Models for Ordinal Response Data

Robin High
Department of Biostatistics
Center for Public Health
University of Nebraska Medical Center
Omaha, Nebraska

Recommendations

- Analyze numerical data with a statistical model having an appropriate continuous or count distribution
- Work with actual numbers, not recoded into ordered categories
  - Reduces unnecessary measurement error
  - Increases the power to detect a significant effect
- When responses are coded into ordered categories are the data made available or practical to collect, what are possible analysis methods?

Outline of Topics

- Types of Ordinal Models
- SAS® Procedures for Ordinal Response Data
- Example Data and Coding Suggestions
- PROC NLMIXED
- Stereotype Model

Types of Ordinal Response Models

- Fully Constrained
  1. Cumulative Logit (CL)
  2. Adjacent Logit (AL)
  3. Continuation Ratio (CR)
- Partial Proportional Odds
  1. Constrained (Linear Adjustment)
  2. Unconstrained (GLM Codes)
- Stereotype Model
  - Related to Generalized Logit Model (not ordinal)
  - Diagnostics
  - Binary and Adjacent Logits are special cases

A few SAS Procedures for Ordinal Data

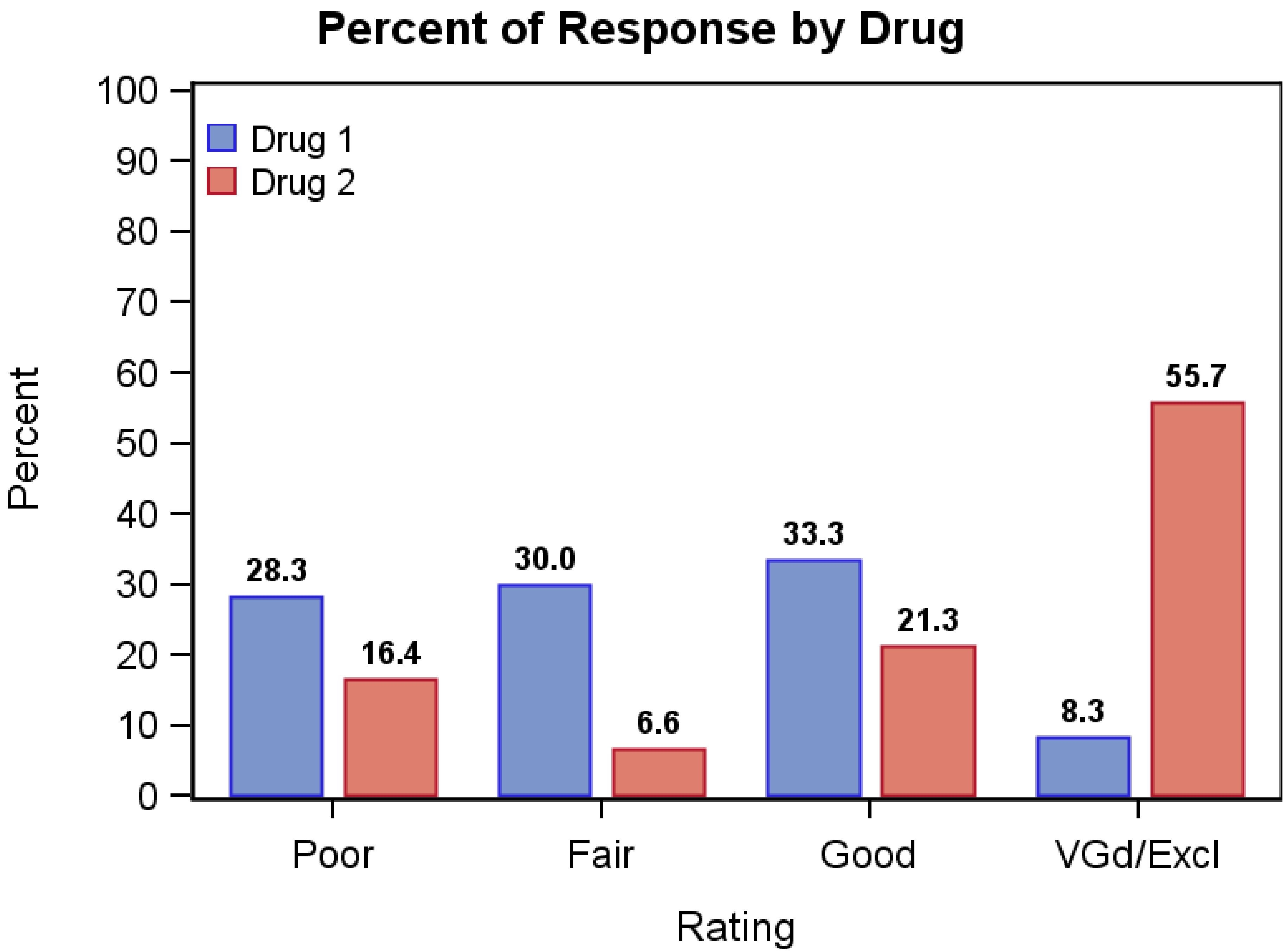
- PROCs LOGISTIC, GENMOD, GLIMMIX
  - Cumulative Logit (Proportional Odds): link=clogit
  - Restructured data set
    - » Continuation Ratio (Allison, "Logistic Regression" 2012)
    - » Unconstrained Partial Proportional Odds (Stokes, et. al., "Categorical Data Analysis," 2000)
- Adjacent Logit
  - PROC CATMOD: Categorical predictors only
  - PROC GENMOD: Loglinear model with added variables
- PROC NLMIXED

