This article discusses strategies for modeling a categorical variable when subjects can select any subset of the categories. With c *outcome categories, the models relate to a* c-*dimensional binary response, with each component indicating whether a particular category is chosen. The strategies are the following: (1) Using logit models directly for the marginal distribution of each component; this accounts for dependence among the component responses but does not treat the dependence as an integral part of the model. (2) Using logit models containing subject random effects to generate the dependence among the components; this approach is limited by implying nonnegative associations having a certain exchangeability. (3) Using loglinear modeling; quasi-symmetric ones are useful but are limited to estimation of within-subject effects. Marginal logit models less fully describe the dependence patterns for the data but require fewer assumptions and focus more directly on the effects of greatest substantive interest.*

Strategies for Modeling a Categorical Variable Allowing Multiple Category Choices

ALAN AGRESTI
University of Florida
IVY LIU
Victoria University of Wellington

1. INTRODUCTION

Surveys sometimes contain qualitative variables for which the respondents may select any number of the outcome categories. For instance, the Gallup organization has periodically asked people their reasons for supporting or opposing the death penalty (see, e.g., their poll release of February 24, 2000, at their Web site and the January/February 1985 issue of *Gallup Report*). Those favoring the death

AUTHORS' NOTE: *The research of the first author was supported by grants from the National Institutes of Health and the National Science Foundation. The authors are grateful to Thomas Loughin of Kansas State University for kindly supplying the Kansas pig farmer data and for making comments on this article. They thank Joseph B. Lang of the University of Iowa for supplying a copy of his S-PLUS function* gllm *for fitting generalized log-linear models. Thanks also to Michael Radelet for mentioning the death-penalty survey applications.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 29 No. 4, May 2001 403-434

penalty could select from reasons such as revenge (an "eye for an eye"), acts as a deterrent, costly to keep them in prison, keeps them from killing again. In such surveys, subsequent analysis can then model the selections in terms of the recorded explanatory variables.

Loughin and Scherer (1998) recently provided an example of such data and proposed certain analyses. They quoted a study (Richert et al. 1993) that asked 262 pig farmers in Kansas questions about their veterinary information sources. For the question "What are your primary sources of veterinary information?" the response categories were (A) professional consultant, (B) veterinarian, (C) state or local extension service, (D) magazines, and (E) feed companies and reps. Farmers sampled were asked to select all categories that were relevant. Table 1 shows the response counts for the outcome categories cross classified with two explanatory variables, the farmers' achieved education (whether they had at least some college education), and the number of pigs they marketed annually (less than 1,000, 1,000 to 2,000, 2,000 to 5,000, more than 5,000). We refer to these explanatory variables as "education" and "size of farm." This $2 \times 4 \times 5$ contingency table contains 453 positive responses of categories from the 262 farmers.

For these data, let $y_k = 1$ if a subject selects category k ($k = 1, \ldots, 5$), and let $y_k = 0$ otherwise. Let $y = (y_1, y_2, \ldots, y_5)$ denote the response profile on the five categories, and note that $0 \leq \Sigma_k y_k \leq 5$. At each combination of education and size of farm, there are 2^5 possible $(y_1, y_2, \ldots, y_5)$ profiles, according to the (yes, no) outcome for the selection of each outcome category. Thus, the selection of outcome categories is most fully viewed as a cross classification of five binary components: variable A indicating whether the respondent said "yes" to source A, variable B indicating whether the respondent said "yes" to source B, and so forth. For instance, among the farmers with no college education and less than 1,000 pigs marketed annually, Table 1 indicates that two farmers received information from source A. In fact, one of these farmers reported source A only (i.e., y = [1,0,0,0,0]), and the other farmer reported all five sources (i.e., y = [1,1,1,1,1]). The complete data set is the $2 \times 4 \times 2^5$ contingency table showing the counts of the possible profiles y for each combination of levels of X_1 = education and X_2 = size of farm; that is, the $X_1 \times X_2 \times A \times B \times C \times D \times E$ cross classification. Table 2 shows this table. Table 1 is marginal to Table 2,

TABLE 1: Farmers' Veterinary Information Sources by Education and Number of Pigs Marketed Annually, for 262 Kansas Pig Farmers

Education	*Number of Pigs*	*Information Source*					*Total Number of Responses*	*Total Number of Subjects*
		A	*B*	*C*	*D*	*E*		
No college	< 1,000	2	13	18	22	17	72	42
	1,000-2,000	2	15	10	11	15	53	27
	2,000-5,000	7	10	10	14	11	52	22
	> 5,000	13	10	7	14	7	51	27
Some college	< 1,000	3	16	21	33	22	95	53
	1,000-2,000	2	10	15	22	10	59	42
	2,000-5,000	1	7	7	7	6	28	20
	> 5,000	14	9	7	8	5	43	29
Total		44	90	95	131	93	453	262

SOURCE: Data supplied by Dr. Thomas Loughin, Kansas State University.
NOTE: Information sources are as follows: A = professional consultant; B = veterinarian; C = state or local extension service; D = magazines; and E = feed companies and reps.

referring to counts of the yes responses in the marginal distributions of the components at each educational level.

This article discusses three ways to use existing methods to model outcome variables such as this that have multiple potential responses. All three strategies analyze the data in the complete form of Table 2, but they differ in terms of the level of aggregation of the probabilities to which the models refer. Section 2 discusses logit models expressed directly in terms of the marginal distributions of the five components, that is, in terms of the distributions for which Table 1 summarizes counts for the yes category. The models deal with distributions that are marginal to the joint distribution for Table 2, for instance, focusing on how the probability of selecting a particular veterinary information source depends on education and size of farm. Section 3 presents logit models with normal random effects that account for nonnegative associations among components of the joint distribution. Those models refer to subject-specific distributions for which Table 2 is marginal but again focus on effects of explanatory variables on the choices of categories. Section 4 discusses loglinear models for the joint distribution of Table 2 itself. Connections exist among the model types; we see, for

TABLE 2: Cross-Classification of Responses (Y = yes, N = no) on Five Items Relating to Primary Sources of Veterinary Information, by Education (no college, at least some college) of Farmer and Annual Number of Pigs Marketed (in thousands)

			Response on D															
			A = Yes								*A = No*							
			B = Yes				*B = No*				*B = Yes*				*B = No*			
			C = Yes		*C = No*		*C = Yes*		*C = No*		*C = Yes*		*C = No*		*C = Yes*		*C = No*	
Education	*Pigs*	*E*	*Y*	*N*	*Y*	*N*	*Y*	*N*	*Y*	*N*	*Y*	*N*	*Y*	*N*	*Y*	*N*	*Y*	*N*
No	< 1	Y	1	0	0	0	0	0	0	0	2	1	1	2	1	1	5	3
		N	0	0	0	0	0	0	0	1	1	0	0	5	4	7	7	0
	1-2	Y	2	0	0	0	0	0	0	0	4	0	0	4	1	0	0	4
		N	0	0	0	0	0	0	0	0	0	0	0	5	0	3	4	0
	2-5	Y	3	0	0	0	0	0	0	0	3	0	0	1	2	0	1	1
		N	1	0	0	0	0	0	0	3	0	0	0	2	0	1	4	0
	> 5	Y	2	0	0	0	0	0	0	0	1	0	1	0	0	1	0	2
		N	1	0	0	2	1	0	1	6	0	1	1	1	0	0	6	0
Some	< 1	Y	3	0	0	0	0	0	0	0	4	0	1	1	0	0	2	11
		N	0	0	0	0	0	0	0	0	4	0	1	2	4	6	14	0
	1-2	Y	0	0	0	0	0	0	0	0	2	0	0	1	0	0	1	6
		N	0	0	0	0	1	0	0	1	2	1	0	4	2	7	14	0
	2-5	Y	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	3
		N	1	0	0	0	0	0	0	0	0	0	0	5	0	4	4	0
	> 5	Y	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2
		N	1	1	0	0	0	1	0	10	0	0	0	4	1	2	4	0

NOTE: Items are as follows: A = professional consultant; B = veterinarian; C = state or local extension service; D = magazines; E = feed companies and reps.

instance, that when a logit model with random effects holds, then so must a quasi-symmetric loglinear model.

