Categorical Data Analysis

Chapter 7

Deyuan Li
School of Management
Fudan University

Fall 2015

Outline

7.1.1 Baseline-Category Logits

Let $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$, $j = 1, 2, ...J$. Then $\sum_{j=1}^{J} \pi_j(\mathbf{x}) = 1$.

Logit models pair each response category with a baseline category (often the last one or the most common one):

$$\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \beta_j' \mathbf{x}, \quad j = 1, 2, ..., J. \tag{1}$$

Of course,

$$\log \frac{\pi_a(\mathbf{x})}{\pi_b(\mathbf{x})} = \log \frac{\pi_a(\mathbf{x})}{\pi_J(\mathbf{x})} - \log \frac{\pi_b(\mathbf{x})}{\pi_J(\mathbf{x})}.$$

With categorical predictors, X^2 and G^2 goodness-of-fit statistics provide a model check when data are not sparse.

When an explanatory variable is continuous or the data are sparse, X^2 and G^2 are still valid for comparing nested models differing by few terms.

7.1.2 Alligator Food Choice Example

分为美国短吻鳄（密河鳄）和中国短吻鳄（扬子鳄）。
不挑食，基本上能捉的都吃：鱼类，昆虫，蜗牛和介虫；龟，哺乳类动物，鸟类和爬行动物；牛和鹿，美洲豹和熊。

7.1.2 Alligator Food Choice Example

TABLE 7.1 Primary Food Choice of Alligators

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

Source: Data courtesy of Clint Moore, from an unpublished manuscript by M. F. Delaney and C. T. Moore.

7.1.2 Alligator Food Choice Example

Y=food choice (fish, invertebrate, reptile, bird, other)
L=Lake (Hancock, Oklawaha, Trafford, George);
G=Gender (male, female);
S=Size (≤ 2.3, > 2.3);

Data are sparse since 219 observations scattered among 80 cells. Thus, G^2 is more reliable for comparing models than testing fit.

Model ( ): no predictor;
Model ($L + S$): considering lake (L) and size (S) effects;
Model ($G + L + S$): considering gender (G), lake (L) and size (S) effects.

7.1.2 Alligator Food Choice Example

TABLE 7.2 Goodness of Fit of Baseline-Category Logit Models for Table 7.1

Model[a]	G^2	X^2	df
( )	116.8	106.5	60
(G)	114.7	101.2	56
(S)	101.6	86.9	56
(L)	73.6	79.6	48
$(L+S)$	52.5	58.0	44
$(G+L+S)$	50.3	52.6	40
Collapsed over G			
( )	81.4	73.1	28
(S)	66.2	54.3	24
(L)	38.2	32.7	16
$(L+S)$	17.1	15.0	12

[a] G, gender; S, size; L, lake of capture. See the text for details.

7.1.2 Alligator Food Choice Example

$G^2[()|(G)] = 116.8 - 114.7 = 2.1$ and
$G^2[(L+S)|(G+L+S)] = 52.5 - 50.3 = 2.2$, each based on df=4, implies simplifying by collapsing the table over gender.

Other analysis, not presented here, show that adding interaction terms including G do not improve the fit significantly.

Table 7.3 lists fitted values for model (L+S) for the collapsed table (see next page).

7.1.2 Alligator Food Choice Example

TABLE 7.3 Observed and Fitted Values for Study of Alligator's Primary Food Choice

Lake	Size of alligator (meters)	Primary Food Choice				
		Fish	Invertebrate	Reptile	Bird	Other
Hancock	≤ 2.3	23 (20.9)	4 (3.6)	2 (1.9)	2 (2.7)	8 (9.9)
	> 2.3	7 (9.1)	0 (0.4)	1 (1.1)	3 (2.3)	5 (3.1)
Oklawaha	≤ 2.3	5 (5.2)	11 (12.0)	1 (1.5)	0 (0.2)	3 (1.1)
	> 2.3	13 (12.8)	8 (7.0)	6 (5.5)	1 (0.8)	0 (1.9)
Trafford	≤ 2.3	5 (4.4)	11 (12.4)	2 (2.1)	1 (0.9)	5 (4.2)
	> 2.3	89 (8.6)	7 (5.6)	6 (5.9)	3 (3.1)	5 (5.8)
George	≤ 2.3	16 (18.5)	19 (16.9)	1 (0.5)	2 (1.2)	3 (3.8)
	> 2.3	17 (14.5)	1 (3.1)	0 (0.5)	1 (1.8)	3 (2.2)

7.1.2 Alligator Food Choice Example

Absolute values of standard Pearson residuals comparing the observed and fitted values exceed 2 in only two of the 40 cells and exceed 3 in none of the cells. So, this fit seems adequate. Fish is the baseline category.

