Chapter 8: multinomial regression and discrete survival analysis

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 770: Categorical Data Analysis

8.1 Baseline category logit models for nominal responses

Let Y be categorical with J levels. Let $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$.

Logit models pair each response $Y = j$ with the baseline category, here $Y = J$:

$$\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j'\mathbf{x}, \text{ for } j = 1, \ldots, J-1.$$

The parameters are $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{J-1})$ and $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J-1})$. If each $\boldsymbol{\beta}_j$ is $p-1$ dimensional, then there are $(J-1) + (p-1)(J-1) = (J-1)p$ parameters to estimate.

For a fixed $\mathbf{x}$, the ratio of probabilities $Y = a$ versus $Y = b$ is given by

$$\frac{\pi_a(\mathbf{x})}{\pi_b(\mathbf{x})} = \exp\left\{(\alpha_a - \alpha_b) + (\boldsymbol{\beta}_a - \boldsymbol{\beta}_b)'\mathbf{x}\right\}.$$

This model reduces to ordinary logistic regression when $J = 2$.

Alligator food!

Lake	Gender	Size (m)	Primary food choice: Fish	Invertebrate	Reptile	Bird	Other
Hancock	Male	≤ 2.3	7	1	0	0	5
		> 2.3	4	0	0	1	2
	Female	≤ 2.3	16	3	2	2	3
		> 2.3	3	0	1	2	3
Oklawaha	Male	≤ 2.3	2	2	0	0	1
		> 2.3	13	7	6	0	0
	Female	≤ 2.3	3	9	1	0	2
		> 2.3	0	1	0	1	0
Trafford	Male	≤ 2.3	3	7	1	0	1
		> 2.3	8	6	6	3	5
	Female	≤ 2.3	2	4	1	1	4
		> 2.3	0	1	0	0	0
George	Male	≤ 2.3	13	10	0	2	2
		> 2.3	9	0	0	1	2
	Female	≤ 2.3	3	9	1	0	1
		> 2.3	8	1	0	0	1

From $n = 219$ alligators caught in four Florida lakes

Let L be lake, G be gender, and S size. Each alligator will have $\mathbf{x} = (L, G, S)$ as a predictor for what they primarily eat. The probability of food source being (fish, invertebrate, reptile, bird, other) is $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$, where $\boldsymbol{\pi} = \boldsymbol{\pi}(\mathbf{x})$ according to the baseline logit model.

```
data gator;
input lake gender size food count ;
datalines;
1 1 1 1 7
1 1 1 2 1
1 1 1 3 0
1 1 1 4 0
1 1 1 5 5
...
4 2 2 1 8
4 2 2 2 1
4 2 2 3 0
4 2 2 4 0
4 2 2 5 1
;
proc logistic; freq count; class lake size gender / param=ref;
  model food(ref='1') = lake size gender lake*size size*gender lake*gender / link=glogit
  aggregate scale=none selection=backward;
```

Backwards elimination

We have

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	lake*size	12	5	0.7025	1.0000
2	size*gender	4	4	1.3810	0.8475
3	lake*gender	12	3	8.0477	0.7814
4	gender	4	2	2.1850	0.7018

The final model has lake and size as additive effects; gender is unimportant to predicting primary food source. GOF and Type III analyses:

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	52.4785	44	1.1927	0.1784
Pearson	58.0140	44	1.3185	0.0765

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
lake	12	35.4890	0.0004
size	4	18.7593	0.0009

GOF statistics

Unless we specify the variables to aggregate over (e.g. `aggregate=(lake size)` in the model statement), the SAS GOF tests use all variables in *the original model* we worked backwards from to determine the saturated model. The original model has three effects: lake, gender, and size.

The saturated model has 16 sets (4 lakes $\times$ 2 genders $\times$ 2 sizes) of 5 probabilities associated with it. Since the probabilities in each row add to one, that implies $16 \times 4 = 64$ parameters total in the saturated model.

However, the *reduced model* from SAS only has the effects lake and size! The number of parameters in the reduced model is 20: 12 lake effects, 4 size effects, and 4 intercepts.

Since we've determined that gender is not important, we should not include gender in the saturated model when determining lack of fit.

$L + S$ fit

We refit the model including only those predictors $L + S$ in the final model:

```
proc logistic; freq count; class lake size / param=ref;
  model food(ref='1') = lake size / link=glogit aggregate scale=none;
```

yielding

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	17.0798	12	1.4233	0.1466
Pearson	15.0429	12	1.2536	0.2391

The $df = 12$ is the number of parameters in the saturated model *aggregated over only lake and gender* minus the number in the reduced regression model. The saturated model has four parameters (five probabilities that add to one) for each level of lake and size: $4 \times 4 \times 2 = 32$ df. The regression model (still) has $p = 20$ effects so there are $32 - 20 = 12$ df for testing model fit.