For the pig farmer data, Loughin and Scherer (1998) proposed a weighted chi-square test and a bootstrap test of the hypothesis that the probability of selecting any given veterinary information source is identical among levels of a predictor variable. As a by-product of presenting modeling strategies, we show in Section 5 how model comparisons provide simpler chi-square tests of this hypothesis. Section 6 compares the modeling strategies, summarizes connections among them, and makes recommendations. The final section suggests possible extensions of the models. Throughout the article, we discuss the use of software, illustrating with SAS, for performing the analyses.

2. A MARGINAL LOGIT MODEL APPROACH

Let Y denote the qualitative response variable, having c categories with multiple potential choices among them. We refer to the c separate component binary variables that determine the outcome of Y as *items*. A 2^c contingency table cross classifies the items, and each subject's response profile $(y_1, y_2, \ldots, y_c)$ contributes to a cell count in this table. Let X denote a column vector of predictor variables. At each of the ℓ possible combinations of settings of X, we assume an independent multinomial distribution for the counts in the cells of the 2^c response profiles. At $X=x$, let $\pi_k(x)$ denote the probability that item k is in the set of categories selected, that is, the probability of responding "yes" on the kth item, $k=1, \ldots, c$. Then, $\{(\pi_k(x), 1-\pi_k(x)), k=1, \ldots, c\}$ are the c one-dimensional marginal distributions for the 2^c cross classification of responses at $X = x$. Table 1 contains sample information corresponding to these one-dimensional margins. For instance, for $X =$ (no college, < 1,000 pigs), $\hat{\pi}_1(x) = 2/42 = .048, \ldots \hat{\pi}_5(x) = 17/42 = .405$. Note that $\Sigma_k \pi_k(x)$ can take value between 0 and c, not 0 and 1.

The representation of the data in the 2^c response profile form of Table 2 converts the original c responses, which are not exclusive or exhaustive as shown in Table 1, into a set of responses that are exclusive and exhaustive. This is necessary because one can use counts in Table 2 but not in Table 1 to estimate the joint multinomial probability distribution defined at the 2^c cells for $(y_1, y_2, \ldots, y_c)$ at each setting of X and consequently the dependence among response choices. In most applications, however, the multinomial probabilities for the ℓ separate 2^c tables that underlie data in the form of Table 2 are not of primary interest. Rather, one would like to describe how X affects whether one selects any given category. This focus relates to the marginal probabilities $\{\pi_k(x)\}$ directly.

Logit models for the marginal probabilities have the form

$$\log\left(\frac{\pi_k(x)}{1-\pi_k(x)}\right) = \alpha_k + \beta_k x, k = 1, \ldots, c, \tag{1}$$

where β_k denotes a row vector of parameters. For a fixed k, this is an ordinary logit model for one component of the c-variate binary response. In this general form, the effect of X varies according to the outcome category k, but in some applications it may be plausible to have the same effects for a certain subset of those categories.

2.1. SIMULTANEOUS MARGINAL LOGIT MODELING

It is simple to fit a particular component of this model (for fixed k) with ordinary software for logit modeling. It is more challenging to fit all components simultaneously. The complete likelihood function refers to the $\ell \times 2^c$ multinomial cell probabilities, but the model itself applies to the marginal probabilities $\{\pi_k(x)\}$ for the items. Yet, fitting the models simultaneously is preferable since (1) some parameters may be common to the separate equations, (2) that fit provides fitted values for Table 2 that are of use for forming residuals and for checks of overall goodness of fit that apply to the entire set of models for the various items, and (3) one can then easily perform inferences that compare parameters for the logit models for different items, taking into account the common sample for those models.

One approach to fitting model (1) maximizes the multinomial likelihood for the $\ell \times 2^c$ table having the form of Table 2 while treating the model formula (1) as a set of constraint equations. Aitchison and Silvey (1958), Haber (1985), and Lang and Agresti (1994) have presented numerical algorithms for maximizing multinomial likelihoods subject to constraints. The latter two articles apply the algorithm to generalized loglinear models having the matrix form

$$C \log A\pi = X\beta. \tag{2}$$

In this context, π is the vector (with $\ell \times 2^c$ elements) of the ℓ sets of multinomial probabilities, one set for each level of X, and β is the entire vector of model parameters; for instance, when model (1) contains distinct parameters for each k, then $\beta = (\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots \alpha_c, \beta_c)'$. The matrix A contains 0 and 1 entries in such a pattern that when applied to π it forms the relevant marginal probabilities $\{\pi_k(x)\}$ to which the model applies; the matrix C contains 0 and 1 and –1 entries in such a pattern that when applied to the log marginal probabilities, it forms the marginal logits for the models. Ordinary loglinear models are the special case of model (2) in which C and A are identity matrices.

Unfortunately, the popular statistical software packages do not yet have procedures available for maximum likelihood (ML) fitting of generalized loglinear models of form (2). However, specialized programs are available. For the results quoted below, we used a function developed by and available from Prof. Joseph B. Lang of the Statistics Department, University of Iowa, for the S-PLUS software (marketed by MathSoft, Inc.; see www.mathsoft.com/splus). The algorithm works best for categorical predictors with small to moderate values for c and ℓ. In that case, one can also test the model fit using ordinary chi-square statistics comparing the observed counts in the $\ell \times 2^c$ table to the fitted values for the model. For the case of fully categorical predictors, one can approximate this ML fit using a simple weighted least squares algorithm, such as PROC CATMOD in SAS provides for (2). This approach breaks down, however, when the data are sparse, with many of the marginal totals being less than about 5.

A computationally simpler approach for estimating parameters in marginal models uses the methodology of generalized estimating equations (GEE) (Liang and Zeger 1986). This method is a multivariate extension of quasi-likelihood methods, in which the estimates are solutions of estimating equations that resemble likelihood equations but need not be, as the method does not fully specify the distribution of the data. With the GEE approach, one must provide structure only for how the variance depends on the mean and for the correlation structure of the c binary responses, without stating a particular distribution. A simple version of this method exploits the fact that good estimates of model parameters can result even if one naively treats the component responses as independent; that is, in this context whether a respondent selects outcome category A is treated as independent of whether the respondent selects outcome category B, and so forth. Although the parameter estimates can be fine under the naive independence assumption, standard errors are not. More appropriate standard errors result from an adjustment the GEE method makes using the empirical dependence for the sample data.

More generally with the GEE approach, rather than using estimates based on treating component responses as independent, one can make a "working guess" about the likely correlation structure of the item responses but again adjust the standard error to reflect what actually occurs for the sample data. For instance, an "exchangeable" working correlation structure under which correlations for all pairs of responses are identical is more flexible and realistic than the naive independence assumption. Even more realistic is an "unstructured" working correlation that permits a separate correlation for each pair of components. When c is large, however, this approach may suffer some efficiency loss because of the large number of extra parameters. In our experience, when the correlations are modest, all such working correlation structures usually lead to similar GEE estimates and standard errors, as the empirical dependence has a large impact on adjustment of the "naive" standard errors. Unless there is reason to expect dramatic differences among the correlations, we recommend using the exchangeable working correlation structure because this is a way of recognizing the dependence at the cost of a single extra parameter.