TABLE 7.4 Estimated Parameters in Logit Model for Alligator Food Choice, Based on Dummy Variable for First Size Category and Each Lake Except Lake George[a]

			Lake		
Logit[b]	Intercept	Size $\leq$ 2.3	Hancock	Oklawaha	Trafford
$\log(\pi_I/\pi_F)$	-1.55	1.46 (0.40)	-1.66 (0.61)	0.94 (0.47)	1.12 (0.49)
$\log(\pi_R/\pi_F)$	-3.31	-0.35 (0.58)	1.24 (1.19)	2.46 (1.12)	2.94 (1.12)
$\log(\pi_B/\pi_F)$	-2.09	-0.63 (0.64)	0.70 (0.78)	-0.65 (1.20)	1.09 (0.84)
$\log(\pi_O/\pi_F)$	-1.90	0.33 (0.45)	0.83 (0.56)	0.01 (0.78)	1.52 (0.62)

[a] SE values in parentheses.

I, invertebrate; *R*, reptile; *B*, bird; *O*, other; *F*, fish.

7.1.2 Alligator Food Choice Example

For example, the prediction equation for log odds of selecting invertebrates instead of fish is

$$\log(\hat{\pi}_I/\hat{\pi}_F) = -1.55 + 1.46s - 1.66z_H + 0.94z_O + 1.12z_T.$$

Here, $s = 1$ for size ≤ 2.3 meters and 0 otherwise; z_H, z_O and z_T are dummy variables.

For a given lake, for small alligators the estimated odds that primary food choice was invertebrates instead of fish are $\exp(1.46) = 4.3$ times the estimated odds for large alligators; the Wald 95% confidence interval is $\exp[1.46 \pm 1.96(0.396)]$.

7.1.2 Alligator Food Choice Example

Other information from Table 7.4 are available.

For example,

$$\begin{aligned}\log(\hat{\pi}_I/\hat{\pi}_O) &= \log(\hat{\pi}_I/\hat{\pi}_F) - \log(\hat{\pi}_O/\hat{\pi}_F)\\ &= \big(-1.55 + 1.46s - 1.66z_H + 0.94z_O + 1.12z_T\big)\\ &\quad - \big(-1.90 + 0.33s + 0.83z_H + 0.01z_O + 1.52z_T\big)\\ &= 0.35 + 1.13s - 2.48z_H + 0.93z_O - 0.39z_T.\end{aligned}$$

7.1.3 Estimated Response Probabilities

(1) implies that

$$\pi_j(\mathbf{x}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j'\mathbf{x})}{1 + \sum_{h=1}^{J-1} \exp(\alpha_h + \boldsymbol{\beta}_h'\mathbf{x})}, \quad j = 1, 2, ..., J, \qquad (2)$$

with $\alpha_J = 0$ and $\beta_J = 0$, so that

From Table 7.4, the estimated probability that a large alligator in Lake Hancock has invertebrates as the primary food choice is

$$\hat{\pi}_I = \frac{e^{-1.55-1.66}}{1 + e^{-1.55-1.66} + e^{-3.31+1.24} + e^{-2.09+0.70} + e^{-1.90+0.83}}$$
$$= 0.023.$$

The estimated probabilities for reptile, birds, other and fish are 0.072, 0.141, 0.194 and 0.570.

7.1.4 Fitting of Baseline-Category Logit Models

For subject $i = 1, 2, ..., n$, let

$\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iJ})$ represent the multinomial trial for subject i;
$y_{ij} = 1$ if response is in category j. $\sum_j y_{ij} = 1$;
$\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ denote explanatory variable for subject i;
$\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{jp})'$ denote parameters for the j-th logit.

Then the contribution to the log likelihood by subject i is

$$\begin{aligned}\log\left[\prod_{j=1}^{J}\pi_j(\mathbf{x}_i)^{y_{ij}}\right] &= \sum_{j=1}^{J-1} y_{ij}\log\pi_j(\mathbf{x}_i) + \left(1-\sum_{j=1}^{J-1}y_{ij}\right)\log\left[1-\sum_{j=1}^{J-1}\pi_j(\mathbf{x}_i)\right] \\ &= \sum_{j=1}^{J-1} y_{ij}\log\frac{\pi_j(\mathbf{x}_i)}{1-\sum_{j=1}^{J-1}\pi_j(\mathbf{x}_i)} + \log\left[1-\sum_{j=1}^{J-1}\pi_j(\mathbf{x}_i)\right].\end{aligned}$$