There is little replication here so the p-values are suspect. However, $17.1 < 2 \times 12$ and $15.0 < 2 \times 12$, so there is no evidence of gross LOF.

Regression parameter estimates

Analysis of Maximum Likelihood Estimates

Parameter		food	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		2	1	-1.5490	0.4249	13.2890	0.0003
Intercept		3	1	-3.3139	1.0528	9.9081	0.0016
Intercept		4	1	-2.0931	0.6622	9.9894	0.0016
Intercept		5	1	-1.9043	0.5258	13.1150	0.0003
lake	1	2	1	-1.6583	0.6129	7.3216	0.0068
lake	1	3	1	1.2422	1.1852	1.0985	0.2946
lake	1	4	1	0.6951	0.7813	0.7916	0.3736
lake	1	5	1	0.8262	0.5575	2.1959	0.1384
lake	2	2	1	0.9372	0.4719	3.9443	0.0470
lake	2	3	1	2.4583	1.1179	4.8360	0.0279
lake	2	4	1	-0.6532	1.2021	0.2953	0.5869
lake	2	5	1	0.00565	0.7766	0.0001	0.9942
lake	3	2	1	1.1220	0.4905	5.2321	0.0222
lake	3	3	1	2.9347	1.1161	6.9131	0.0086
lake	3	4	1	1.0878	0.8417	1.6703	0.1962
lake	3	5	1	1.5164	0.6214	5.9541	0.0147
size	1	2	1	1.4582	0.3959	13.5634	0.0002
size	1	3	1	-0.3513	0.5800	0.3668	0.5448
size	1	4	1	-0.6307	0.6425	0.9635	0.3263
size	1	5	1	0.3316	0.4483	0.5471	0.4595

Theoretical and fitted models

The theoretical model is

$$\log\left(\frac{\pi_I}{\pi_F}\right) = \alpha_2 + \beta_{21} I\{L=1\} + \beta_{22} I\{L=2\} + \beta_{23} I\{L=3\} + \beta_{24} I\{S=1\}$$

$$\log\left(\frac{\pi_R}{\pi_F}\right) = \alpha_3 + \beta_{31} I\{L=1\} + \beta_{32} I\{L=2\} + \beta_{33} I\{L=3\} + \beta_{34} I\{S=1\}$$

$$\log\left(\frac{\pi_B}{\pi_F}\right) = \alpha_4 + \beta_{41} I\{L=1\} + \beta_{42} I\{L=2\} + \beta_{43} I\{L=3\} + \beta_{44} I\{S=1\}$$

$$\log\left(\frac{\pi_O}{\pi_F}\right) = \alpha_5 + \beta_{51} I\{L=1\} + \beta_{52} I\{L=2\} + \beta_{53} I\{L=3\} + \beta_{54} I\{S=1\}$$

The estimated model is

$$\log\left(\frac{\hat{\pi}_I}{\hat{\pi}_F}\right) = -1.55 - 1.66 I\{L=1\} + 0.94 I\{L=2\} + 1.12 I\{L=3\} + 1.46 I\{S=1\}$$

$$\log\left(\frac{\hat{\pi}_R}{\hat{\pi}_F}\right) = -3.31 + 1.24 I\{L=1\} + 2.46 I\{L=2\} + 2.93 I\{L=3\} - 0.35 I\{S=1\}$$

$$\log\left(\frac{\hat{\pi}_B}{\hat{\pi}_F}\right) = -2.09 + 0.70 I\{L=1\} - 0.65 I\{L=2\} + 1.09 I\{L=3\} - 0.63 I\{S=1\}$$

$$\log\left(\frac{\hat{\pi}_O}{\hat{\pi}_F}\right) = -1.90 + 0.82 I\{L=1\} + 0.01 I\{L=2\} + 1.52 I\{L=3\} + 0.33 I\{S=1\}$$

Interpretation

Note that $e^{\beta_{ji}}$ is how the odds of eating food in category j ($j = 2, 3, 4, 5$) changes (relative to eating fish) with levels of lake relative to George ($i = 1, 2, 3$) or alligator size relative to large ($i = 4$).

For example $e^{\beta_{32}}$ is how the odds of eating primarily reptiles ($j = 3$) changes for lake Oklawaha ($i = 2$) versus lake George, holding size constant. Here, we estimate $e^{2.46} \approx 11.7$. There's probably proportionately more reptiles (relative to fish) in Oklawaha than George!

Similarly, $e^{\beta_{44}}$ is how the odds of eating primarily birds ($j = 4$) changes for smaller alligators ($i = 4$), holding lake constant. We estimate this as $e^{-0.63} \approx 0.53$. The odds of eating primarily birds (relative to fish) increases by $e^{0.63} \approx 1.88$ for large alligators.