The GEE method is not likelihood based; that is, it does not specify a particular distribution for the $\ell \times 2^c$ table. Thus, likelihood-based methods are not available for testing fit, comparing models, and testing hypotheses about parameters. Instead, inference uses Wald statistics constructed using the asymptotic normality of the estimates together with their estimated covariance matrix. For instance, an estimate divided by its standard error is a Wald statistic for testing that a parameter equals 0; this test statistic has an asymptotic standard normal null distribution, and its square has an asymptotic chi-square distribution with $df = 1$. In SAS, PROC GENMOD (the procedure for generalized linear modeling with a variety of response distributions, including binomial and Poisson) can implement the GEE approach for a variety of working correlation structures for the dependence among the c items.

2.2. EXAMPLE OF SIMULTANEOUS MARGINAL MODELING

We now illustrate by applying marginal models to Table 2. For a farmer with education i and size of farm j, let $\pi_k(ij)$ denote the probability of responding "yes" on the kth item. For a baseline, we begin with the null logit model,

$$\log\left(\frac{\pi_k(ij)}{1-\pi_k(ij)}\right)=\alpha_k, i=1,2,\ j=1,\ldots,4,\ k=1,\ldots,5. \tag{3}$$

This states that for each item k, the probability of responding "yes" on that item is the same for each (education, size of farm) combination. The standard chi-square goodness-of-fit tests comparing the counts in Table 2 to the fitted values that maximize the likelihood for model (3) yield likelihood-ratio (deviance) statistic $G^2 = 78.3$ and Pearson statistic $X^2 = 73.4$; these each have $df = 35$ since model (3) describes 40 logits, namely, logits for the 5 items at each of the 8 (education, size of farm) combinations, by 5 parameters. Both statistics provide strong evidence of lack of fit. Table 2 is very sparse, so we make this conclusion cautiously, but in any case one would want to consider more complex models that include explanatory variables. Table 3 summarizes results for several such models.

Consider first the model

$$\log\left(\frac{\pi_k(ij)}{1-\pi_k(ij)}\right)=\alpha_k+\beta_{ik}^{E}+\beta_{jk}^{S}, \tag{4}$$

having additive education and size of farm effects for each item. Identifiability requires constraints such as $\beta_{2k}^{E} = \beta_{4k}^{S}$ for all k. This model assumes a lack of interaction between education and size of farm in their effects on the choice of each item but allows those effects

TABLE 3: Comparison of Fits of Several Marginal Models for Table 2; Differences Reported in Likelihood-Ratio Statistic (deviance) and in *df* Are Relative to Baseline Model With no Effects of Education or Size of Farm

Effects	*Deviance*	df	*Change in Deviance*	*Change in* df
None	78.3	35	—	—
lin.S	34.5	30	43.8	5
E	69.4	30	8.9	5
lin.S + E	26.3	25	52.0	10
lin.S + E + lin.S * E	20.7	20	57.6	15
S	27.2	20	51.1	15
S + E	18.7	15	69.6	20
S + E + S * E	0.0	0	78.3	35

NOTE: S = size of farm, E = education, lin.S = linear effect of size of farm.

to differ for each item. It appears to fit relatively well ($G^2 = 18.7$ and $X^2 = 17.8$, $df = 15$). The more complex model permitting interaction is saturated ($df = 0$) since it requires 15 parameters to describe the interaction, $(2 - 1) \times (4 - 1) = 3$ for each item. The model deleting the education terms from (4) has $G^2 = 27.2$ ($df = 20$), so there is only modest evidence of an education effect, controlling for size of farm (change in $G^2 = 8.4$, $df = 5$, p value = .13). By contrast, the model deleting the size of farm terms from model (4) has $G^2 = 69.4$ ($df = 30$), so there is very strong evidence of a size of farm effect.

Inspection of the $\{\hat{\beta}^S_{jk}\}$ for model (4) suggested roughly a linear trend across size of farm levels ($j = 1,2,3,4$) for each item. Thus, we next fitted the special case of (4),

$$\log\left(\frac{\pi_k(ij)}{1-\pi_k(ij)}\right) = \alpha_k + \beta^E_{ik} + \beta^S_k x_j, \quad (5)$$

using the equally spaced scores $\{x_j = j\}$ for size of farm. This model also fits relatively well ($G^2 = 26.3$, $df = 25$), with change in G^2 equal to 7.6 with $df = 10$ compared with the model (4) treating size of farm as qualitative rather than quantitative. Adding interaction terms permits the linear size of farm effects to vary by educational level but does not provide a significantly improved fit.

Again, there is not serious damage from deleting the education effects from model (5); the resulting fit has $G^2 = 34.5$ and $X^2 = 31.6$ (df = 30). Table 4 shows the parameter estimates and standard errors for that model. There is a strong positive size effect for item A (professional consultant), with the probability of selecting that category increasing substantially with size of farm. The estimated probability of selecting A changes from $\exp(-4.493 + 1.069(1))/[1 + \exp(-4.493 + 1.069(1))] = .03$ for the smallest farms to $\exp(-4.493 + 1.069(4))/[1 + \exp(-4.493 + 1.069(4))] = .45$ for the largest; by contrast, the change is from .30 to .40 for item B (veterinarian), .41 to .29 for item C (state or local extension service), .57 to .43 for item D (magazines), and .41 to .28 for item E (feed companies and reps), with only the last of these four changes approaching statistical significance in the usual sense.

Similar results occur with the computationally simpler GEE approach. Model parameter estimates and standard errors are very similar to those using ML, but model comparison is not as convenient because of the lack of a likelihood function. Table 4 also shows the GEE estimates for the model with linear size of farm effects and no education effect. These estimates resulted from the exchangeable working correlation structure, but very similar results occurred using either the independence or unstructured working correlations. In fact, the estimated working correlation matrix in the unstructured case has correlation estimates varying only between –0.004 and 0.212. Table 5 shows SAS code (PROC GENMOD) for the GEE analysis.

Since the education effects are nonsignificant, one could also estimate the size of farm effects by fitting the linear trend models after collapsing Table 2 over education. Results are substantively similar to those reported in Table 4.

2.3. SIMULTANEOUS VERSUS SEPARATE FITTING OF MARGINAL MODELS

Since ML fitting of these models simultaneously for the various items requires special software, one might question whether it is worth the trouble compared with simply fitting ordinary logit models separately for the five items. Indeed, fitting them separately, one obtains estimates that provide the same conclusions; for instance,

TABLE 4: Estimates for Marginal Model and Random Effects Model With Linear Size of Farm Effects

	Marginal Model				*Random Effects Model*	
	ML Estimates		*GEE Estimates*			
Parameter	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*	*Estimate*	*Standard Error*
α_1	–4.493	.566	–4.499	.646	–4.780	.616
α_2	–0.978	.274	–0.828	.281	–0.885	.312
α_3	–0.211	.267	–0.153	.274	–0.147	.304
α_4	0.475	.256	0.487	.270	0.555	.298
α_5	–0.151	.265	–0.081	.274	–0.068	.305
β_1^S	1.069	.174	1.081	.198	1.144	.191
β_2^S	0.147	.106	0.079	.110	0.080	.123
β_3^S	–0.168	.109	–0.189	.112	–0.214	.125
β_4^S	–0.185	.102	–0.221	.108	–0.248	.119
β_5^S	–0.204	.106	–0.239	.113	–0.268	.126

slope estimates are similar to those reported in Table 4. There is a slight efficiency loss, as standard errors of ML estimates are 1% to 2% larger for the separate fitting. Presumably, this efficiency loss could be larger if responses on the five items were more strongly correlated.

An advantage of simultaneous fitting using ML or the GEE approach is having the capability of common parameters for different items. To illustrate, the estimates in Table 4 suggest that the linear size of farm effect is similar for items C, D, and E. The model that has this common effect and a null education effect is slightly more parsimonious than model (5) but has essentially the same values for goodness-of-fit statistics ($G^2 = 34.5$, $df = 32$). The common ML slope estimate of $\beta_3^S = \beta_4^S = \beta_5^S$ is –.184, and the standard error of .063 compares with standard errors of about .11 for the separate-effects model.