7.1.4 Fitting of Baseline-Category Logit Models

Assume observations are independent, then the log likelihood is

$$
\begin{aligned}
&\log \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \pi_j(\mathbf{x}_i)^{y_{ij}} \right] \\
&= \sum_{i=1}^{n} \left\{ \sum_{j=1}^{J-1} y_{ij}(\alpha_j + \boldsymbol{\beta}'_j \mathbf{x}_i) - \log\left[1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \boldsymbol{\beta}'_j \mathbf{x}_i)\right]\right\} \\
&= \sum_{j=1}^{J-1} \left[\alpha_j \left(\sum_{i=1}^{n} y_{ij} \right) + \sum_{k=1}^{p} \beta_{jk} \left(\sum_{i=1}^{n} x_{ik} y_{ij} \right)\right] \\
&\quad - \sum_{i=1}^{n} \log\left[1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \boldsymbol{\beta}'_j \mathbf{x}_i)\right].
\end{aligned}
$$

7.1.4 Fitting of Baseline-Category Logit Models

The log likelihood is concave and the Newton-Raphson method yields the ML parameter estimates.

Their asymptotic standard errors are square roots of diagonal elements of the inverse information matrix.

Outline

7.2.1 Cumulative logits

Denote

$$P(Y \leq j|\mathbf{x}) = \pi_1(\mathbf{x}) + \cdots \pi_j(\mathbf{x}), \quad j = 1, 2, ..., J.$$

Then *cumulative logits* are defined as

Note that if Y is not ordinal categorical response variable, it makes no sense to consider $P(Y \leq j|\mathbf{x})$.

7.2.2 Proportional Odds Model

Proportional odds models:

$$\text{logit}[P(Y \le j|\mathbf{x})] = \alpha_j + \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, 2, ..., J-1.$$

Each cumulative logit has its own intercept but same effects β.

Two comments: (1). α_j increases as j increases.
(2). $\text{logit}[P(Y \le j|\mathbf{x}_1)] - \text{logit}[P(Y \le j|\mathbf{x}_2)] = \boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2)$.

Similar to Section 7.1.4, the likelihood function is

$$\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\pi_j(\mathbf{x}_i)^{y_{ij}}\right] = \prod_{i=1}^{n}\left[\prod_{j=1}^{J}\left(P(Y \le j \mid \mathbf{x}_i) - P(Y \le j-1 \mid \mathbf{x}_i)\right)^{y_{ij}}\right]$$

$$= \prod_{i=1}^{n}\left[\prod_{j=1}^{J}\left(\frac{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_i)} - \frac{\exp(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{x}_i)}\right)^{y_{ij}}\right],$$

viewed as a function of $(\{\alpha_j\}, \boldsymbol{\beta})$. $\Rightarrow$ obtain the ML estimators.

7.2.4 Mental Impairment Example

Table 7.5 comes from a study of mental health for a random sample of adult residents.

Y=Mental Impairment (well, mild symptom, moderate symptom, impairment);

x_1=number of important life events such as birth of child, new job, divorce, or death of family;

x_2=SES (socioeconomic status, 1=high, 0=low).

7.2.4 Mental Impairment Example

TABLE 7.5 Mental Impairment by SES and Life Events

Subject	Mental Impairment	SES[a] x_2	Life Events x_1	Subject	Mental Impairment	SES[a] x_2	Life Events x_1
1	Well	1	1	21	Mild	1	9
2	Well	1	9	22	Mild	0	3
3	Well	1	4	23	Mild	1	3
4	Well	1	3	24	Mild	1	1
5	Well	0	2	25	Moderate	0	0
6	Well	1	0	26	Moderate	1	4
7	Well	0	1	27	Moderate	0	3
8	Well	1	3	28	Moderate	0	9
9	Well	1	3	29	Moderate	1	6
10	Well	1	7	30	Moderate	0	4
11	Well	0	1	31	Moderate	0	3
12	Well	0	2	32	Impaired	1	8
13	Mild	1	5	33	Impaired	1	2
14	Mild	0	6	34	Impaired	1	7
15	Mild	1	3	35	Impaired	0	5
16	Mild	0	1	36	Impaired	0	4
17	Mild	1	8	37	Impaired	0	4
18	Mild	1	2	38	Impaired	1	8
19	Mild	0	5	39	Impaired	0	8
20	Mild	1	5	40	Impaired	0	9

[a] 0, low; 1, high.

7.2.4 Mental Impairment Example

The main-effects model is

$$\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \quad j = 1, 2, 3.$$

Table 7.6 is taken from SAS output. ($J = 4$).