Example: Ordinal Response

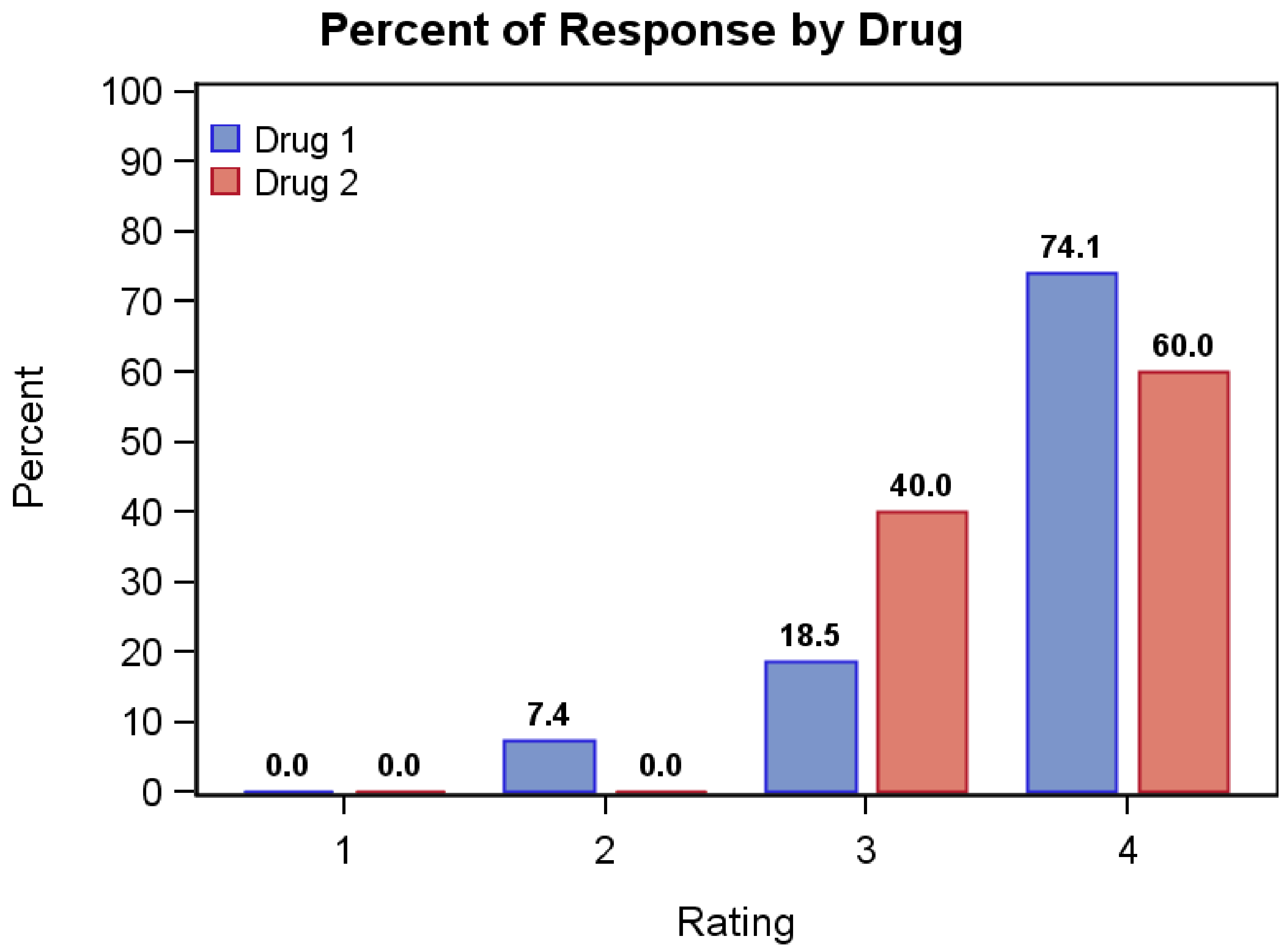
Counts		Rating (y)				
		1	2	3	4	
		Poor	Fair	Good	VrGd/ Excl	All
Drug	x					
1	1	17	18	20	5	60
2	0	10	4	13	34	61