2.4. PAIRWISE COMPARISONS OF CATEGORY SELECTION PROBABILITIES

So far, we have estimated effects of explanatory variables on whether any particular response category is chosen. One may also

TABLE 5: Example of SAS Code for Using PROC GENMOD to Implement GEE Method for Marginal Model and PROC NLMIXED to Implement ML for Random Effects Model With Linear Size of Farm Effects

```
data example;
  input    case    educ    size    a    b    c    d    e    response;
  datalines; * 5 lines for each subject, for (a, b, c, d, e) outcomes;
     1       1       1       1     0    0    0    0    1
     1       1       1       0     1    0    0    0    1
     1       1       1       0     0    1    0    0    1
     1       1       1       0     0    0    1    0    1
     1       1       1       0     0    0    0    1    1
......
   262       2       4       1     0    0    0    0    0
   262       2       4       0     1    0    0    0    0
   262       2       4       0     0    1    0    0    0
   262       2       4       0     0    0    1    0    1
   262       2       4       0     0    0    0    1    0
;
proc    genmod    data=example; class case;
   model response = a b c d e a*size  b*size  c*size  d*size  e*size
        / dist=bin noint; * binomial distribution, no intercept;
   repeated subject=case / type=exch corrw; * exchangeable corr;
run;
 ;
proc nlmixed data=example;
eta = lambda + alpha1*a+alpha2*b+alpha3*c+alpha4*d+alpha5*e
   +beta1*size*a+beta2*size*b+beta3*size*c+beta4*size*d+beta5*size*e;
p = exp(eta)/(1 + exp(eta)); * specifies logit link, since logit(p)=eta;
model y ~ binary (p) ;
random lambda ~ normal(0, sigma*sigma) subject=case;
run;
```

NOTE: PROC GENMOD = procedure for generalized linear modeling; GEE = generalized estimating equations; ML = maximum likelihood.

want to compare the probabilities of selection of various categories at fixed values of the explanatory variables. For model (5) with null education effects, for instance, at size of farm j, the odds of selecting category a are $\exp[(\alpha_a - \alpha_b) + (\beta_a^S - \beta_b^S)x_j]$ multiplied by the odds of selecting category b. To illustrate, from Table 4 the estimated odds of selecting professional consultant (category A) are $\exp[(-4.493 - 0.475) + (1.069 + 0.185)j] = \exp[-4.97 + 1.25j]$ multiplied by the estimated odds of selecting magazines (category D); thus, the estimated probability of selecting magazines is higher than the estimated probability of selecting professional consultant at all size levels except for the largest farms ($j = 4$), where the estimated probabilities are essentially the same.

3. A LOGIT RANDOM EFFECTS MODEL APPROACH

The models of the previous section apply directly to the marginal distributions for the various items, so the effects comparing the probabilities of selecting different items are "population-averaged." For instance, a fitted probability estimate $\hat{\pi}(ij)$ refers to *all* subjects at education level i and size of farm j for item k, and the model does not attempt to study potential heterogeneity among those subjects in this probability. Alternatively, one can model "subject-specific" probabilities and allow such heterogeneity by incorporating subject terms in the model. We next discuss logit random effects models of this type.

3.1. MODELING WITH RANDOM INTERCEPTS

For subject s, $s = 1, \ldots, n$, let x_s denote the value of the predictors X and let $\pi_k(x_s; s)$ denote the probability that the kth outcome category is in the selected set. A subject-specific model analog of the marginal model (1) is

$$\log\left(\frac{\pi_k(x_s; s)}{1-\pi_k(x_s; s)}\right) = \lambda_s + \alpha_k + \beta_k x_s, k = 1, \ldots, c, \quad (6)$$

with a constraint such as $\Sigma_s \lambda_s = 0$. The usual approach to fitting models of this type makes the "local independence" assumption that, given the subject term λ_s, the responses on the c items are independent. Marginally, averaging over subjects, nonnegative dependence occurs between responses on different items due to variability among $\{\lambda_s\}$. For instance, subjects with a large positive value of λ_s have a relatively high probability of a yes response for each item, whereas subjects with a large negative value of λ_s have a relatively low probability of a yes response for each item, resulting in overall positive associations for the population of subjects.

Introducing subject-specific terms increases the parameter space dramatically. It is sensible to reduce the number of parameters by using a *random effects* approach with $\{\lambda_s\}$. This approach assumes a parametric distribution for $\{\lambda_s\}$ (e.g., Anderson and Aitkin 1985). The standard assumption is that $\{\lambda_s\}$ are independent from a normal $N(0, \sigma)$ distribution, with σ an unknown parameter. Model (6) is then the special case of a *generalized linear mixed model* that is referred to as the *logistic-normal model.* To obtain the likelihood function, one eliminates $\{\lambda_s\}$ (which are unobserved) by integrating with respect to their normal distribution. This requires numerical integration methods, such as Gauss-Hermite quadrature, which replaces the integral by an approximating finite sum. The likelihood function then depends on the fixed effects parameters $\{\alpha_k, \beta_k\}$ as well as the parameter σ for the random effects distribution. This is much more manageable than a likelihood that treats $\{\lambda_s\}$ as fixed effects and has the number of parameters on the same order as the number of subjects.

As usual, one obtains the ML estimates of parameters by maximizing the likelihood function. The estimate of σ is itself of interest as a description of subject heterogeneity. When $\sigma = 0$, the model simplifies to an ordinary logit model that treats the c repeated responses by a subject as independent, that is, in the same way as if the observations occurred for c separate subjects. In that case, the model implies a marginal model of form (1) having the same $\{\alpha_k, \beta_k\}$ parameters.

Various software can fit binary-response models containing random effects, although few currently provide ML parameter estimates rather than approximations that can be rather crude. We used PROC NLMIXED in SAS, which provides ML estimates for generalized linear mixed models based on the Gauss-Hermite quadrature approximation for the likelihood function. In fitting the models reported below in section 3.3, we assumed normally distributed random effects with common σ at each (education, size of farm) combination.

3.2. EXCHANGEABILITY FOR CONDITIONAL ODDS RATIOS

Section 2.1 noted that the GEE marginal approach is often employed with an exchangeable working correlation structure as a way of representing with a single parameter the association structure among the component responses. The random effects approach also has a sort of exchangeable structure but in terms of conditional odds ratios rather than correlations.

For the response profile $(y_1, y_2, \ldots, y_c)$ for an arbitrary subject, let $g(\lambda, \alpha, \beta) = \Pi_k[1 + \exp(\lambda + \alpha_k + \beta_k x)]$ and let F denote the $N(0, \sigma)$ cumulative distribution function. For a pair of components a and b, let $t_{ab} = \Sigma_k y_k - y_a - y_b$. For the marginal distribution (averaged over subjects) at covariate values X, one can directly show that the odds ratio between responses a and b, given the other responses, equals

$$\frac{\left[\int [e^{2\lambda} e^{t_{ab}\lambda} / g(\lambda, \alpha, \beta)] dF(\lambda)\right]\left[\int [e^{t_{ab}\lambda} / g(\lambda, \alpha, \beta)] dF(\lambda)\right]}{\left[\int [e^{\lambda} e^{t_{ab}\lambda} / g(\lambda, \alpha, \beta)] dF(\lambda)\right]^2}.$$

Note that this depends on the other responses only through t_{ab}. In particular, the conditional odds ratio is the same for all pairs (a, b) with a common value of t_{ab}, and this is true regardless of the form of the random effects distribution F.

This basic random effects model has enjoyed much success for a variety of types of data. However, the implications of nonnegative marginal log odds ratios among the component responses, averaged over subjects, and exchangeability in the conditional odds ratios can be severe limitations. We further discuss this point in the next subsection.