Let's answer some more questions

How does the odds of choosing invertebrates over fish change from small to large alligators in a given lake? Answer:

$$\frac{\frac{\pi_I}{\pi_F}(S=1, L=l)}{\frac{\pi_I}{\pi_F}(S=2, L=l)} = e^{\beta_{24}}.$$

From the regression coefficients we have $e^{1.4582} = 4.298$. The odds of primarily eating invertebrates over fish are four times greater for smaller alligators than larger alligators. Is this significant? Yes, $p = 0.0002$ for $H_0 : \beta_{24} = 0$. What about a 95% CI?

A 95% CI is part of the output automatically generated by PROC LOGISTIC.

Odds ratios

Odds Ratio Estimates

Effect	food	Point Estimate	95% Wald Confidence Limits	
lake 1 vs 4	2	0.190	0.057	0.633
lake 1 vs 4	3	3.463	0.339	35.343
lake 1 vs 4	4	2.004	0.433	9.266
lake 1 vs 4	5	2.285	0.766	6.814
lake 2 vs 4	2	2.553	1.012	6.437
lake 2 vs 4	3	11.685	1.306	104.508
lake 2 vs 4	4	0.520	0.049	5.490
lake 2 vs 4	5	1.006	0.219	4.608
lake 3 vs 4	2	3.071	1.174	8.032
lake 3 vs 4	3	18.815	2.111	167.717
lake 3 vs 4	4	2.968	0.570	15.447
lake 3 vs 4	5	4.556	1.348	15.400
size 1 vs 2	2	4.298	1.978	9.339
size 1 vs 2	3	0.704	0.226	2.194
size 1 vs 2	4	0.532	0.151	1.875
size 1 vs 2	5	1.393	0.579	3.354

So $e^{1.4582} = 4.298$ with a 95% CI of $(1.98, 9.34)$.

Reptiles vs. birds

How about reptiles over birds?

$$\frac{\frac{\pi_R}{\pi_B}(S=1, L=l)}{\frac{\pi_R}{\pi_B}(S=2, L=l)} = e^{\beta_{34}-\beta_{44}} = e^{-0.35-(-0.63)} \approx 1.3.$$

This is an exponentiated contrast, but I'd suggest simply refitting the model with "birds" as the reference category to get a CI:

```
proc logistic; freq count; class lake size / param=ref;
* type 4 is birds and type 3 is reptiles;
 model food(ref='4') = lake size / link=glogit aggregate scale=none;
```

and pull out

```
                              Odds Ratio Estimates

                                          Point          95% Wald
          Effect          food         Estimate      Confidence Limits
          size 1 vs 2     3               1.322        0.272       6.421
```

The odds of primarily eating primarily reptiles over birds are 1.3 times greater for small alligators than large ones. Does this mean that small (or large) alligators eat more reptiles than birds? Hint: what if the odds are 13 and 10? What if they are 0.13 and 0.10?

In terms of probabilities...

Odds are 13 and 10:

$$1.3 = \frac{\left[\frac{13/14}{1/14}\right]}{\left[\frac{10/11}{1/11}\right]},$$

implies more reptiles than birds for small and large alligators!

Odds are 0.13 and 0.10:

$$1.3 = \frac{\left[\frac{13/113}{100/113}\right]}{\left[\frac{1/11}{10/11}\right]},$$

implies more birds than reptiles for small and large alligators!

Odds ratios tell you nothing about the actual probabilities underlying the events of interest.

Fitted multinomial probabilities

Figure 8.1, p. 297: note that the curves have to add up to one. As the alligator gets bigger, she increasingly chooses "fish" and "other" over "invertebrates" (worms, snails, bugs, etc.) Would you?

Let $\mathbf{x}$ be a fixed covariate vector and say n observations are sampled at $\mathbf{x}$. Then $\mathbf{n} = (n_1, \ldots, n_J) \sim \text{mult}(n, \boldsymbol{\pi}(\mathbf{x}))$ where $\boldsymbol{\pi}(\mathbf{x}) = (\pi_1(\mathbf{x}), \ldots, \pi_J(\mathbf{x}))$ and

$$\pi_j(\mathbf{x}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j'\mathbf{x})}{1 + \sum_{h=1}^{J-1} \exp(\alpha_h + \boldsymbol{\beta}_h'\mathbf{x})}.$$

For example, each row in the alligator food table is a different multinomial vector $\mathbf{n} = (n_1, n_2, n_3, n_4, n_5)$ corresponding to a unique $\mathbf{x}$ yielding probabilities $\boldsymbol{\pi}(\mathbf{x})$ through the baseline logit model.

8.2 Cumulative logit models for ordinal responses

Let Y be *ordinal* with J categories. The *proportional odds model* stipulates

$$\log \frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})} = \alpha_j + \boldsymbol{\beta}'\mathbf{x} \text{ for } j = 1, \ldots, J-1.$$

There are only $(J-1)+(p-1)$ parameters to estimate rather than $p(J-1)$ with the nominal model.