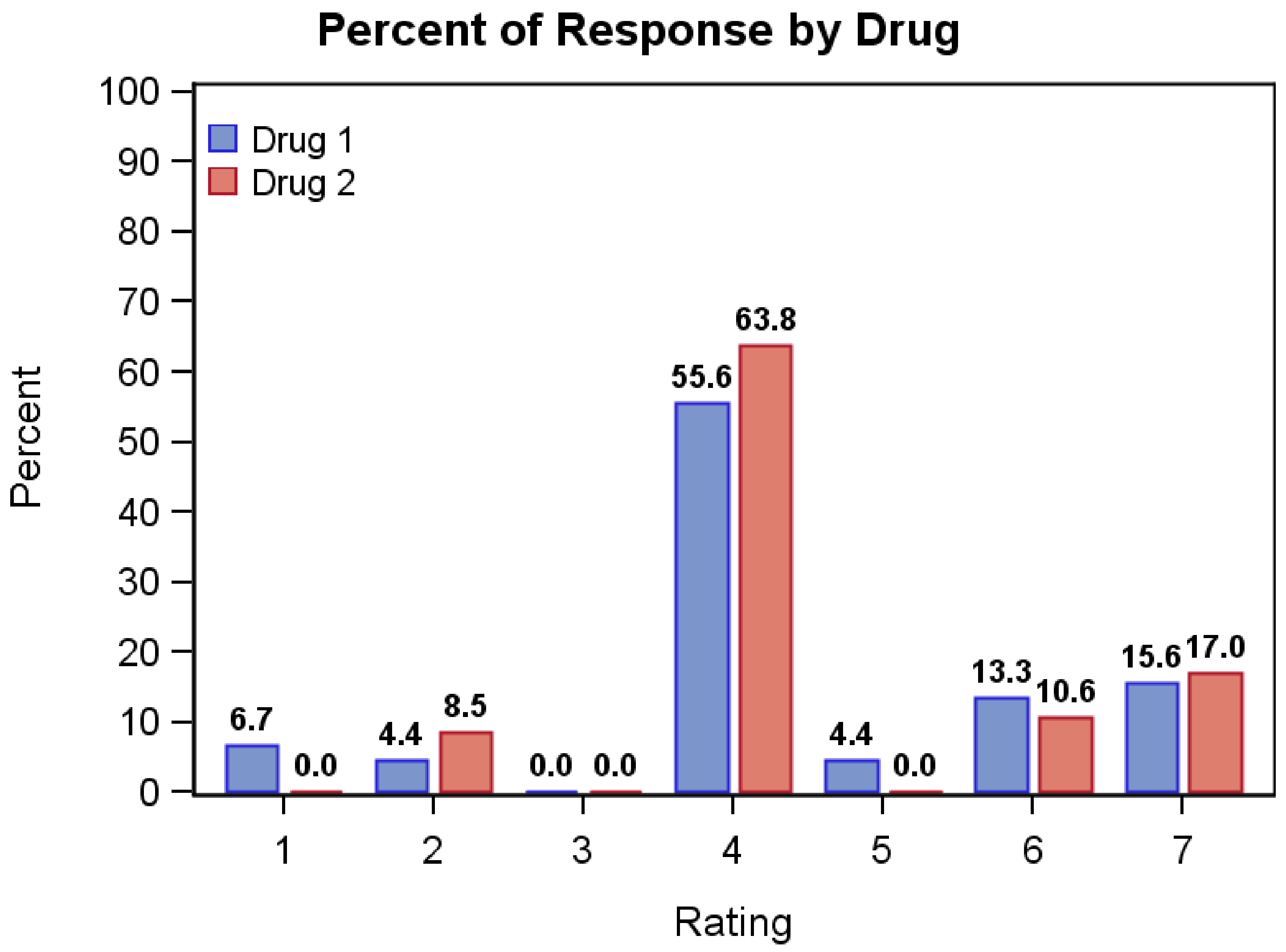
Response: J = 4 categories
Explanatory: Drug is categorical with 2 levels
x: dummy coded representation of Drug

Graphical Representation



Problematic Distributions





Responses at the Edge

Indifference

Response Probabilities for Ordinal Data

Multinomial Distribution with J categories ($J \geq 3$)

Cumulative Logit

Adjacent Logit

Descending order

$p_1 = Pr(y=4)$

$p_2 = Pr(y=3)$

$p_3 = Pr(y=2)$

$p_4 = Pr(y=1)$

Continuation Ratio

Ascending order

$p_1 = Pr(y=1)$

$p_2 = Pr(y=2)$

$p_3 = Pr(y=3)$

$p_4 = Pr(y=4)$

Constraints

$0 < p_i < 1 \quad i = 1,2,\ldots,J$

$\sum p_i = 1$

Ordinal Response Models with J Levels

- Binary comparisons by placing adjacent values of the J responses into mutually exclusive subsets, A and B, defined by J-1 cut-points