TABLE 7.6 Output for Fitting Cumulative Logit Model to Table 7.5

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
2.3255	4	0.6761

Parameter	Estimate	Std Error	Like. Ratio 95% Conf Limits		Chi-Square	Pr > Chi Sq
Intercept1	−0.2819	0.6423	−1.5615	0.9839	0.19	0.6607
Intercept2	1.2128	0.6607	−0.0507	2.5656	3.37	0.0664
Intercept3	2.2094	0.7210	0.8590	3.7123	9.39	0.0022
life	−0.3189	0.1210	−0.5718	−0.0920	6.95	0.0084
ses	1.1112	0.6109	−0.0641	2.3471	3.31	0.0689

7.2.4 Mental Impairment Example

The parameter estimates yield estimated logits and hence the estimates of $P(Y \leq j)$, $P(Y > j)$ or $P(Y = j)$. *(In fact, it means $P(Y \leq j|\mathbf{x})$).*

We illustrate for subjects at the mean life events scores of $x_1 = 4.275$ with low SES ($x_2 = 0$). Since $\hat{\alpha}_1 = -0.282$, the estimated probability of response *well* is

$$\hat{P}(Y = 1) = \hat{P}(Y \leq 1) = \frac{\exp[-0.282 - 0.319(4.275)]}{1 + \exp[-0.282 - 0.319(4.275)]} = 0.16.$$

Outline

7.3 Ordinal Responses: Cumulative Link Models

Let G be a distribution function and denote G^{-1} as its inverse function. The *cumulative link* model is

$$G^{-1}[P(Y \leq j|\mathbf{x})] = \alpha_j + \boldsymbol{\beta}'\mathbf{x},$$

or equivalent to

$$P(Y \leq j|\mathbf{x}) = G[\alpha_j + \boldsymbol{\beta}'\mathbf{x}].$$

The logit link function is $G^{-1}(u) = \log(u/(1-u))$.

7.3.1 Types of cumulative links

Cumulative probit model:

$$\Phi^{-1}[P(Y \le j|\mathbf{x})] = \alpha_j + \boldsymbol{\beta}'\mathbf{x}.$$

(i.e. $G = \Phi$ standard normal distribution function).

Proportional hazards model:

$$\log\{-\log[1 - P(Y \le j|\mathbf{x})]\} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}.$$

It has the property (survival analysis)

$$P(Y > j|\mathbf{x_1}) = [P(Y > j|\mathbf{x_2})]^{\exp[\boldsymbol{\beta}'(\mathbf{x_1}-\mathbf{x_2})]}.$$

The link function is called *complementary log-log link*.

7.3.3 Life Table Example

Table 7.7 shows the life-length distribution for US residents in 1981, by race and gender.

TABLE 7.7 Life-Length Distribution of U.S. Residents (Percent),[a] 1981

	Males		Females	
Life Length	White	Black	White	Black
0−20	2.4 (2.4)	3.6 (4.4)	1.6 (1.2)	2.7 (2.3)
20−40	3.4 (3.5)	7.5 (6.4)	1.4 (1.9)	2.9 (3.4)
40−50	3.8 (4.4)	8.3 (7.7)	2.2 (2.4)	4.4 (4.3)
50−60	17.5 (16.7)	25.0 (26.1)	9.9 (9.6)	16.3 (16.3)
Over 65	72.9 (73.0)	55.6 (55.4)	84.9 (84.9)	73.7 (73.7)

[a] Values in parentheses are fit of proportional hazards (i.e., complementary log-log link) model.

Source: Data from *Statistical Abstract of the United States* (Washington, DC: U.S. Bureau of the Census, 1984), p. 69.

7.3.3 Life Table Example

Y=Life length (group 1, 2, 3, 4, 5);
G=Gender (1=female, 0=male);
R=Race (1=black, 0=white).

The sample distribution function increases slowly at small to moderate ages but increases sharply at older ages. This suggests the log-log link.

Table 7.7 contains the fitted distributions for the model

$$\log\{-\log[1 - P(Y \le j | G = g, R = r)]\} = \alpha_j + \beta_1 g + \beta_2 r.$$

Estimates: $\hat{\beta}_1 = -0.658$ and $\hat{\beta}_2 = 0.626$. The fitted cdf's satisfy

$$P(Y > j | G = 0, R = r) = [P(Y > j | G = 1, R = r)]^{\exp(0.658)}.$$

Outline

7.4.1 Alternative-categories logits

Cumulative probabilities are not necessary for ordinal responses. We discuss alternative logit model and a simpler model that resembles ordinary regression.

The adjacent-categories logits are

These logits are a basic set equivalent to the baseline-category logits. The connections are

$$\log \frac{\pi_j}{\pi_J} = \log \frac{\pi_j}{\pi_{j+1}} + \log \frac{\pi_{j+1}}{\pi_{j+2}} + \cdots + \log \frac{\pi_{J-1}}{\pi_J}$$

and

$$\log \frac{\pi_j}{\pi_{j+1}} = \log \frac{\pi_j}{\pi_J} - \log \frac{\pi_{j+1}}{\pi_J}.$$

7.4.1 Alternative-categories logits

The adjacent-categories logit model is

$$\log \frac{\pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, 2, ..., J-1.$$

Its equivalent baseline-category logit model is

$$\begin{aligned}\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} &= \sum_{k=j}^{J-1} \alpha_k + \boldsymbol{\beta}'(J-j)\mathbf{x} \\ &= \alpha_j^* + \boldsymbol{\beta}'\mathbf{u}_j, \quad j = 1, ..., J-1\end{aligned}$$

with $\mathbf{u}_j = (J-j)\mathbf{x}$.