The odds for $Y \leq j$ is allowed to change with j through α_j. However, the effect of covariates $\mathbf{x}$ on odds $Y \leq j$ is *independent of* j. Note that $P(Y \leq J)/(Y > J)$ is $1/0$ and undefined.

This model reduces to ordinary logistic regression when $J = 2$.

Model, restated

Restated, the odds of $Y \leq j$ at $\mathbf{x}_1$ divided by the odds of $Y \leq j$ at $\mathbf{x}_2$ are, under the model:

$$\log \frac{P(Y \leq j|\mathbf{x}_1)/P(Y > j|\mathbf{x}_1)}{P(Y \leq j|\mathbf{x}_2)/P(Y > j|\mathbf{x}_2)} = \boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2).$$

This is the log *cumulative odds ratio*.

The odds of making response $\leq j$ at $\mathbf{x}_1$ are $e^{\boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2)}$ times the odds at $\mathbf{x}_2$, *independent of the level* j.

Note that e^{β_j} is how the odds of $Y \leq j$ change when increasing the predictor x_j by one.

Mental impairment example

$Y = 1, 2, 3, 4$ is degree of impairment (well, mild symptom formation, moderate symptom formation, impaired) for $n = 40$ randomly sampled people in Alachua County, Florida.

We wish to relate Y to $L =$ number and severity of important life events (new baby, new job, divorce, death in family within 3 years), $S =$ socioeconomic status (low=0 or high=1).

Y	S	L	Y	S	L	Y	S	L	Y	S	L
1	1	1	1	1	9	1	1	4	1	1	3
1	0	2	1	1	0	1	0	1	1	1	3
1	1	3	1	1	7	1	0	1	1	0	2
2	1	5	2	0	6	2	1	3	2	0	1
2	1	8	2	1	2	2	0	5	2	1	5
2	1	9	2	0	3	2	1	3	2	1	1
3	0	0	3	1	4	3	0	3	3	0	9
3	1	6	3	0	4	3	0	3			
4	1	8	4	1	2	4	1	7	4	0	5
4	0	4	4	0	4	4	1	8	4	0	8
4	0	9									

SAS code

```
data impair;
input mental ses life;
datalines;
1 1 1
1 1 9
...
4 0 8
4 0 9
;
proc logistic;
   model mental = life ses / aggregate scale=none;
```

Output:

```
                    Response Profile

          Ordered                          Total
            Value          mental      Frequency
                1               1             12
                2               2             12
                3               3              7
                4               4              9

Probabilities modeled are cumulated over the lower Ordered Values.

          Score Test for the Proportional Odds Assumption

          Chi-Square        DF      Pr > ChiSq
              2.3255         4          0.6761
```

8.2.5 GOF test vs. more general model

The test of the proportional odds assumption tests the fitted model against the alternative

$$\log \frac{P(Y \leq j|\mathbf{x})}{P(Y > j|\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j'\mathbf{x} \text{ for } j = 1, \ldots, J-1.$$

The proportional odds model is a special case where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_{J-1} = \boldsymbol{\beta}$. The drop in model parameters is $p(J-2)$, here $2(4-2) = 4$ *df*. We accept that the simpler cumulative logit model fits, and find no gross LOF from the Pearson GOF:

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	57.6833	52	1.1093	0.2732
Pearson	57.0248	52	1.0966	0.2937

Number of unique profiles: 19

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.9442	2	0.0069
Score	9.1431	2	0.0103
Wald	8.5018	2	0.0143

Parameter estimates

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 1	1	-0.2818	0.6231	0.2045	0.6511
Intercept 2	1	1.2129	0.6511	3.4700	0.0625
Intercept 3	1	2.2095	0.7171	9.4932	0.0021
life	1	-0.3189	0.1194	7.1294	0.0076
ses	1	1.1111	0.6143	3.2719	0.0705

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
life	0.727	0.575	0.919
ses	3.038	0.911	10.126

The fitted model is

$$\begin{aligned}
\log\left\{\frac{P(Y=1)}{P(Y=2,3,4)}\right\} &= -0.28 - 0.32\ \text{life} + 1.11\ \text{ses} \\
\log\left\{\frac{P(Y=1,2)}{P(Y=3,4)}\right\} &= 1.21 - 0.32\ \text{life} + 1.11\ \text{ses} \\
\log\left\{\frac{P(Y=1,2,3)}{P(Y=4)}\right\} &= 2.21 - 0.32\ \text{life} + 1.11\ \text{ses}
\end{aligned}$$

Interpretation

Note that $\alpha_1 < \alpha_2 < \alpha_3$ must hold because this series of odds can only increase. The event of interest is $Y \leq j$, i.e. being "less impaired."

The odds of being "less impaired" increases by $e^{1.11} = 3.0$ for high socioeconomic status versus low (for fixed number of life events). The odds of being "less impaired" decreases by a factor of $e^{-0.32} = 0.73$ for every additional life event that occurred in the previous 3 years (for fixed socioeconomic status).