- Specify subsets A and B with j as cut-point
  - Cum Logit A: $i > j$ vs B: $i \leq j$ for $j = J-1,..,2,1$
  - Adj Logit A: $i = j+1$ vs B: $i = j$ for $j = J-1,..,2,1$
  - Cont Ratio A: $i = j$ vs B: $i > j$ for $j = 1,2,..,J-1$

Ordinal Response Models

- Define p_A and p_B as subsets of $\{p_1, ..., p_J\}$
- Linear Predictor

  $$LOG(p_A / p_B) = \alpha_j + ß*x \quad \text{for } j = 1 .. J-1$$

- Cumulative and Adjacent Logit Models
  - Define A and B so that a positive/negative estimate of ß implies a positive/negative association with ordered response levels
- Continuation Ratio Model
  - Define A and B so that sign of estimate for ß corresponds to the sign of coefficient for hazard ratio in survival analysis

Binary Data Review

- Code binary responses as y= 1 / 2

- SAS default is lowest coded value as target level:
  that is, compare level 1 vs level 2

- Add “descending” option to compare level 2 vs level 1 (e.g., on the PROC LOGISTIC, GENMOD statement)
  - Positive ß coefficient defines odds ratio > 1

Odds Ratio from 2x2 Table

Counts		y				
		2	1	All	Odds	Odds Ratio
Drug	Code					
1	1	30	20	50	30/20 = 1.50	2.67
2	0	18	32	50	18/32 = 0.56	

Odds Ratio = 2.67 > 1

The odds of y=2 are more likely with Drug 1 than the odds of y=2 with Drug 2

PROCs LOGISTIC and NLMIXED with Binary data

```
DATA testD;
INPUT drug y count;
DATALINES;
  1    2   30
  1    1   20
  2    2   18
  2    1   32
;
PROC LOGISTIC DATA=testD descending;
CLASS drug / param=glm;
MODEL y = drug / expb;
FREQ count;
RUN;
```

PROCs LOGISTIC and NLMIXED with Binary data

```
PROC NLMIXED DATA=testD;
PARMS b0 .1 b1 .1 ;             * 1. initialize estimates;
eta = b0 + b1*(drug=1);         * 2. linear predictor ;
p1 = 1 /(1+exp(-eta));          * 3. probabilities;
p2 = 1-p1;
lk =(p1**(y=2))*(p2**(y=1)); * desc; * 4. Binary likelihood ;
IF (lk > 1e-8) THEN lglk = LOG(lk);
       ELSE lglk=-1e100;
MODEL y ~ general(lglk);
REPLICATE count;
ESTIMATE “Odds Ratio" EXP(b1); * 5. estimate with parameters;
RUN;
```

PROC NLMIXED for Ordinal Response Models

Types of Statements

1. Initial Values of Estimates (PARMS)
2. J-1 Linear Predictors (called eta_j)
3. J-1 Response Probabilities (for $p_1,\dots,p_{J-1}$ | find p_J by subtraction)
4. Maximum Likelihood Estimation (Likelihood eq. / MODEL / REPLICATE)
5. Odds Ratios (ESTIMATE)
6. Predicted Probabilities (PREDICT)

Basic NLMIXED Code for J=4 Response Levels

```
PROC NLMIXED DATA=indat;
PARMS <enter initial values of parameters >;
* Linear predictors, eta1, eta2, eta3;
* probabilities p1, p2, p3 as functions of eta1, eta2, eta3 / p4 by subtraction;
* Multinomial Likelihood Equation;
lk = (p1**(y=4))*(p2**(y=3))*(p3**(y=2))*(p4**(y=1)); * descending response;
* Compute loglikelihood;
IF (lk > 1e-8) then lglk = LOG(lk); else lglk=-1e100; * computational safety ;
* Estimate model;
MODEL y ~ general(lglk);
REPLICATE count;  * for categorical explanatory data entered as cell counts;
ESTIMATE "Drug 1 vs 2" EXP(drg) ; * odds ratios;
ID p1 p2 p3 p4;
PREDICT p1 =prd (keep= y drug count p1 p2 p3 p4) ;
RUN;
```

PROC NLMIXED

Advantages

- Similar code for three ordinal response models
- Work with one data structure (no restructuring required)
- Start with the Cumulative Logit Model
- Modifications follow with type of linear predictors and model
- Linear predictors may have both categorical and continuous data
- Extra insight into the analysis

PROC NLMIXED

Disadvantages

- Write out linear predictors for complex models, esp. with categorical data

- Can be tedious to discover coding errors which are often reduced with short variable names

