for sparse data based on a simulation study. When there are dependent strata as in Table 2, the generalized MH estimators can still be used. The motivation of the naive assumption of independence across strata comes from work by Liang and Zeger (1986) showing that such naive estimators for repeated measurement data can perform well.

Though we motivated the generalized MH estimators by treating K separate $r \times 2$ tables as independent, the variances and covariances estimators based on this would be inappropriate. Also, the MH estimators between items are dependent. In Section 2, we proposed the nonparametric bootstrap procedure to estimate the variances and covariances for the MH estimators. Another simulation study shows that the bootstrap estimates perform reasonably well.

REFERENCES

Agresti, A. and Liu. I. (2001). Strategies for Modeling a Categorical Variable Allowing Multiple Category Choices. *Sociological Methods and Research* **29**, 403-434.

Andersen, E. B. (1980). *Discrete Statistical Methods with Social Science Applications.* North-Holland, New York.

Bilder, C. and Loughin, T. (2002). Testing for Conditional Multiple Marginal Independence. *Biometrics* **58**, 200-208.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Greenland, S. (1989). Generalized Mantel-Haenszel Estimators for K $2 \times J$ Tables. *Biometrics* **45**, 183-191.

Liang, K.-Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Mantel, N. and Haenszel, W. (1959). Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. *Journal of the National Cancer Institute* **22**, 719-748.

Table 1: The marginal UTI data

	Oral	Condom	L. cond.	Spermicide	Diaphragm	Total responses	Total women
	Contraceptive						
Age ≥ 24							
UTI							
No	18	9	8	7	0	42	24
Yes	8	9	2	3	2	24	14
Age < 24							
UTI							
No	55	41	37	27	0	160	85
Yes	75	68	33	22	5	203	116

Table 2: The marginal Linguistics data

	Items							Total	Total
	a	b	c	d	e	f	g	responses	utterances
Rater 1									
Rating									
1	8	7	2	2	1	0	1	21	8
2	32	22	7	2	6	0	3	72	32
3	8	1	3	0	0	0	1	13	10
Rater 2									
Rating									
1	10	8	8	4	5	8	0	43	11
2	18	6	10	11	8	11	1	65	19
3	18	9	4	3	8	7	0	49	20
Rater 3									
Rating									
1	7	1	3	0	4	2	0	17	7
2	11	4	6	1	8	4	0	34	13
3	23	7	8	3	13	8	2	64	30
Rater 4									
Rating									
1	2	2	2	2	0	0	0	8	2
2	11	7	2	4	1	1	0	26	12
3	11	6	1	5	0	0	1	24	36
Rater 5									
Rating									
1	1	0	0	0	0	0	0	1	1
2	8	6	5	0	1	1	0	21	23
3	5	11	4	0	1	1	0	22	26
Rater 6									
Rating									
1	14	18	6	14	14	17	0	83	18
2	12	10	1	9	11	9	0	52	14
3	12	14	4	7	9	11	1	58	18

Table 3: The complete UTI data

Age $\geq$ 24																
Contraceptive																
Oral	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Condom	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
L. cond.	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Spermicide	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Diaphragm	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
UTI																
No	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	0
Yes	0	0	0	0	0	0	0	0	2	0	1	1	1	0	0	1
Contraceptive																
Oral	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Condom	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
L. cond.	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Spermicide	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Diaphragm	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
UTI																
No	14	0	1	0	0	0	0	0	1	0	0	0	0	0	2	0
Yes	5	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Age < 24																
Contraceptive																
Oral	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Condom	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
L. cond.	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Spermicide	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Diaphragm	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
UTI																
No	0	0	1	0	0	0	0	0	2	0	1	0	8	0	18	0
Yes	0	0	1	0	0	0	0	0	14	0	3	0	10	0	12	1
Contraceptive																
Oral	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Condom	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
L. cond.	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Spermicide	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Diaphragm	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
UTI																
No	42	0	1	0	0	0	0	0	1	0	0	0	5	0	6	0
Yes	44	3	0	0	0	0	0	0	15	1	2	0	7	0	3	0

Table 4: The bootstrap variances and covariances estimates of $\{L_{12}^{j},\, j = 1, \ldots, 5\}$.

	L_{12}^1	L_{12}^2	L_{12}^3	L_{12}^4	L_{12}^5
L_{12}^1	0.079	−0.050	−0.045	−0.048	0.011
L_{12}^2	−0.050	0.068	0.051	0.045	−0.007
L_{12}^3	−0.045	0.051	0.081	0.051	−0.006
L_{12}^4	−0.048	0.045	0.051	0.104	−0.012
L_{12}^5	0.011	−0.007	−0.006	−0.012	0.152

Table 5: A 95% confidence interval for $\log \theta_{12}^{j} - \log \theta_{12}^{j'}$.

			j		
	1	2	3	4	5
j'	Oral	Condom	L. cond.	Spermicide	Diaphragm
1. Oral		(−1.6140, 0.3342)	(−0.3876, 1.5724)	(−0.5117, 1.5589)	(−3.5851, −1.7931)
2. Condom			(0.8074, 1.6572)	(0.6022, 1.7248)	(−2.9973, −1.1011)
3. L. cond.				(−0.6335, 0.4959)	(−4.2517, −2.3113)
4. Spermicide					(−4.2498, −2.1756)

Table 6: The generalized MH estimates and their bootstrap standard errors (in parentheses) for the data in table 2

				item j			
	a	b	c	d	e	f	g
	pronunciation of consonants	pronunciation of vowels	word stress	sentence stress	rhythm	intonation	rate
L_{12}^{j}	−0.00	1.19	0.70	0.28	−0.10	0.88	−0.39
	(0.81)	(0.50)	(0.53)	(0.40)	(0.47)	(0.50)	(1.07)
L_{13}^{j}	1.34	1.47	1.21	1.49	0.73	1.36	−1.23
	(0.73)	(0.48)	(0.58)	(0.49)	(0.44)	(0.45)	(1.17)
L_{23}^{j}	1.34	0.27	0.52	1.20	0.83	0.48	−0.84
	(0.52)	(0.30)	(0.47)	(0.50)	(0.50)	(0.43)	(1.35)