3.3. EXAMPLES OF RANDOM EFFECTS MODELS

We now illustrate the random effects form of model for Table 2. For farmer s having education level i and size of farm j, let $\pi_k(ij; s)$ denote the probability that the selected set of response categories includes the kth one. The subject-specific model analog of the marginal null model (3) is

$$\log\left(\frac{\pi_k(ij; s)}{1-\pi_k(ij; s)}\right) = \lambda_s + \alpha_k, i=1, 2, j=1, \ldots, 4, k=1, \ldots, 5. \quad (7)$$

Marginally (averaged over subjects), this model is not equivalent to the marginal model (3) but is a special case of it since the dependence structure is required to result from a normal mixing on the logit scale. It has residual $df = 242$ since the $(2^5 - 1) = 31$ multinomial probabilities for each of the 8 (ij) combinations are predicted by five $\{\alpha_k\}$ parameters and the σ parameter from the normal random effects distribution (i.e., $df = 8(31) - 6 = 242$).

Table 6 shows a variety of more complex random effects models and the change in the deviance compared with this simple model. The model permitting interaction does not fit significantly better than the main effects model

$$\log\left(\frac{\pi_k(ij;s)}{1-\pi_k(ij;s)}\right) = \lambda_s + \alpha_k + \beta_{ik}^E + \beta_{jk}^S, i = 1, 2, j = 1, \ldots, 4, k = 1, \ldots, 5, \quad (8)$$

the change in deviance being 21.3 with $df = 15$. As with marginal modeling, size effects are adequately described by a linear trend, with the model

$$\log\left(\frac{\pi_k(ij;s)}{1-\pi_k(ij;s)}\right) = \lambda_s + \alpha_k + \beta_{ik}^E + \beta_k^S x_j, i = 1, 2, j = 1, \ldots, 4, k = 1, \ldots, 5, \quad (9)$$

having increase in deviance of 7.8 with $df = 10$ compared to (8). Again, the education effects are not significant. Table 4 shows the estimates for the model (denoted by lin.S in Table 6) containing only linear size of farm effects. Table 5 shows the SAS code for using PROC NLMIXED to fit this model. For this model, the estimated standard deviation of the random effects equals $\hat{\sigma} = .65$. The estimated sizes of farm effects and standard errors tend to be slightly larger than those for the corresponding marginal model but provide the same substantive results. Generally, the larger the random effects variability (and, consequently, the stronger the association among the repeated responses), the greater the differences tend to be between subject-specific and population-averaged effects. The subject-specific ones tend to be larger in absolute value (Zeger, Liang, and Albert 1988), but so do their standard errors, and when the models fit decently they tend to send similar messages regarding significance.

One can also fit random effects models that permit $\{\lambda_s\}$ to have a different variance for each size of farm. For these data, results are substantively similar. Incidentally, with heterogeneous variances, the

TABLE 6: Comparison of Fits of Several Random Effects Models for Table 2; Differences Reported in Likelihood-Ratio Statistic (deviance) and in *df* Are Relative to Baseline Model With No Effects of Education or Size of Farm

Effects	*Deviance*	df	*Change in Deviance*	*Change in* df
None	476.8	242	—	—
lin.S	414.0	237	62.8	5
E	467.7	237	9.1	5
lin.S + E	405.3	232	71.5	10
lin.S + E + lin.S * E	398.3	227	78.6	15
S	406.1	227	70.7	15
S + E	397.6	222	79.3	20
S + E + S * E	376.2	207	100.6	35

NOTE: S = size of farm; E = education; lin.S = linear effect of size of farm.

model without predictor effects does not satisfy model (3), whereby there are also no effects marginally. When the random effect distributions differ at different levels of education, homogeneous conditional distributions at a fixed level of the random effect do not imply homogeneous marginal distributions after averaging with respect to the subject distributions.

Unfortunately, what we have ignored so far is that all the random effects models actually fit several cells in Table 2 very poorly. Model (9) has goodness-of-fit statistics $G^2 = 405.3$ and $X^2 = 455.1$ ($df = 232$). The data collapsed over education are less sparse, but the fit is still poor ($G^2 = 329.6$, $df = 113$). For example, inspection of the data shows that none of the subjects responded "no" for all five of the source items, whereas model (7) predicts a probability of .15 of this for each (education, size of farm) combination.

This example shows a serious problem with the random effects model for this sort of application, namely, a violation of the "local independence" assumption, given the random effect. All subjects selected at least one response category, whereas many would choose none of them if the independence assumption truly held. By contrast, in Table 2 the cells corresponding to a response of yes for one item and no for the other four items occur more commonly than the random effects models predict. Given that someone does not select a particular four of the sources, they appear to be much more likely to select the fifth item than if local independence held. In most situations, respondents may psychologically feel obligated to select at least one category.

3.4. PAIRWISE COMPARISONS OF CATEGORY SELECTION PROBABILITIES

As with the marginal modeling approach, one can also use random effects models to compare the probabilities of selection of various categories at fixed values of the explanatory variables. For instance, from the estimates in Table 4 for the model with only linear size of farm effects, for farmers with the largest farms the estimated odds of selecting professional consultant (category A) are exp[−4.780 + 1.144(4) − .555 + .248(4)] = 1.26 multiplied by the estimated odds of selecting magazines (category D); again, for the other sizes of farms, category D seems a more likely choice than category A.

4. LOGLINEAR MODEL APPROACHES

We next consider loglinear models for data of the form of Table 2, as well as connections between them and random effects models. Consider c items and ℓ independent multinomial samples with sample sizes $\{n_i, i = 1, \ldots, \ell\}$ at ℓ levels of a composite variable X corresponding to the cross classification of predictor variables. Loglinear models focus on the expected frequencies $\{\mu_i(y_1, \ldots, y_c) = n_i P(A = y_1, B = y_2, \ldots, E = y_c | X = i)\}$ for the possible sequences $y = (y_1, y_2, \ldots, y_c)$ of responses.

When the null random effects model (7) with no education or size of farm effect has $\sigma = 0$, it is equivalent to the loglinear model by which the responses on the items are mutually independent, with the same probabilities for each (education, size of farm) combination. That is, for each (ij) in Table 2, the probability of a sequence y of responses for the items satisfies

$$P(A = y_1, B = y_2, C = y_3, D = y_4, E = y_5 | X_1 = i, X_2 = j) = \\ P(A = y_1)P(B = y_2)P(C = y_3)P(D = y_4)P(E = y_5).$$

This structure satisfies the loglinear model

$$\log \mu_i(y_1, \ldots, y_c) = \alpha_i + \sum_k \beta_k y_k,\ i = 1, \ldots, \ell. \tag{10}$$

For $c = 5$, this model is symbolized by (A, B, C, D, E, X) in the common loglinear notation for the sufficient marginal distributions that determine the model fit. We mention this simplistic model purely as a baseline since it is almost always implausible in practice. For Table 2, $G^2 = 487.3$ ($df = 243$).

Next, consider the random effects model with education and size of farm effects but of a completely unstructured form. When this model has $\sigma = 0$, it is equivalent to the loglinear model by which the responses on the items are mutually independent within each level of the predictors but with possibly different probabilities for each level. This is the loglinear model

$$\log \mu_i(y_1, \ldots, y_c) = \alpha_i + \sum_k \beta_{ik} y_k,\ i = 1, \ldots, \ell, \tag{11}$$

that specifies conditional independence of the items given X, symbolized when $c = 5$ by (AX, BX, CX, DX, EX). For Table 2, this has $G^2 = 387.8$ ($df = 208$).