7.4.2 Job Satisfaction Example

TABLE 7.8 Job Satisfaction and Income, Controlling for Gender

		Job Satisfaction			
Gender	Income (dollars)	Very Dissatisfied	A Little Satisfied	Moderately Satisfied	Very Satisfied
Female	< 5000	1	3	11	2
	5000−15,000	2	3	17	3
	15,000−25,000	0	1	8	5
	$> 25{,}000$	0	2	4	2
Male	< 5000	1	1	2	1
	5000−15,000	0	3	5	1
	15,000−25,000	0	0	7	3
	$> 25{,}000$	0	1	9	6

Source: 1991, General Social Survey, National Opinion Research Center.

The income score $x = (1, 2, 3, 4)$ and g=1(female) or 0(male). The model is

$$\log(\pi_j/\pi_{j+1}) = \alpha_j + \beta_1 x + \beta_2 g, \quad j = 1, 2, 3.$$

7.4.2 Job Satisfaction Example

This model is equivalent to

$$\log(\pi_j/\pi_4) = \alpha_j^* + \beta_1(4-j)x + \beta_2(4-j)g, \quad j = 1, 2, 3.$$

The ML estimation implies that $\hat{\beta}_1 = -0.389$ (SE=0.155) and $\hat{\beta}_2 = 0.045$ (SE=0.314).

Two comments:

- $\hat{\beta}_1 < 0$ means the odds of lower job satisfaction decreases as income increases;
- $G^2 = 12.6$ with $df = 19$, implies the linear trend model for income effect and a lack of interaction between income and gender seems adequate.

7.4.3 Continuation-ratio logits

Continuation-ratio logits are defined as

$$\log \frac{\pi_j}{\pi_{j+1} + \cdots + \pi_J}, \quad j = 1, 2, ..., J - 1$$

or as

$$\log \frac{\pi_{j+1}}{\pi_1 + \cdots + \pi_j}, \quad j = 1, 2, ..., J - 1.$$

The continuation-ratio logit model is

7.4.4 Developmental toxicity study with pregnant mice

TABLE 7.9 Outcomes for Pregnant Mice in Developmental Toxicity Study

Concentration (mg/kg per day)	Response		
	Nonlive	Malformation	Normal
0 (controls)	15	1	281
62.5	17	0	225
125	22	7	283
250	38	59	202
500	144	132	9

[a]Based on results in C. J. Price et al., *Fund. Appl. Toxicol.* **8**:115–126 (1987). I thank Louise Ryan for showing me these data.

7.4.4 Developmental toxicity study with pregnant mice

We fitted the continuation-ratio logit models

$$\log \frac{\pi_1(x_i)}{\pi_2(x_i) + \pi_3(x_i)} = \alpha_1 + \beta_1 x_i, \quad \log \frac{\pi_2(x_i)}{\pi_3(x_i)} = \alpha_2 + \beta_2 x_i.$$

The ML estimates are $\hat{\beta}_1 = 0.0064$ (SE=0.0004) and $\hat{\beta}_1 = 0.0173$ (SE=0.0012).

The likelihood-ratio statistics are $G^2 = 5.78$ for $j = 1$ and $G^2 = 6.06$ for $j = 2$, each based on $df = 3$.

Their sum, $G^2 = 11.84$ (or similarly $X^2 = 9.76$), with $df = 6$, summarizes the fit.

7.4.5 Mean response models for ordered response

For scores $v_1 \leq v_2 \leq \cdots \leq v_J$, let $M(\mathbf{x}) = \sum_j v_j \pi_j(\mathbf{x})$ denotes the mean response. The model

$$M(\mathbf{x}) = \alpha + \boldsymbol{\beta}' \mathbf{x}$$

assumes a linear relationship between the mean and the explanatory variables.

With $J = 2$, it is the linear probability model (Section 4.2.1).

7.4.6 Job satisfaction example revisited

Data are in Table 7.8. Modeling the mean of Y=job satisfaction using income x and gender g (1=female, 0=male).

For simplicity, we use job satisfaction scores and income scores (1,2,3,4).

The model has ML fit,

$$\hat{M} = 2.59 + 0.181x - 0.030g,$$

with SE=0.069 for income and 0.145 for gender.

Outline

7.5.1 Using Multinomial Models to Test Conditional Independence

In Section 6.3.2 we introduced the Cochran-Mantel-Haenszel (CMH) test of conditional independence for $2 \times 2 \times K$ tables.