Put another way, for high ses the odds of being *more impaired* is only $1/3$ that of low ses (so low ses is bad). The odds of being more impaired increases by $1/0.727 = 1.38$ for every additional life event.

Low SES is equivalent to about 3.5 life events: $[e^{0.3189}]^{3.5} \approx 3.05$.

8.2.3 Latent variable motivation*

It is useful to think of each individual having an underlying *continuous* "impairment" score Y^*. This *latent* continuous variable determines the observed level of impairment via cutoffs

$$\begin{array}{ccc} Y^* < \alpha_1 & \Rightarrow & Y = 1 \\ \alpha_1 < Y^* < \alpha_2 & \Rightarrow & Y = 2 \\ \alpha_2 < Y^* < \alpha_3 & \Rightarrow & Y = 3 \\ \alpha_3 < Y^* & \Rightarrow & Y = 4 \end{array}$$

The latent score has a regression model

$$Y^* = -\beta_1 \text{ life} - \beta_2 \text{ ses} + \epsilon,$$

where ϵ is subject-to-subject error and distributed standard logistic

$$f(\epsilon) = \frac{e^\epsilon}{(1 + e^\epsilon)^2}.$$

Latent variable formulation

This formulation is equivalent to the proportional odds model. To see this, note that the CDF of the logistic distribution is $F(\epsilon) = \frac{e^{\epsilon}}{(1+e^{\epsilon})}$. Then

$$
\begin{aligned}
P(Y=1) &= P(Y^* \le \alpha_1) \\
&= P(-\beta_1 \text{life} - \beta_2 \text{ses} + \epsilon \le \alpha_1) \\
&= P(\epsilon \le \alpha_1 + \beta_1 \text{life} + \beta_2 \text{ses}) \\
&= \frac{e^{\alpha_1 + \beta_1 \text{life} + \beta_2 \text{ses}}}{(1 + e^{\alpha_1 + \beta_1 \text{life} + \beta_2 \text{ses}})}
\end{aligned}
$$

yielding

$$
\log\left\{\frac{P(Y=1)}{P(Y=2,3,4)}\right\} = \alpha_1 + \beta_1 \text{life} + \beta_2 \text{ses}.
$$

Repeat for $P(Y \le 2)$ and $P(Y \le 3)$.
See Figure 8.5 (p. 304).

Generalizations

- **8.3 & 8.3.1** discusses other models

$$P(Y \leq j|\mathbf{x}) = F(\alpha_j + \boldsymbol{\beta}'\mathbf{x}),$$

  where F is probit or complimentary log-log. These can also be fit in PROC LOGISTIC (LINK=CPROBIT or LINK=CCLOGLOG) and may improve fit over proportional odds (i.e. the cumulative logit model).

- 8.3.8 adds *covariate-specific* dispersion:

$$P(Y \leq j|\mathbf{x}) = F\left(\frac{\alpha_j + \boldsymbol{\beta}'\mathbf{x}}{\exp(\boldsymbol{\gamma}'\mathbf{x})}\right).$$

  This model can also improve model fit and can be fit with some work in PROC NLMIXED. See Figure 8.7 (p. 313).

Let $Y = 1, \ldots, J$ be ordered stages that one *must* pass through in order starting with the first (e.g. egg, larva or caterpillar, pupa or chrysalis, and adult butterfly). Often the categories are time periods (e.g. years 1, 2, 3, 4). Let

$$h_j(\mathbf{x}) = P(Y = j | Y \geq j).$$

This probability is termed the *hazard* of ending up in stage $Y = j$. If $Y = j$ indicates death in time period j, then this is the risk of dying right at j given that you've made it up to j.

Let $P(Y = j) = \pi_j(\mathbf{x})$. Then

$$h_j(\mathbf{x}) = \frac{\pi_j(\mathbf{x})}{\pi_j(\mathbf{x}) + \pi_{j+1}(\mathbf{x}) + \cdots + \pi_J(\mathbf{x})}.$$

Hazard regression

The logit model specifies

$$\log\left\{\frac{h_j(\mathbf{x})}{1-h_j(\mathbf{x})}\right\} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}.$$

This is an example of a hazard regression model.

Note that

$$\frac{h_j(\mathbf{x})}{1-h_j(\mathbf{x})} = \frac{P(Y=j)/P(Y\geq j)}{P(Y>j)/P(Y\geq j)} = \frac{\pi_j}{\pi_{j+1}+\pi_{j+2}+\cdots+\pi_J}.$$

This latter expression is called a *continuation ratio*.