More complex loglinear models than these focus on the association and interaction structure among the responses on the five items. These are appropriate if one wishes to study such structure, but in most applications this would not be the primary focus. A disadvantage of standard loglinear models is that they do not yield simple summaries for the effects of predictors, such as is provided by the parameters in marginal logit and random effects logit models. They also do not provide a direct comparison of probabilities of responding yes on various items at fixed settings of predictors, such as described in sections 2.4 and 3.4 for marginal models and random effects models.

4.1. QUASI-SYMMETRY MODELS

There is a type of loglinear model that does permit making this type of comparison of category selection probabilities. This is the *quasi-symmetry model*. Adapting an idea from Tjur (1982) that was discussed also by Agresti (1995), we derive this model starting with a random effects approach that is not dependent on a particular parametric family for the random effects. This model is more promising for describing multiple selection data than any of the standard loglinear models.

We start with a random effects model without specifying the structure of the predictor effects. For subject s at setting i of a set of predictors X, consider the logit model for selecting category k,

$$\log\left(\frac{\pi_k(i;s)}{1-\pi_k(i;s)}\right) = \lambda_s + \beta_{ik}, k = 1, \ldots, c. \tag{12}$$

(Here, since no structure is yet assumed for the effects, we have absorbed α_k in the β_{ik} term, for notational simplicity in the following argument.) Under the assumption of local independence, the model-based probability of a particular sequence $y = (y_1, y_2, \ldots, y_c)$ of responses for a subject having random effect value $\lambda_s = \lambda$ is

$$\prod_{k=1}^{c}\left[\frac{\exp(\lambda+\beta_{ik})}{1+\exp(\lambda+\beta_{ik})}\right]^{y_k}\left[\frac{1}{1+\exp(\lambda+\beta_{ik})}\right]^{1-y_k}.$$

Let $F_i(\bullet)$ denote an arbitrary cumulative distribution function for λ, possibly changing form across predictor levels. Then, the marginal probability of the sequence y, averaging over the population of subjects at level i, equals

$$\exp\left(\sum_{k=1}^{c}\beta_{ik}y_k\right)\int_{\lambda}\frac{\exp\{\lambda(\Sigma_k y_k)\}}{\Pi_k\{1+\exp(\lambda+\beta_{ik})\}}dF_i(\lambda).$$

Consider now the $\ell \times 2^c$ contingency table that cross classifies the ℓ possible predictor values with the 2^c possible sequence of responses $y = (y_1, \ldots, y_c)$. Since this integral depends on y only through $\Sigma_k y_k$, the corresponding expected frequencies $\{\mu_i(y_1, \ldots, y_c)\}$ in that contingency table have structure that is a special case of the loglinear model

$$\log \mu_i(y_1, \ldots, y_c) = \alpha_i + \sum_k \beta_{ik}y_k + \gamma(i; \sum_k y_k), i = 1, \ldots, \ell. \tag{13}$$

The final term of this model represents, for each i, a symmetric interaction that takes the same value for any permutation of a possible sequence of responses; for instance, when $c = 5$, this parameter takes value $\gamma(i;\ 1)$ for each of the sequences $\{(1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1)\}$. If $\{\beta_{i1} = \ldots = \beta_{ic}\}$ for a particular i, then the response probabilities are symmetric at that level; for instance, when $c = 5$, each of the sequences $\{(1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1)\}$ has the same probability. The $\gamma(\,)$ interaction parameters account for the dependence among the c responses. They represent the way this model generalizes the loglinear conditional independence model (11), which is the special case in which all these interaction parameters equal zero. The name *quasi symmetry* for this model refers to the symmetric interaction pattern in this model being modified by main effect terms $\{\beta_{ik}\}$ that may vary by k. For each level i of the predictors, this model resembles a loglinear latent class model in the sense that it satisfies a type of quasi-independence, conditional on the sum of the responses. In fact, the loglinear latent class model provides a very similar approach for these data, which we do not explore here. Also closely related is the semiparametric approach of using random effects structure such as (6) but with a discrete distribution having a fixed number of support points for the random effects distribution (Lindsay, Clogg, and Grego 1991).

Identifiability of the item parameters in model (13) requires a constraint such as $\beta_{ic} = 0$ for each i; that is, the model can describe only within-subject effects of the form $\{\beta_{ia} - \beta_{ib}\}$. For instance, one cannot use quasi-symmetry models to describe how the probability of selecting a category depends on size of farm, but one can use them to describe how, for a given size of farm, the odds of selecting one category compares with the odds of selecting another one. In this sense, this approach corresponds to the conditional ML approach of dealing with the $\{\lambda_s\}$ subject terms in model (6), which eliminates them by conditioning on their sufficient statistics (Rasch 1961); Tjur (1982) also explored the connection between conditional ML and quasi symmetry.

The quasi-symmetry model requires c interaction parameters to account for the dependence induced by subject heterogeneity, compared to the single parameter (σ) in the normal random effects model. For Table 2, for instance, the quasi-symmetry model derived from the logistic-normal null model (7) has three more parameters since it has four more interaction parameters but one fewer main effect parameter (reflecting the restriction to within-subjects effects). Thus, its residual *df* equals 239 instead of 242. Its fit is much better, however, with $G^2 = 263.4$ compared with $G^2 = 476.8$ for the logistic-normal model.

We made no assumption about the random effects distribution F_i to derive the quasi-symmetric model (13). Thus, marginally (averaging over subjects), the general logistic-normal random effects model (6) is a special case of it. The quasi-symmetry model places only the restriction that the interaction structure be symmetric rather than the particular structure corresponding to a normal or any other distribution of random effect. As a consequence, it has the potential for a much better fit. For instance, the likelihood equations for quasi-symmetry model (13) with $c = 5$ imply that the fitted counts for the item response sequences (0, 0, 0, 0, 0) and (1, 1, 1, 1, 1) are identical to the observed counts (at each setting of X), whereas there are large discrepancies between observed and fitted counts for these cells in Table 2 using the logistic-normal model.

Table 7 summarizes goodness of fit for various loglinear models fitted to Table 2. The quasi-symmetry model derived from the random effects model (8) with main effect factors for education and size of farm is

$$\log \mu_{ij}(y_1, \ldots, y_5) = \alpha_{ij} + \sum_k (\beta_{ik} + \beta_{jk}) y_k + \gamma\,(ij; \sum_k y_k),\ i = 1, 2, j = 1, \ldots, 4. \tag{14}$$

It has $G^2 = 150.7$ ($df = 188$) compared with $G^2 = 397.6$ ($df = 222$) for the logistic-normal model. The model deleting the education effects fits essentially as well ($G^2 = 153.1$, $df = 192$). The sizes of farm effects obtained with this simpler model are necessarily the same as those obtained if we collapse the table over education and then fit the quasi-symmetry model

$$\log \mu_j(y_1, \ldots, y_5) = \alpha_j + \sum_k \beta_{jk} y_k + \gamma\,(j; \sum_k y_k), j = 1, \ldots, 4. \tag{15}$$

which has $G^2 = 90.1$ ($df = 88$). This model fits the table well except for 3 of the 128 cells that have adjusted residuals exceeding 3 in absolute value. As in other cases, an even simpler model having linear effects for size of farm fits essentially as well ($G^2 = 92.8$, $df = 96$).

TABLE 7: Goodness of Fit of Several Loglinear Models for Table 2, With X_1 = Education, X_2 = Size of Farm, and X = the 8-Category Composite of X_1 and X_2

Effects	*Deviance*	df
(A, B, C, D, E, X)	487.3	243
(AX, BX, CX, DX, EX)	387.8	208
Quasi symmetry with none	263.4	239
Quasi symmetry with X_2	153.1	192
Quasi symmetry with $X_1 + X_2$	150.7	188

4.2. PAIRWISE COMPARISONS OF CATEGORY SELECTION PROBABILITIES

Compared with standard loglinear models that concentrate on summarizing association and interaction structure, a model such as (15) has the advantages of parsimony and $\{\hat{\beta}_{jk}\}$ estimates that provide simple comparison of the items within each size of farm category. For the constraint $\beta_{j5} = 0$ for all j, Table 8 reports the $\{\hat{\beta}_{jk}\}$. For instance, for farmers with the largest farms, the estimated odds of selecting professional consultant (category A) are $\exp(1.32 - 0.98) = 1.40$ multiplied by the estimated odds of selecting magazines (category D). The corresponding estimate using the linear trend model is 1.42. By contrast, again for other sizes of farms, category D seems much more likely than category A.