In this section, we discuss the conditional independence test for $I \times J \times K$ tables.

Treating Z as a nominal control factor, we discuss four cases with (Y, X) as (ordinal, ordinal), (ordinal, nominal), (nominal, ordinal), (nominal, nominal).

7.5.1 Using Multinomial Models to Test Conditional Independence

- *Y ordinal*, *X ordinal*. Let $\{x_i\}$ be ordered scores. The model

  has the same linear trend for the X effect in each partial table. XY conditional independence $\Leftrightarrow H_0 : \beta = 0$. Wald, score, likelihood ratio statistics, as $n \to \infty$, converge to chi-squared distribution with $df = 1$.

- *Y ordinal*, *X nominal*. Treating X as a factor, the model is

  with $\beta_I = 0$. XY conditional independence $\Leftrightarrow H_0 : \beta_1 = \cdots = \beta_I$. Large sample chi-squared tests have $df = I - 1$.

7.5.1 Using Multinomial Models to Test Conditional Independence

- *Y nominal*, *X ordinal*. For ordered scores $\{x_i\}$, the model is

  Conditional independence is $H_0 : \beta_1 = \cdots = \beta_{J-1} = 0$. Large sample chi-squared tests have $df = J - 1$.

- *Y nominal*, *X nominal*. Treating X as a factor, the model is

  with $\beta_{Ij} = 0$ for each j. Conditional independence is $H_0 : \beta_{1j} = \cdots = \beta_{Ij}$ for $j = 1, ..., J - 1$. Large sample chi-squared tests have $df = (I - 1)(J - 1)$.

7.5.1 Using Multinomial Models to Test Conditional Independence

Table 7.10 summarizes the four tests.

TABLE 7.10 Summary of Models for Testing Conditional Independence

Y-X	Model	Conditional Independence	df
Ord-Ord	$\text{logit}[P(Y \le j)] = \alpha_j + \beta x_i + \beta_k^Z$	$\beta = 0$	1
-Nom	$\text{logit}[P(Y \le j)] = \alpha_j + \beta_i + \beta_k^Z$	$\beta_1 = \cdots = \beta_I$	$I - 1$
Nom-Ord	$\log\left[\frac{P(Y=j)}{P(Y=J)}\right] = \alpha_{jk} + \beta_j x_i$	$\beta_1 = \cdots = \beta_{J-1} = 0$	$J - 1$
-Nom	$\log\left[\frac{P(Y=j)}{P(Y=J)}\right] = \alpha_{jk} + \beta_{ij}$	all $\beta_{ij} = 0$	$(I-1)(J-1)$

7.5.2 Job satisfaction example revisited

Data: see Table 7.8; scores $\{3, 10, 20, 35\}$.

TABLE 7.11 Summary of Model-Based Likelihood-Ratio Tests of Conditional Independence for Table 7.8

Satisfaction	Income	G^2 Fit	df	Test Statistic	df	P-value
Ordinal	Ordinal	13.95	19	5.7	1	0.017
	Nominal	10.51	17	9.1	3	0.028
	Not in model	19.62	20	—	—	—
Nominal	Ordinal	11.74	15	7.6	3	0.054
	Nominal	7.09	9	12.3	9	0.198
	Not in model	19.37	18	—	—	—

Likelihood-ratio statistic $19.62 - 13.95 = 5.67$ with $df = 20 - 19 = 1$ implies strong evidence of an effect (i.e. reject H_0 conditional independence.) Models that treat either or both variables as nominal do not provide such strong evidence.

7.5.3 Generalized CMH tests for $I \times J \times K$ Tables

Let n_{ijk} be the number of observations in cell (i, j, k) and

$$\mathbf{n}_k = (n_{11k},\ n_{12k},\ \cdots,\ n_{1,J-1,k},\ \cdots,\ n_{I-1,J-1,k})'.$$

Denote $\boldsymbol{\mu}_k = E(\mathbf{n}_k)$ under H_0 : conditional independence, namely

$$\hat{\boldsymbol{\mu}}_k = \left(n_{1+k}n_{+1k},\ n_{1+k}n_{+2k},\ \ldots,\ n_{I-1,+,k}n_{+,J-1,k}\right)'/n_{++k}$$

and let $\mathbf{V}_k$ be the null covariance matrix of $\mathbf{n}_k$, where

$$\mathrm{cov}(n_{ijk}, n_{i'j'k}) = \frac{n_{i+k}(\delta_{ii'}n_{++k} - n_{i'+k})n_{+jk}(\delta_{jj'}n_{++k} - n_{+j'k})}{n_{++k}^2(n_{++k} - 1)}$$

with $\delta_{ab} = 1$ when $a = b$ and $\delta_{ab} = 0$ otherwise. Let

$$\boldsymbol{n} = \sum \boldsymbol{n}_k, \quad \hat{\boldsymbol{\mu}} = \sum \hat{\boldsymbol{\mu}}_k, \quad \boldsymbol{V} = \sum \boldsymbol{V}_k.$$

7.5.3 Generalized CMH tests for $I \times J \times K$ Tables

The generalized CMH statistic for nominal X and Y is

$$\mathrm{CMH} = (\boldsymbol{n} - \hat{\boldsymbol{\mu}})' \boldsymbol{V}^{-1} (\boldsymbol{n} - \hat{\boldsymbol{\mu}}).$$

Its large-sample chi-squared distribution has $df = (I-1)(J-1)$.