The model thus specifies

$$\log\left\{\frac{\pi_j}{\pi_{j+1}+\pi_{j+2}+\cdots+\pi_J}\right\} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}.$$

Proportional hazards

If we specify a cumulative log-log link instead,

$$h_j(\mathbf{x}) = 1 - \exp\{-\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x})\},$$

$$\begin{aligned}
P(Y \geq j) &= P(Y \geq 1, Y \geq 2, \ldots, Y \geq j) \\
&= P(Y \geq j | Y \geq j-1) \cdots P(Y \geq 2 | Y \geq 1) \\
&= \frac{P(Y \geq j)}{P(Y \geq j-1)} \frac{P(Y \geq j-1)}{P(Y \geq j-2)} \cdots \frac{P(Y \geq 2)}{P(Y \geq 1)} \\
&= [e^{-e^{\alpha_{j-1}}}]^{e^{\boldsymbol{\beta}'\mathbf{x}}} [e^{-e^{\alpha_{j-2}}}]^{e^{\boldsymbol{\beta}'\mathbf{x}}} \cdots [e^{-e^{\alpha_1}}]^{e^{\boldsymbol{\beta}'\mathbf{x}}} \\
&= \left[e^{-\sum_{i=1}^{j-1} e^{\alpha_i}}\right]^{e^{\boldsymbol{\beta}'\mathbf{x}}} \quad \text{for fixed } \mathbf{x}.
\end{aligned}$$

Let $S_{\mathbf{x}}(j) = P(Y \geq j | \mathbf{x})$. Then

$$S_{\mathbf{x}}(j) = S_0(j)^{e^{\boldsymbol{\beta}'\mathbf{x}}},$$

where $S_0(j) = e^{-\sum_{i=1}^{j-1} e^{\alpha_i}}$, the proportional hazards model.

Generalizations

Both models are written

$$h_j(\mathbf{x}) = F(\alpha_j + \boldsymbol{\beta}'\mathbf{x}).$$

Generalizations:

- If the affect of covariates changes with time (or stage), we can generalize to

  $$h_j(\mathbf{x}) = F(\alpha_j + \boldsymbol{\beta}_j'\mathbf{x}).$$

  This can be fit as a series of nested binomial regression models.
- If time-dependent covariates $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J\}$ are measured (e.g. blood pressure, amount of television watched, etc.) then we can fit

  $$h_j(\mathbf{x}) = F(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_j).$$

In general, it *is not* straightforward to fit these models in SAS; see http://support.sas.com/faq/045/FAQ04512.html.

Fitting

To form the likelihood note that

$$P(Y = j|\mathbf{x}) = h_j(\mathbf{x}) \prod_{k=1}^{j-1} (1 - h_k(\mathbf{x})).$$

Then

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{n} P(Y = j|\mathbf{x}).$$

Also note that

$$h_J(\mathbf{x}) = P(Y = J|Y \geq J) = 1.$$

Recall for the logit model $h_j(\mathbf{x}) = \frac{e^{\alpha_j + \boldsymbol{\beta}'\mathbf{x}}}{1+e^{\alpha_j + \boldsymbol{\beta}'\mathbf{x}}}$.

The proportional odds (cumulative logit) model for this type of data is also applicable and provides a different type of inference.

Example

Consider a widely-analyzed data set first presented by Feigl and Zelen (1965) on $n = 33$ leukemia patients. The outcome is $Y = 1$ for death within the year after diagnosis, $Y = 2$ for death within the second year, and $Y = 3$ for within 3 or more years (only one made it to 4 years). The predictors are $x_1 = 0$ for AG$-$ and $x_1 = 1$ for AG$+$ and $x_2 = \log(\text{wbc})$, log white blood cell count. AG$+$ indicates the presence of Auer rods and/or significant granulature of leukemic bone marrow cells.

PROC NLMIXED has routines built in to maximize certain types of likelihoods, and is especially useful when random effects are present. We will use it to build and maximize the continuation ratio (hazard regression) likelihood.

```
data leuk1;
 input x1 x2 y @@;
 datalines;
1    6.62 3 1    7.74 2 1    8.36 2 1    7.86 3 1    8.69 1 1    9.25 3
1    9.21 3 1    9.74 1 1    8.59 1 1    8.85 3 1    9.14 2 1   10.37 1
1   10.46 1 1   10.85 1 1   11.51 1 1   11.51 1 1   11.51 2 0    8.38 2
0    8.00 2 0    8.29 1 0    7.31 1 0    9.10 1 0    8.57 1 0    9.21 1
0    9.85 1 0   10.20 1 0   10.23 1 0   10.34 1 0   10.16 1 0    9.95 1
0   11.27 1 0   11.51 1 0   11.51 1
;
proc nlmixed; * effect of beta constant across stages;
 parms a1=-7 a2=-6 b1=-3 b2=1; * started with a1=0 a2=1 b1=0 b2=0;
 p1=exp(a1+x1*b1+x2*b2); p2=exp(a2+x1*b1+x2*b2);
 if (y=1) then z=(p1/(1+p1));
 if (y=2) then z=(1/(1+p1))*(p2/(1+p2));
 if (y=3) then z=(1/(1+p1))*(1/(1+p2));
 if (z>1e-8) then ll=log(z); else ll=-1e100;
 model y ~ general(ll);
```

