

PubH 7407, Spring 2007: Exam II

There are five problems, each with multiple parts. Please start a new page for each problem. Put your name at the top of each page.

1. **Brief answer.** Answer the following questions.

(a) A logistic-normal GLMM with one predictor looks like

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \beta_1 x_{ij} + u_i, \quad u_1, \dots, u_n \stackrel{iid}{\sim} N(0, \sigma^2).$$

Describe a large sample test for $H_0 : \sigma = 0$.

(b) The baseline-category logit model for a multinomial response variable Y with 5 categories is given by

$$\log \left\{ \frac{P(Y = j)}{P(Y = 5)} \right\} = \alpha_j + \beta_j x, \quad j = 1, 2, 3, 4$$

for a continuous predictor x . From this model find an expression for

$$\log \left\{ \frac{P(Y = 3)}{P(Y = 2)} \right\}.$$

(c) The cumulative logit (proportional odds) model for the response in part (b) is given by

$$\log \left\{ \frac{P(Y \leq j)}{P(Y > j)} \right\} = \alpha_j + \beta x, \quad j = 1, 2, 3, 4.$$

From this model find an expression for $P(Y = 3)$.

(d) What does a log-linear model for a contingency table actually model?

(e) Starting with $[A][B][C]$, list all the possible hierarchical log-linear models one can fit to an $I \times J \times K$ table with factors A , B , and C .

2. **True/false.** Write **true** or **false** for each statement.

(a) In the log-linear model $[AB][BC]$, A is independent of C .

(b) For a categorical response $Y \in \{1, 2, \dots, J\}$, both the baseline-category logit and cumulative logit (proportional odds) models reduce to the logistic regression model when $J = 2$.

(c) The baseline-category logit model should only be fit to data involving alligators.

(d) The cumulative logit (proportional odds) model is a sensible choice for a nominal multinomial outcome.

(e) Regression effects from a model fit using the GEE approach to account for correlation in clustered responses are interpreted the same way as in a GLMM.

Output from PROC GENMOD:

```

Exchangeable Working
Correlation

Correlation    0.2168926903

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter Estimate    Standard    95% Confidence
                    Error          Limits          Z Pr > |Z|
Intercept  -0.3201    0.4111   -1.1259    0.4858   -0.78    0.4363
treat      -0.5540    0.2330   -1.0106   -0.0974   -2.38    0.0174

```

- From the GEE approach, what is the estimated correlation among responses from the same clinic?
- Carefully interpret the estimated treatment effects from both approaches in terms of odds ratios.
- Code to fit the GLM without the random effects yields the following output.

```

Fit Statistics

-2 Log Likelihood          358.2
AIC (smaller is better)   362.2

```

Which model is preferred according to AIC? What does this tell you about responses within a clinic?

- Formally test $H_0 : \sigma = 0$ in the GLMM at the $\alpha = 0.05$ level. Note that $P(\chi_1^2 > 3.84) = 0.05$. (Yes, you can figure it out from this.)
4. Let Y be political ideology on an ordinal scale from 1 = very liberal to 5 = very conservative. Consider the covariates $x_1 =$ gender (1 = female, 0 = male) and $x_2 =$ political party (1 = Democrat, 0 = Republican).
- A main effects model with a cumulative logit link gives the annotated output shown below.

Parameter	DF	Estimate	SE
Intercept1	1	-2.53	0.15
Intercept2	1	-1.54	0.13
Intercept3	1	0.17	0.12
Intercept4	1	1.01	0.12
gender femal	1	0.12	0.13
gender male	0	0.00	0.00
party dem	1	0.96	0.13
party rep	0	0.00	0.00

```

LR Statistics For Type 3 Analysis
Source DF chi-square Pr > ChiSq
gender 1 0.84 0.36
party 1 56.85 <.0001

```

Explain how gender affects political ideology with an estimated odds ratio and 95% CI.

- Give the hypotheses to which the LR statistic refers and interpret the results of the test.

(c) When an interaction term is added we get the output show below.

Parameter		DF	Estimate	SE
Intercept1		1	-2.67	0.16
Intercept2		1	-1.68	0.15
Intercept3		1	0.04	0.13
Intercept4		1	0.88	0.14
gender	femal	1	0.37	0.18
gender	male	0	0.00	0.00
party	dem	1	1.27	0.20
party	rep	0	0.00	0.00
gender*party	femal dem	1	-0.51	0.24
gender*party	femal rep	0	0.00	0.00
gender*party	male dem	0	0.00	0.00
gender*party	male rep	0	0.00	0.00

Provide estimated odds ratios and 95% confidence intervals characterizing how gender affects political ideology. Is the interaction significant at the 5% level?

5. A contingency table has the factors A , B , C , D , and E . The log-linear model $[AB][BC][AC][CDE]$ fits the data well.
- What conditional independence does this model imply?
 - Draw an association graph for this model.
 - We are only interested in factors A, B , and C . Can we collapse the table over D and E and still retain valid log-linear model inferences among A, B , and C ? For example, do we have to worry about Simpson's paradox occurring?