For $K = 1$, CMH=$[(n-1)/n]X^2$, where X^2 is the Pearson statistic (3.10).

7.5.3 Generalized Cochran-Mantel-Haenszel tests for $I \times J \times K$ Tables

For ordinal X and Y, set ordered scores $\{u_i\}$ and $\{v_j\}$ and define $T_k = \sum_i \sum_j u_i v_j n_{ijk}$. Evidence of a positive trend occurs if each stratum T_k exceeds its null expectation.

Given the marginal totals in each stratum, under conditional independence

$$E(T_k) = \Big[\sum_i u_i n_{i+k}\Big]\Big[\sum_j v_j n_{+jk}\Big]/n_{++k},$$

$$var(T_k) = \frac{1}{n_{++k}-1}\Big[\sum_i u_i^2 n_{i+k} - \frac{(\sum_i u_i n_{i+k})^2}{n_{++k}}\Big]\Big[\sum_j v_j^2 n_{+jk} - \frac{(\sum_j v_j n_{+jk})^2}{n_{++k}}\Big].$$

It converges to a chi-squared distribution with $df = 1$ under conditional independence as $n \to \infty$. For $K = 1$, this is the M^2 statistic (3.15).

Outline

7.6 Discrete-Choice Multinomial Logit Models

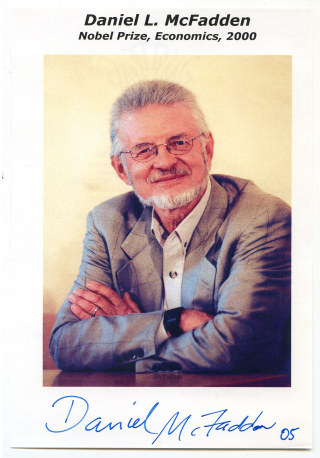
An important application of multinomial logit models is determining effects of explanatory variables on a subject's choice from a discrete set of options.

For example,

- the choice of transportation system to take to work (drive, bus, subway, walk, bicycle);
- the choice of housing (buy house, buy condominium, rent);
- the choice of primary shopping location (downtown, mall, catalogs, Internet).

Models for response variables consisting of a discrete set of choices are called *discrete-choice models* (It is popular in Econometrics).

7.6 Discrete-Choice Multinomial Logit Models



Daniel Little McFadden (born July 29, 1937) is an econometrician who shared the 2000 Nobel Memorial Prize in Economic Sciences with James Heckman for the development of theory and methods for analyzing discrete choice. He was the E. Morris Cox Professor of Economics at the University of California, Berkeley.

7.6.1 Discrete-choice modeling

In many discrete-choice applications, an explanatory variable takes different values for different choices. For example, as predictor of choice of transportation system, cost and time to each destination take different values for each option. Explanatory variables of this type are characteristics of the choices.

These variables differ from the usual ones, for which values remain constant across the choice set. Such variables, called characteristics of the chooser, include income, education, and others.

McFadden (1974) proposed a discrete-choice models for explanatory variables that are characteristics of the choices.

7.6.1 Discrete-choice modeling

For subject i and response choice j, let $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, ..., x_{ijp})'$ and let $\mathbf{x}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{i,q_i})$, where C_i is the choice of set for subject i and $q_i = \#C_i$.

Then, conditional on the choice set C_i for subject i, the probability model for probability of selecting option j is

$$\pi_j(\mathbf{x}_i) = \frac{\exp(\beta' \mathbf{x}_{ij})}{\sum_{h \in C_i} \exp(\beta' \mathbf{x}_{ih})}. \tag{22}$$

For each pair of choices a and b, this model has the logit form

$$\log[\pi_a(\mathbf{x}_i)/\pi_b(\mathbf{x}_i)] = \beta'(\mathbf{x}_{ia} - \mathbf{x}_{ib}).$$

7.6.2 Discrete-choice and multinomial logit models

Model (22) can also incorporate explanatory variables that are characteristics of chooser. This seems surprising, since (22) have a single parameter for each explanatory variable (i.e. only β, not β_j).