We obtain

The NLMIXED Procedure

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
a1	-6.7090	3.4093	33	-1.97	0.0575	0.05	-13.6454	0.2273	-3.67E-8
a2	-5.8987	3.2094	33	-1.84	0.0751	0.05	-12.4282	0.6309	-1.07E-8
b1	-2.6455	0.9875	33	-2.68	0.0114	0.05	-4.6545	-0.6364	-4.32E-8
b2	0.9677	0.3813	33	2.54	0.0161	0.05	0.1919	1.7436	-4.49E-7

Interpretation

Clearly both AG factor and log(wbc) affect the probability of moving from stage to stage. Given that a subject has made it to a given stage, the odds of dying in that stage (instead of moving on) are estimated to significantly decrease by a factor of $e^{-2.6455} = 0.071$ when x_1 changes from 0 to 1. The odds of dying increase by $e^{0.9677} = 2.63$ for each unit increase in log(wbc).

Model	-2 Log L	AIC
Hazard regression, logistic, AG+WBC $\boldsymbol{\beta}$ same across stages	39.2	47.2
Hazard regression, logistic, AG+WBC $\boldsymbol{\beta}_j$ changes $j = 1, 2$	38.2	50.2
Hazard regression, logistic, AG+WBC+AG*WBC $\boldsymbol{\beta}$ same across stages	39.0	49.0
Proportional odds (cumulative logit) AG+WBC	39.9	47.9
Proportional odds (cumulative logit) AG+WBC+AG*WBC	39.7	49.7
Hazard regression, cumulative log-log, AG+WBC $\boldsymbol{\beta}$ same across stages	64.3	56.3

Comments

- The proportional odds model is trivially fit: `proc logistic; model y=x1 x2;`.
- We can test the logistic continuation ratio model with the effect of the covariates changing with stage by comparing the decrease in `-2 Log L` to the increase in parameters. The simpler model has (β_1, β_2) increased to $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$, a $df = 2$ parameter difference. $39.2 - 38.2 = 1.0$; $P(\chi^2_2 > 1.0) = 0.61$; the simpler (constant β) model is preferred.
- This confirms the best choice from AIC: the additive logistic hazard regression model with AG and log(wbc).

8.5 Discrete choice models

Let Y be nominal with J levels. Associated with each level $Y = j$ are aspects of $Y = j$ that might affect the probability $P(Y = j)$. There also might be subject-specific covariates.

Example: Choosing breakfast. Let $Y = 1$ indicate nothing (breakfast is skipped), $Y = 2$ be cereal, and $Y = 3$ eggs. For each individual $i = 1, \ldots, n$, there are two covariates: x_{ij} is how long choice j takes to fix and eat and z_i is a crude hunger level ($z_i = 0$ for not hungry, $z_i = 1$ for hungry).

i	x_{i1}	x_{i2}	x_{i3}	z_i	j
1	0	15	25	1	3
2	0	10	15	0	1
3	0	5	25	0	2
4	0	15	10	1	3
5	0	5	25	1	1
6	0	20	45	1	1
7	0	10	10	1	3
8	0	10	20	0	1
9	0	15	15	1	2
10	0	10	25	1	1

SAS data format for PROC MDC

```
data breakfast;
input id decision nothing cereal eggs time hungryn hungryc hungrye;
datalines;
1  0 1 0 0  0  1  0  0
1  0 0 1 0 15  0  1  0
1  1 0 0 1 15  0  0  1
2  1 1 0 0  0  0  0  0
2  0 0 1 0 10  0  0  0
2  0 0 0 1 15  0  0  0
3  0 1 0 0  0  0  0  0
3  1 0 1 0  5  0  0  0
3  0 0 0 1 25  0  0  0
4  0 1 0 0  0  1  0  0
4  0 0 1 0 15  0  1  0
4  1 0 0 1 10  0  0  1
5  1 1 0 0  0  1  0  0
5  0 0 1 0  5  0  1  0
5  0 0 0 1 25  0  0  1
6  1 1 0 0  0  1  0  0
6  0 0 1 0 20  0  1  0
6  0 0 0 1 45  0  0  1
7  0 1 0 0  0  1  0  0
7  0 0 1 0 10  0  1  0
7  1 0 0 1 10  0  0  1
8  1 1 0 0  0  0  0  0
8  0 0 1 0 10  0  0  0
8  0 0 0 1 20  0  0  0
9  0 1 0 0  0  1  0  0
9  1 0 1 0 15  0  1  0
9  0 0 0 1 15  0  0  1
10 1 1 0 0  0  1  0  0
10 0 0 1 0 20  0  1  0
10 0 0 0 1 30  0  0  1
;
```

Discrete choice model

Let $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$ be the times for person i. Ignoring hunger, a simple *discrete choice model* for these data looks like:

$$\pi_j(\mathbf{x}_i) = P(Y_i = j|\mathbf{x}_i) = \frac{\exp(\beta x_{ij})}{\sum_{h=1}^{3} \exp(\beta x_{ih})}.$$

The odds of choosing eggs over nothing for person i is function of how much longer it takes to cook eggs for this person

$$\frac{\pi_3}{\pi_1}(\mathbf{x}_i) = e^{\beta(x_{i3} - x_{i1})}.$$

Can modify to allow the actual available choices to differ by person! For example, some people never eat eggs; for that person the denominator would sum only over $h = 1, 2$.