- Produce your own model diagnostics and statistical graphs

1. Cumulative Logit for J=4

- Descending: P(y=4) = p_1 ... P(y=1) = p_4

- Subsets A and B

  A: $i > j$ vs B: $i \leq j$ for $j = 3,2,1$

- With J=4, make 3 comparisons

  j=3 A: 4 vs B: 1,2,3

  j=2 A: 3,4 vs B: 1,2

  j=1 A: 2,3,4 vs B: 1

- Comparisons made across all 4 levels of response

Cumulative Logit

With J=4 have three Linear Predictors

```
eta1 = a1 + drg*(drug=1);
eta2 = a2 + drg*(drug=1);
eta3 = a3 + drg*(drug=1);
```

Model cumulative response probabilities as functions of these three linear predictors

```
cp1= 1 / (1 + EXP(-eta1));
cp2= 1 / (1 + EXP(-eta2));
cp3= 1 / (1 + EXP(-eta3));
```

Cumulative Logit Response Probabilities

Get individual probabilities by subtraction

```
p1 = cp1;
p2 = cp2 - cp1;
p3 = cp3 - cp2;
p4 =  1 - cp3;
```

[recall that p_1 = Pr(y=4)]

Important to set initial values in NLMIXED PARMS statement

a1 < a2 < a3

```
PARMS a1 -1 a2 0 a3 1;
```

Cumulative Logit with PROC LOGISTIC

3 Linear Predictors

4 vs 123: 0.07 + (-1.77)*(drug=1)

34 vs 12: 1.47 + (-1.77)*(drug=1)

234 vs 1: 2.42 + (-1.77)*(drug=1)

Cumulative Odds Ratio = EXP(-1.77) = 0.170

for all 3 comparisons

Example Binary Comparison

j=3		Rating (y)				
		4	1,2,3	Tot	Odds	Odds Ratio
Drug						
1	1	5	55	60	5/55 = 0.091	0.072
2	0	34	27	61	34/27 = 1.259	