Although at first glance quasi-symmetry models seem to have unusual form, it is straightforward to fit them with software for generalized linear models. Table 9 illustrates, showing the use of SAS (PROC GENMOD) to fit model (14) and the related model with null education effects.

5. TESTING INDEPENDENCE BETWEEN A PREDICTOR AND A MULTIRESPONSE VARIABLE

Loughin and Scherer (1998) developed a large-sample weighted chi-square test and a small-sample bootstrap test of the hypothesis that the probability of selecting any given category is identical among the levels of a predictor. In the context of the models presented in this

TABLE 8: Estimates of Within-Subject Item Comparisons $\{\beta_{jk} - \beta_{j5}\}$, by Size of Farm j (number of pigs), for Quasi-Symmetry Model (15)

Size of Farm	β_{j1}	β_{j2}	β_{j3}	β_{j4}
< $1,000	–4.13	–0.53	0.00	0.63
1,000-2,000	–2.97	0.00	0.00	0.46
2,000-5,000	–1.47	0.00	0.00	0.42
> $5,000	1.32	0.74	0.25	0.98

NOTE: Estimates obtained setting $\beta_{j5} = 0$, so refer to differences between each item and item 5.

TABLE 9: Example of SAS Code for Quasi-Symmetry Model (14) and Corresponding Model Without Education Effects

```
data qs;
   input    educ    size    a    b    c    d    e    count ;
   sym = a + b + c + d + e; * defines symmetric interaction factor;
   datalines; * 1 line for each of 256 counts in Table 2;
0    1    1    1    1    1    1    1
0    1    1    1    1    0    1    0
0    1    1    1    0    1    1    0
......
1    4    0    0    1    0    0    2
1    4    0    0    0    1    0    4
1    4    0    0    0    0    0    0
;
proc    genmod    data=qs; * fit quasi symmetry model (14);
   class sym size;
   model count = sym|educ|size size|a size|b size|c size|d size|e
      educ|a educ|b educ|c educ|d educ|e
      / dist=poi link=log; * Poisson loglinear model;
run;
proc genmod    data=qs;    * Previous model without education effect ;
   class sym size;
   model count = sym|educ|size size|a size|b size|c size|d size|e
               / dist=poi link=log;
run;
```

article, one can test this hypothesis directly by likelihood-ratio tests comparing models with and without the predictor effect.

We illustrate by testing the hypothesis that, for each item, the probability of selection is independent of education. One cannot do this by applying ordinary chi-square statistics to Table 1 collapsed over size of farm (number of pigs) since the 453 entries in that table are not independent. One can, however, conduct a likelihood-ratio test comparing the marginal logit model

$$\log\left(\frac{\pi_k(i)}{1-\pi_k(i)}\right)=\alpha_k+\beta_{ik}^E, i=1,2, k=1,\ldots,5, \tag{16}$$

(with the constraints $\{\beta_{2k}=0, k=1,\ldots,5\}$) to the simpler model in which also $\beta_{1k}^E=0, k=1,\ldots,5$. The test statistic, which is the difference in G^2 goodness-of-fit statistics for the two models, equals 9.1 with $df=5$ ($p=.11$). It corresponds to testing independence simultaneously in each of the five 2×2 marginal tables obtained by cross classifying education with each of the (yes, no) possible responses for the items.

Alternatively, one could conduct a subject-specific test of this hypothesis by comparing the random effects model

$$\log\left(\frac{\pi_k(i;s)}{1-\pi_k(i;s)}\right)=\lambda_s+\alpha_k+\beta_{ik}^E, i=1,2, k=1,\ldots,5, \tag{17}$$

to the simpler model with $\beta_{1k}^E=0, k=1,\ldots,5$. The test statistic equals 9.2 with $df=5$ ($p=.10$). The separate models have $G^2=295.6$ ($df=51$) and $G^2=304.8$ ($df=56$). Like the random effects analyses discussed previously, this one is hampered by the poor fit of the more complex of the two models.

Finally, yet another approach is based on comparing the quasi-symmetry model

$$\log \mu_i(y_1,\ldots,y_5)=\alpha_i+\sum_k \beta_{ik}y_k+\gamma(\sum_k y_k), i=1,2, k=1,\ldots,5, \tag{18}$$

to the simpler model replacing each β_{ik} by β_k. The difference in G^2 statistics of 8.5 again has $df=5$ ($p=.13$). In summary, all three approaches provide similar results regarding the statistical significance of the effect of education on these items. For further discussion of tests of this sort using marginal models, see Agresti and Liu (1999).

6. COMMENTS AND COMPARISONS OF MODELING STRATEGIES

This article has discussed models that apply at three quite different levels of aggregation. Subject-specific models, such as the random effects model (8), operate at the finest level. Averaging over subjects yields data in the form of Table 2, to which loglinear models apply. Averaging further to obtain certain two-way marginal tables yields data in the form of Table 1, to which the parameters in marginal models such as (4) apply. Of the three model types, which offers the greatest promise for analyzing categorical variables allowing multiple category choices? This section discusses the pros and cons of each model type and makes a recommendation. First, we summarize relationships and connections among the models.

In general, a random effects model of logit form does not imply a marginal model of logit form and vice versa. In fact, if a logit random effects model holds with $\sigma > 0$, then the implied marginal model does not have logit form, although there are approximate relationships between their parameter estimates (Zeger, Liang, and Albert 1988). The only time they are exactly compatible is when the random effects model holds with $\sigma = 0$, in which case marginally the responses are independent, and the same parameters apply as in the logit model. This special case also implies a quasi-symmetry model without the interaction parameters, or simply a loglinear model of conditional independence of the item responses, given the predictors. In practice, when the random effects model has a relatively small value of $\hat{\sigma}$, estimates and standard errors are very similar from the three approaches. Regardless of the value of $\hat{\sigma}$, in our experience the substantive results are the same with each approach.

The random effects models of section 3 have the advantage that they provide a mechanism for generating a joint distribution for the dependence among responses on the various items. Because of their simple joint distribution, they have relatively few parameters for the contingency table of form Table 2. A disadvantage, however, is that this simple structure for the joint distribution may be inappropriate for most applications of this type. Subjects who respond positively to one item may be less likely to respond positively to others, whereas random effects models imply independence locally and nonnegative associations marginally.

The marginal models of section 2 have the advantage that, in not assuming subject-specific structure to generate a joint distribution, they are applicable even if assumptions such as local independence are violated. The GEE version of the marginal approach does use a working correlation matrix for the item responses to attempt to improve efficiency, but typically similar results occur for the model estimates and standard errors regardless of the choice for that matrix. The ML version of the marginal approach need not make any assumption at all about the joint distribution, although it is possible to do so to make the model more parsimonious (Lang and Agresti 1994). When the ML marginal approach does not include a model for the joint distribution, it does still account for the dependence among the item responses in obtaining estimates and standard errors; it essentially uses a saturated structure for it, and thus it has a much lower value of residual *df* than for the random effects models (compare, for instance, *df* values in Tables 3 and 6).