However, multinomial logit model (2) has discrete-choice form (22) after replacing such an explanatory variable by J artificial variables. For instance, for a single explanatory variable, let x_i denote its value for subject i.

For $j = 1, 2, ..., J$, let δ_{jk} equal to 1 when $k = j$ and 0 otherwise, and let

$$\mathbf{z}_{ij} = (\delta_{j1}, \cdots, \delta_{jJ}, \delta_{j1}x_i, \cdots, \delta_{jJ}x_i)'.$$

7.6.2 Discrete-choice and multinomial logit models

Let $\boldsymbol{\beta} = (\alpha_1, \cdots, \alpha_J, \beta_1, \cdots, \beta_J)'$. Then $\boldsymbol{\beta}' \mathbf{z}_{ij} = \alpha_j + \beta_j x_i$, and (2) is (with $\alpha_J = \beta_J = 0$ for identifiability)

$$\pi_j(x_i) = \frac{\exp(\alpha_j + \beta_j x_i)}{\exp(\alpha_1 + \beta_1 x_i) + \cdots + \exp(\alpha_J + \beta_J x_i)}$$
$$= \frac{\exp(\boldsymbol{\beta}\mathbf{z}_{ij})}{\exp(\boldsymbol{\beta}\mathbf{z}_{i1}) + \cdots + \exp(\boldsymbol{\beta}\mathbf{z}_{iJ})}$$

This has form (22).

So, discrete choice models can contain characteristic of chooser and the choices. Thus model (22) is very general (*this may be one of the reasons for McFadden to win Nobel prize, 2000*) and model (2) is the special case.

7.6.3 Shopping choice example

McFadden (1974) used multinomial logit models to describe how residents of Pittsburg, Pennsylvania chose a shopping destination.

$Y = 1, 2, 3, 4, 5$ denotes 5 possible destination (city zones);

Two explanatory variables: *P*- price of the trip (time and cost); *S*- shopping opportunity (retail employment in the zone);

The ML estimates of model parameters were -1.06 (SE=0.28) for price of the trip and 0.84 (SE=0.23) for shopping opportunity. So,

$$\log(\hat{\pi}_a/\hat{\pi}_b) = -1.06(P_a - P_b) + 0.84(S_a - S_b).$$

Conclusion: a destination is relatively more attractive as the trip price decreases and as the shopping opportunity increases.

The conclusion is not surprising.

7.6.2 Discrete-choice and multinomial logit models

TABLE A.10 SAS Code for Baseline-Category Logit Models with Alligator Data in Table 7.1

```
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 / param=ref;
 model food(ref='1') = lake size / link=glogit
     aggregate scale=none;
proc catmod; weight count;
 population lake size gender;
 model food = lake size / pred=freq pred=prob;
```

7.6.2 Discrete-choice and multinomial logit models

TABLE A.11 SAS Code for Cumulative Logit and Probit Models with Mental Impairment Data in Table 7.5

```
data impair;
input mental ses life;
datalines;
1 1 1
…
4 0 9
;
proc genmod ;
  model mental = life ses / dist = multinomial link = clogit lrci type3;
proc logistic;
  model mental = life ses / link = probit;
```

7.6.2 Discrete-choice and multinomial logit models

TABLE A.12 SAS Code for Adjacent-Categories Logit and Mean Response Models and CMH Analysis of Job Satisfaction Data in Table 7.8

```
data jobsat;
input gender income satisf count @@;
count2 = count + .01;
datalines;
1 1 1 1  1 1 2 3  1 1 3 11  1 1 4 2
...
0 4 1 0  0 4 2 1  0 4 3  9  0 4 4 6
;
proc catmod order = data; * ML analysis of adj - cat logit (ACL) model;
    weight count;
    population gender income;
    model satisf =
        (1 0 0  3 3,  0 1 0 2 2,  0 0 1 1 1,
         1 0 0  6 3,  0 1 0 4 2,  0 0 1 2 1,
         1 0 0  9 3,  0 1 0 6 2,  0 0 1 3 1,
         1 0 0 12 3,  0 1 0 8 2,  0 0 1 4 1,
         1 0 0  3 0,  0 1 0 2 0,  0 0 1 1 0,
         1 0 0  6 0,  0 1 0 4 0,  0 0 1 2 0,
         1 0 0  9 0,  0 1 0 6 0,  0 0 1 3 0,
         1 0 0 12 0,  0 1 0 8 0,  0 0 1 4 0)
          /ml pred = freq;
proc catmod order = data; weight count2; * WLS analysis of ACL model;
  response alogits; population gender income; direct gender income;
  model satisf = _response_ gender income;
proc catmod; weight count; * mean response model;
  population gender income; response mean; direct gender income;
  model satisf = gender income / covb;
proc freq; weight count;
  tables gender*income*satisf / cmh scores = table;
```