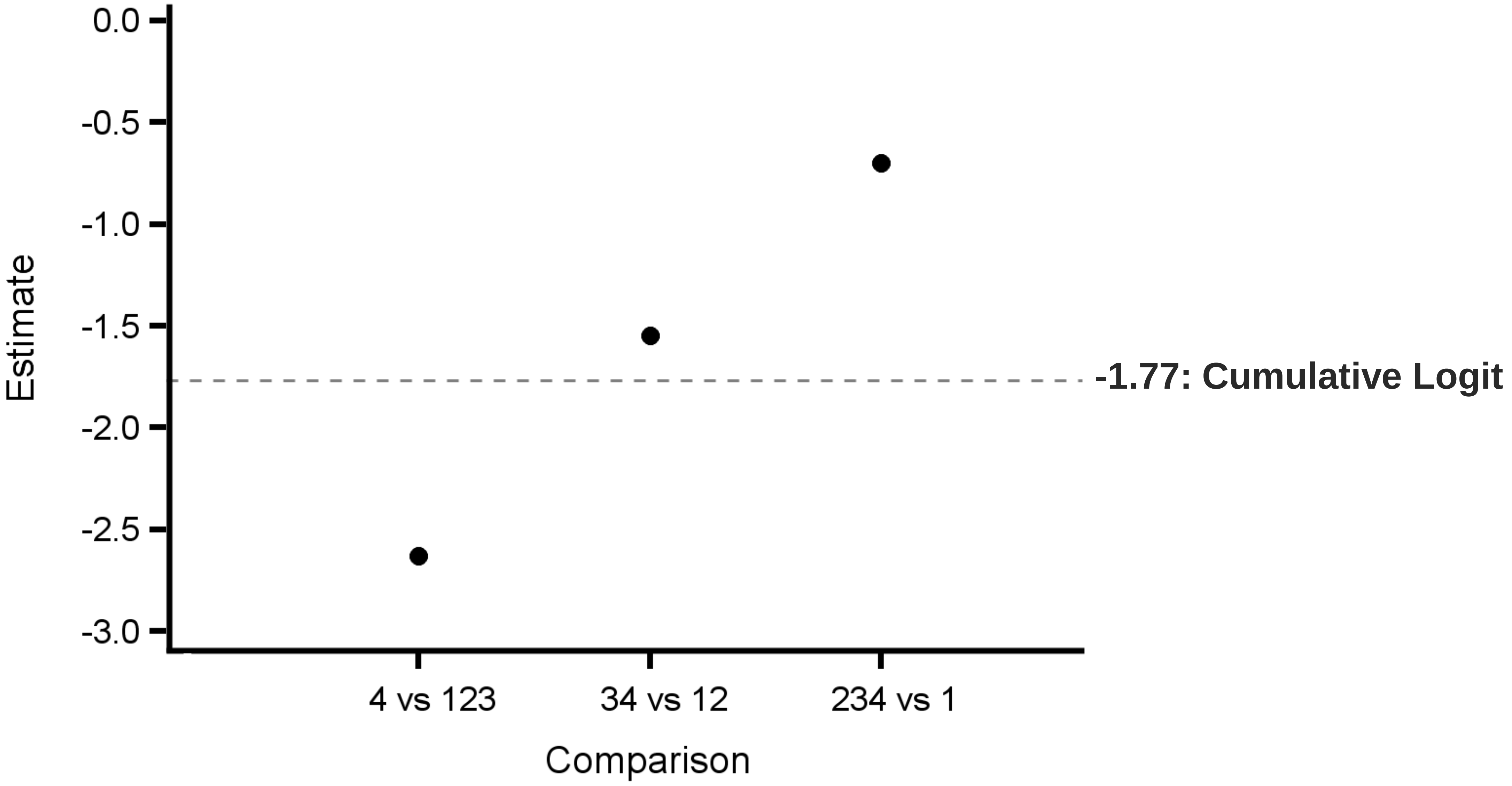
Summary of Binary Comparisons

Separate estimates and odds ratios for aggregated 2x2 tables

2x2 Table	Estimate	Odds Ratio
1: 4 vs 123	-2.63	0.072
2: 34 vs 12	-1.55	0.213
3: 234 vs 1	-0.70	0.496

Compare to -1.77 found with cumulative logit model

Plot Estimate by Comparison



Partial Proportional Odds (Constrained)

Linear adjustment: need 4 or more response levels (J ≥ 4)

Modify 3 linear predictors with a linear adjustment

```
eta1 = a1 + (drg + 0*d1 )*(drug=1);
eta2 = a2 + (drg + 1*d1 )*(drug=1);
eta3 = a3 + (drg + 2*d1 )*(drug=1);
```

All other NLMIXED statements are the same

Partial Proportional Odds
Constrained Linear Adjustment

Compare	Estimate	PPOM Odds Ratio	2x2 Odds Ratio
4 v 123	-2.53	0.079	0.072
34 v 12	-1.61	0.200	0.213
234 v 1	-0.69	0.502	0.496

Partial Proportional Odds (Unconstrained)

Suppose linear adjustment is not reasonable

Unconstrained adjustments, d1b and d1c, to modify estimate of drg

```
eta1 = a1 + (drg           )*(drug=1);
eta2 = a2 + (drg + d1b     )*(drug=1);
eta3 = a3 + (drg      + d1c)*(drug=1);
```

With unconstrained model, the number of new estimates is J-2

Partial Proportional Odds (Unconstrained)

Linear Predictors with two independent variables

```
gender = F=female / M=male
drug   = 1/2

eta1 = a1 + gnd*(gender='F') + (drg             )*(drug=1);
eta2 = a2 + gnd*(gender='F') + (drg + d1b       )*(drug=1);
eta3 = a3 + gnd*(gender='F') + (drg       + d1c)*(drug=1);
```

gnd retains proportional odds

drg does not retain proportional odds

Example: PROC LOGISTIC (V. 12.1) with unequalslopes; see Chapter 9, Stokes, et. al., "Categorical Data Analysis," 3rd Ed. (2012)

2. Adjacent Logits with J=4

- Descending: P(y=4) = p_1 ... P(y=1) = p_4
- Subsets A and B

  A: i = j+1 vs B: i = j for j = 3,2,1

- Compare adjacent response categories

  j=3 A: 4 vs B: 3

  j=2 A: 3 vs B: 2

  j=1 A: 2 vs B: 1

Compare Adjacent Responses (descending)

j=3		y				
		4	3	All	Odds	Odds Ratio
Drug	Code					
1	1	5	20	25	5/20= 0.25	0.092
2	0	34	13	34	34/13= 2.62	

Adjacent Logit

J=4, 3 Linear Predictors (same as cumulative logit)

```
eta1 = a1 + drg*(drug=1);
eta2 = a2 + drg*(drug=1);
eta3 = a3 + drg*(drug=1);
```

Response probabilities $\{p_1, p_2, p_3, p_4\}$ are a different function of the linear predictors

Adjacent Logit Response Probabilities

```
total = 1 + EXP(eta3) + EXP(eta2+eta3) + EXP(eta1+eta2+eta3);
```

Individual response probabilities

```
p4 = 1 / total;
p3 = EXP(eta3)*p4;
p2 = EXP(eta2)*p3;
p1 = EXP(eta1)*p2;
```

See appendix from paper from SGF 2013 proceedings:

paper 445-2013

Adjacent Logits

Estimate = -0.809

Odds Ratio = 0.445   (local odds ratio)

Compare with the cumulative logit model

Estimate = -1.77

Odds Ratio = 0.170 (cumulative odds ratio)

3. Continuation Ratio

- Ordered response levels indicate progression through sequential stages: 1, 2, 3, 4, ..
- Every subject begins at level 1
- No skipped levels
- Forward trend is not reversed