The quasi-symmetric loglinear models of section 4 do not impose such severe structure as the random effects models on the joint distribution of the items. Thus, they often tend to provide a better fit, as we observed for the data analyzed in this article. However, they are limited to describing within-subject effects, which is a severe limitation not shared by the marginal and random effects models. One cannot use them to describe how the probability of selecting a particular category depends on values of the explanatory variables. Moreover, although the quasi-symmetry model was motivated beginning with a logit random effects model, in its loglinear form it is not obvious how to interpret its parameters, and it takes careful thought to see how those parameters relate to those in the logit model. Thus, we do not believe that loglinear models are as useful as logit random effects models, even though some of them often fit better.

In our opinion, overall the marginal models are the most useful of these three types. They are simplest in the sense of treating the joint distribution as a nuisance and not spending much effort in modeling it. Unlike the random effects models, marginal models can fit well when local independence is badly violated. Unlike the quasi-symmetric loglinear models, one can use marginal models also for describing effects of predictors. Perhaps most important, they seem to focus on the matters of primary importance: Does the probability of choosing a particular category depend on certain explanatory variables? For fixed values of those predictors, how do the probabilities compare of selecting the various categories? Also, the overall (i.e., population-averaged) rates would seem more relevant in most applications than the subject-specific ones.

Of the two ways of fitting marginal models, we have a slight preference for the ML approach. With it, one has a likelihood function and the consequent related methods such as likelihood-based tests and confidence intervals and goodness-of-fit tests. However, the GEE approach is computationally much simpler and more widely available in current software. It provides the simplest tool at present for attacking this problem.

A limitation of all the modeling approaches is the potential complications due to sparseness of the data. Tests of model goodness of fit are best suited to a modest number c of response categories with at most a few categorical predictors. When the number of cells in the $\ell \times 2^c$ contingency table is very large, goodness-of-fit statistics for models may not have distribution close to chi-squared, but ordinary ML inference applies reasonably well to estimating the main effect parameters and to comparing nested models as long as the marginal counts at combinations of the predictors and the response categories are not too small. Sparseness is less of an issue for the GEE approach with marginal models than for the other approaches discussed in this article.

7. POSSIBLE EXTENSIONS

All the modeling approaches we have discussed can handle extensions of the multiple-category-choice problem. For instance, suppose two variables each can have multiple responses, such as Y_1 asking what types of movies one likes (comedy, drama, romance, horror, science fiction) and Y_2 asking what types of books one likes to read (fiction, biography, history, homemaking, self-help, other nonfiction). If Y_1 has c_1 categories and Y_2 has c_2 categories, the models then apply to a $2^{c_1} \times 2^{c_2}$ contingency table at each setting of explanatory variables X.

The marginal modeling approach applies simultaneously to study effects of X on components of each response variable and to study marginal associations between pairs of the components. For instance, one might test for independence between Y_1 and Y_2, at each setting of X, in the sense of simultaneous pairwise marginal independence between each component of Y_1 and each component of Y_2. The marginal modeling approach specifies independence simultaneously for c_1c_2 separate 2×2 tables, one for each such pair, at each setting of X. One can obtain ML fitting of such a model with the methodology of Lang and Agresti (1994) alluded to in section 2, but this becomes infeasible as c_1 and c_2 increase or the number of variables increases. In those cases, the GEE approach is preferable.

There are various ways of extending the random effects models for a bivariate response with multiple possible choices for each response. One approach uses a multivariate normal vector of $c_1 + c_2$ random effects, one for each component of Y_1 and one for each component of Y_2. The null hypothesis specifies a correlation of 0 between each pair of a random effect for Y_1 and a random effect for Y_2, whereas under the alternative, those correlations are unspecified. If responses among items of Y_1 are likely to be positively correlated and if responses among items of Y_2 are also likely to be positively correlated, one could simplify this approach by introducing a single random effect for each component of Y_1 (i.e., a special case of the previous model in which the first c_1 random effects are perfectly correlated) and a single random effect for each component of Y_2. Then, one would compare the model in which these two random effects are uncorrelated with one in which they may be correlated. Such a simpler model has the potential for a power improvement, if it fits reasonably well; in practice, however, it would likely often face the lack of fit problems exhibited by random effects models in this article. See Coull and Agresti (2000) for examples of multivariate logit models with vectors of random effects.

Finally, this article has concentrated on models that describe the probability of selection for any particular response category. More generally, in some applications it may be of interest to model the actual subset of categories selected. One could construct multinomial logit models to describe how this choice depends on explanatory variables, but when c is large, this approach is hindered by the large number of parameters in modeling 2^c categories with $2^c - 1$ separate logit formulas.

REFERENCES

Agresti, Alan. 1995. "Logit Models and Related Quasi-Symmetric Loglinear Models for Comparing Responses to Similar Items in a Survey." *Sociological Methods and Research* 24:68-95.

Agresti, Alan and I-Ming Liu. 1999. "Marginal Modeling of a Categorical Variable Allowing Arbitrarily Many Category Choices." *Biometrics* 55:936-43.

Aitchison, John and S. D. Silvey. 1958. "Maximum-Likelihood Estimation of Parameters Subject to Restraints." *Annals of Mathematical Statistics* 29:813-28.

Anderson, Dorothy and Murray Aitkin. 1985. "Variance Components Models With Binary Response: Interviewer Variability." *Journal of the Royal Statistical Society, Series B* 47:203-10.

Coull, Brent A. and Alan Agresti. 2000. "Random Effects Modeling of Multiple Binary Responses Using the Multivariate Binomial Logit-Normal Distribution." *Biometrics* 56:73-80.

Haber, Michael. 1985. "Maximum Likelihood Methods for Linear and Log-Linear Models in Categorical Data." *Computational Statistics and Data Analysis* 3:1-10.

Lang, Joseph and Alan Agresti. 1994. "Simultaneously Modeling Joint and Marginal Distributions of Multivariate Categorical Responses." *Journal of the American Statistical Association* 89:625-32.

Liang, Kung-Yee and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73:13-22.

Lindsay, Bruce, Clifford Clogg, and John Grego. 1991. "Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis." *Journal of the American Statistical Association* 86:96-107.

Loughin, Thomas M. and Peter N. Scherer. 1998. "Testing for Association in Contingency Tables With Multiple Column Responses." *Biometrics* 54:630-37.

Miller, Michael E., Charles S. Davis, and J. Richard Landis. 1993. "The Analysis of Longitudinal Polytomous Data: Generalized Estimating Equations and Connections With Weighted Least Squares." *Biometrics* 49:1033-44.

Rasch, Georg. 1961. "On General Laws and the Meaning of Measurement in Psychology." *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability* 4:321-33.

Richert, B. T., M. D. Tokach, R. D. Goodband, and J. L. Nelssen. 1993. "Integrated Swine Systems: 'The Animal Component'—Phase 1; The Kansas State University Survey." KSU Swine Day 1993, 172-76.

Tjur, Tue. 1982. "A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model." *Scandinavian Journal of Statistics* 9:23-30.

Zeger, Scott L., Kung-Yee Liang, and Paul S. Albert. 1988. "Models for Longitudinal Data: A Generalized Estimating Equation Approach." *Biometrics* 44:1049-60.

Alan Agresti is a distinguished professor in the Department of Statistics, University of Florida, Gainesville. His research interests are in generalized linear modeling of dis-

crete data, particularly methods of categorical data analysis, with emphasis on applications in the social sciences and biomedical sciences. His books include Categorical Data Analysis *(John Wiley, 1990),* An Introduction to Categorical Data Analysis *(John Wiley, 1996), and* Statistical Methods for the Social Sciences *(with B. Finlay, 3rd ed., Prentice Hall, 1997).*

Ivy Liu is a lecturer in the School of Mathematical and Computing Science, Victoria University, Wellington, New Zealand. She received her Ph.D. in statistics from the University of Florida in 1995. Her research interests are in categorical data analysis for sparse data, especially in various extended types of Mantel-Haenszel estimation methods for stratified categorical data.