Note, only preparation time affects choice! One might want to also include an overall *preference*, e.g. some people don't like cereal!

SAS code & output

```
proc mdc data=breakfast;
 model decision=time / type=clogit nchoice=3;
 id id; * clogit is conditional logit here, not cumulative as in proc logistic;
run;
                              The MDC Procedure

                           Conditional Logit Estimates
                              Parameter Estimates

                                        Standard                  Approx
        Parameter      DF      Estimate      Error    t Value    Pr > |t|
        time            1       -0.0684     0.0407      -1.68      0.0930
```

Although not significant, the odds of choosing one breakfast over another increases by 7% for every minute *less* it takes to cook; $e^{0.0684} \approx 1.07$.

This model is *much* simpler than the baseline-category logit model!

Different proportions like different breakfasts

The previous model implies that if preparation was *the same* for nothing, cereal, or eggs, each would be chosen with probability one-third. However, the three choices are likely preferred in different proportions when time is not a factor. Consider the model:

$$\pi_j(\mathbf{x}_i) = P(Y_i = j|\mathbf{x}_i) = \frac{\exp(\beta_{0j} + \beta_1 x_{ij})}{\sum_{h=1}^{3} \exp(\beta_{0h} + \beta x_{ih})}.$$

Need to set one 'intercept' equal to zero, say $\beta_{01} = 0$.

SAS code & output

```
proc mdc data=breakfast; * nothing is baseline;
 model decision=time cereal eggs / type=clogit nchoice=3;
 id id;
run;
```

The MDC Procedure

Conditional Logit Estimates

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
time	1	-0.2496	0.1417	-1.76	0.0783
cereal	1	1.7441	1.5853	1.10	0.2713
eggs	1	3.7103	2.2059	1.68	0.0926

Holding preparation time constant, choosing eggs is $e^{3.71} \approx 41$ times more likely than nothing. When time is not held constant we have for person i

$$\frac{\pi_3}{\pi_1}(\mathbf{x}_i) = e^{\beta_{03}-\beta_{01}} e^{\beta(x_{i3}-x_{i1})}.$$

The odds of choosing one breakfast over another increases by 28% for every minute *less* it takes to cook; $e^{0.2496} \approx 1.28$. Again, it does not matter which two breakfasts we consider when discussing odds.

Hunger can affect breakfast choice

Finally, we can include how hungry someone is. Hunger should affect different choices differently.

$$\pi_j(\mathbf{x}_i, z_i) = P(Y_i = j|\mathbf{x}_i, z_i) = \frac{\exp(\beta_{0j} + \beta_1 x_{ij} + \beta_{2j} z_i)}{\sum_{h=1}^{3} \exp(\beta_{0h} + \beta_1 x_{ih} + \beta_{2h} z_i)}.$$

Again, set $\beta_{21} = 0$.

The hunger effect is modeled exactly as it is in a baseline-category logit model. Hunger affects odds of choosing one choice over another differently, depending on the two breakfast choices we are comparing.

SAS code & output

```
proc mdc data=breakfast;
 model decision=time cereal eggs hungryc hungrye / type=clogit nchoice=3;
 id id;
run;
```

The MDC Procedure

Conditional Logit Estimates

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
time	1	-0.2236	0.1320	-1.69	0.0902
cereal	1	1.1297	1.6483	0.69	0.4931
eggs	1	-11.1631	1386	-0.01	0.9936
hungryc	1	0.6924	1.9175	0.36	0.7180
hungrye	1	15.2171	1386	0.01	0.9912

Interpretation? Note that there are only 10 individuals here.

Comments

- Discrete choice models are appropriate when aspects of the choices themselves affect the probability of them being chosen (e.g. time taken, distance traveled, cost, ease of use, etc.)
- Multinomial baseline-category logits are appropriate when aspects of the choosers affect the probability of choosing among the choices (e.g. gender, age, how hungry, etc.)
- Both aspects can be incorporated into PROC MDC.
- Special case is when $\mathbf{x}_1 = \cdots \mathbf{x}_n = \mathbf{x}$ for all i. For example, the time spent preparing cereal and eggs is the same for all people.
- The discrete-choice model has fewer parameters and simpler interpretation than baseline-category logit models.