- Examples
  - Categorical times to event
  - Level of skill achieved

Continuation Ratio for J=4

- Ascending Order: $P(y=1) = p_1 \ldots P(y=4) = p_4$
- Define A and B

  A: i = j vs B: i > j for j = 1,2,3

- For J=4, compare levels

  j=1 A: 1 vs B: 2,3,4
  j=2 A: 2 vs B: 3,4
  j=3 A: 3 vs B: 4

- Estimate conditional probability of the response p_A given the specified response level or higher, p_B

Restructure Data

Counts	Poor	Fair	Good	Excl
Drug				
1	17	18	20	5
2	10	4	13	34

Rating	y2					
	0	1	All	0	1	
Stage	N	N	N	Row %	Row %	
1 vs 234						
1	17	43	60	28.3	71.7	j=1 first comparison
2	10	51	61	16.4	83.6	
2 vs 34						
1	18	25	43	41.9	58.1	j=2 second comparison
2	4	47	51	7.8	92.2	

Multinomial Distribution

Row counts distributed among J categories (y_1, y_2, …, y_J)

For J=4

$$N = y_1 + y_2 + y_3 + y_4$$

with probabilities $\{p_1, p_2, p_3, p_4\}$

pdf

$Pr(Y_1=y_1, Y_2=y_2, Y_3=y_3, Y_4=y_4)$

$$= \frac{N!}{y_1!\, y_2!\, y_3!\, y_4!} (p_1^{y_1})*(p_2^{y_2})*(p_3^{y_3})*(p_4^{y_4})$$

Multinomial PDF factored into a product of binomials

$$pdf = Bin(N, y_1, \pi_1) * Bin(N-y_1, y_2, \pi_2) * Bin(N-y_1-y_2, y_3, \pi_3)$$

Where the binomial probabilities are functions of response probabilities:

$$\pi_1 = p_1$$

$$\pi_2 = p_2/(1-p_1)$$

$$\pi_3 = p_3/(1-p_1-p_2)$$

$$p_1 = \pi_1$$

$$p_2 = \pi_2 * (1 - p_1)$$

$$p_3 = \pi_3 * (1 - p_1 - p_2)$$

$$p_4 = 1 - (p_1 + p_2 + p_3)$$

Multinomial PDF factored into a product of binomials

Calculations are outlined:

Alan Agresti, (1984) “Analysis of Ordinal Categorical Data” 1st ed., Chapter 6, problem 3, p. 118.

NLMIXED: Continuation Ratio

J=4, have 3 Linear Predictors

```
eta1 = a1 + drg*(drug=1);
eta2 = a2 + drg*(drug=1);
eta3 = a3 + drg*(drug=1);
```

NLMIXED probability estimates

```
p1 =  1 / (1 + exp(-eta1));
p2 = (1 / (1 + exp(-eta2))) * (1 - p1);
p3 = (1 / (1 + exp(-eta3))) * (1 - p1 - p2);
p4 =  1 - (p1 + p2 + p3);
```

Continuation Ratio

- p_A / p_B is closely related to the hazard ratio in discrete-time survival analysis

- Comparisons of the response levels could also be made in descending order
  - 4 vs 1,2,3 / 3 vs 1,2 / 2 vs 1
  - Odds ratio not the inverse of the ascending version
  - I would not do this if natural order is 1->2->3->4

Stereotype Model

First consider the General Baseline Logit Model

Four response levels (J=4)

Three 2-level explanatory variables $\{x_1, x_2, x_3\}$ coded 0/1

$$\text{Log}(p_1 / p_4) = \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3$$

$$\text{Log}(p_2 / p_4) = \alpha_2 + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3$$

$$\text{Log}(p_3 / p_4) = \alpha_3 + \beta_{31}x_1 + \beta_{32}x_2 + \beta_{33}x_3$$

Estimate 12 parameters

Stereotype Model

Modify computations of parameters

$$\text{Log}(p_1 / p_4) = \alpha_1 + \varphi_1 * (ß_1 x_1 + ß_2 x_2 + ß_3 x_3)$$

$$\text{Log}(p_2 / p_4) = \alpha_2 + \varphi_2 * (ß_1 x_1 + ß_2 x_2 + ß_3 x_3)$$

$$\text{Log}(p_3 / p_4) = \alpha_3 + \varphi_3 * (ß_1 x_1 + ß_2 x_2 + ß_3 x_3)$$

$$\text{Log}(p_4 / p_4) = 0 + \varphi_4 * (ß_1 x_1 + ß_2 x_2 + ß_3 x_3)$$

For identifiability fix $\varphi_1 = 1$ and $\varphi_4 = 0$

Estimate 8 parameters

Stereotype: NLMIXED Linear Predictors

x1, x2, x3 all dummy coded (0/1) in DATA step

```
eta1 = a1 +    1*(b1*x1 + b2*x2 + b3*x3);
eta2 = a2 + phi2*(b1*x1 + b2*x2 + b3*x3);
eta3 = a3 + phi3*(b1*x1 + b2*x2 + b3*x3);
eta4 =  0 +    0*(b1*x1 + b2*x2 + b3*x3);
```

eta4 == 0

entering it may help to see features of model

Parameters to Estimate

J = number of response levels

p = number of explanatory variables

(number of regression coefficients)

Cumulative Logit: (J-1) + p

Stereotype Model: 2*J – 3 + p

Generalized Logit: (J–1) + (J–1)* p

Number of Parameters to Estimate

Resp Levels	No. of Explanatory Vars, (# regr coefs)											
	1			2			3			4		
	CL	ST	GL	CL	ST	GL	CL	ST	GL	CL	ST	GL
3	3	4	4	4	5	6	5	6	8	6	7	10
4	4	6	6	5	7	9	6	8	12	7	9	15
5	5	8	8	6	9	12	7	10	16	8	11	20
6	6	10	10	7	11	15	8	12	20	9	13	25

Stereotype: Probabilities for J=4 Responses

```
total = exp(eta1) + exp(eta2) + exp(eta3) + exp(eta4);
      = exp(eta1) + exp(eta2) + exp(eta3) + 1;


p1 = exp(eta1) / total;
p2 = exp(eta2) / total;
p3 = exp(eta3) / total;
p4 = exp(eta4) / total;   [ also, p4 = 1 / total ; ]
```

Distinguishability

- The J values of φ provide a measure of the distinguishability of the response categories
- If two or more adjacent values of φ are similar, evidence suggests that these categories are indistinguishable
- Tested in NLMIXED with ESTIMATE statements

Example with J=6 Ordinal Responses

Counts	y						
	1	2	3	4	5	6	All
Level 1	28	32	16	29	21	23	149
Level 2	6	8	6	25	13	18	76

Stereotype Model: Linear Predictors

```
eta1 = a1 +    1*drg*(level=1);
eta2 = a2 + phi2*drg*(level=1);
eta3 = a3 + phi3*drg*(level=1);
eta4 = a4 + phi4*drg*(level=1);
eta5 = a5 + phi5*drg*(level=1);
eta6 =  0 +    0*drg*(level=1);
```

Stereotype: Estimate Statements

```
ESTIMATE 'Phi2=1' 1-phi2 df = 200;
ESTIMATE 'Phi3=1' 1-phi3 df = 200;

ESTIMATE 'Phi4=0' phi4   df = 200;
ESTIMATE 'Phi5=0' phi5   df = 200;
```

Test Significance of phi values

Label	Estimate	Standard Error	Probt
phi2=1	0.181	0.343	0.60
phi3=1	-0.075	0.337	0.82
phi4=0	0.432	0.432	0.32
phi5=0	0.119	0.433	0.78

Example: Six Responses Levels aggregated into a 2x2 table

Counts	y: 4,5,6	y: 1,2,3	
	1	0	All
Level 1	73	76	149
Level 2	56	20	76

Could run binomial regression with PROC LOGISTIC

Stereotype Model: Binary Logit

Modify Linear Predictors: enter 0s and 1s for φ

```
eta1 = a1 + 1*b1*(level=1);
eta2 = a2 + 1*b1*(level=1);
eta3 = a3 + 1*b1*(level=1);
eta4 = a4 + 0*b1*(level=1);
eta5 = a5 + 0*b1*(level=1);
eta6 =  0 + 0*b1*(level=1);
```

Stereotype Model: Adjacent Logit

Linear Predictors: enter equally spaced integer φ

```
eta1 = a1 + 3*drg*(drug=1);
eta2 = a2 + 2*drg*(drug=1);
eta3 = a3 + 1*drg*(drug=1);
eta4 =  0 + 0*drg*(drug=1);
```

Issues with StereoType

- Stereotype model is not inherently ordinal

- For 1 explanatory variable (categorical or numeric), results are the same as generalized logit model

- For ordinal data analysis, computed Φ coefficients need to have a decreasing order, bounded between 1 and 0:

$$\varphi_1 = 1 > \varphi_2 > \varphi_3 > \varphi_4 = 0$$

- How to interpret if $\varphi_3 > \varphi_2$?

Concluding Comments

- Types of ordinal response models estimated with NLMIXED
  - Cumulative Logit: Write statements to match results with models estimated from PROC LOGISTIC
  - Adjacent Logit: NLMIXED provides computations for both numerical and categorical explanatory data
  - Continuation Ratio: Compare logit link results with cloglog link and also PROC PHREG (ties=exact on MODEL statement)
  - With one input data set, the only difference in NLMIXED statements are the probability calculations
  - Include Partial Proportional Odds

    (Constrained or Unconstrained)

Concluding Comments

- Stereotype
  - Examine ordinal model response categories for distinguishability
  - Adjacent Logit and Binary Logit are special cases

- Chapters 3 & 4 of “Analysis of Ordinal Categorical Data” by Alan Agresti (Wiley, 2010)
- Chapter 6 of “Logistic Regression Using SAS” by Paul Allison