Predicting an Ordinal Outcome: Options and Assumptions

Mark Lunt
ARC Epidemiology Unit
University of Manchester

June 12, 2001

Context

- Subjects with rheumatoid arthritis develop damage to the joints.
- We wish to predict the severity of damage.
- Damage scale:

  1. No damage
  2. Joint space narrowing
  3. Slight evidence of erosion
  4. Clear evidence of erosion
  5. Worse than 4, not as bad as 6
  6. No further damage to joint possible

What do you do with ordinal data ?

1. Dichotomise: use logistic regression

2. Pretend there is an interval scale: use linear regression

3. Ignore the ordering: fit a multinomial model

4. Use methods specifically for ordinal data

Types of Ordinal Data

- Grouped Continuous
  - There is a underlying continuous variable.
  - Not measured exactly, only to certain fixed ranges.
  - E.g. Age 15-24, 25-34, 35-50 etc.
- Assessed
  - Subjective judgement made by an individual.
  - E.g. strongly disagree, disagree, neither agree nor disagree, agree, strongly agree.
  - May or may not be an underlying continuous latent variable.
  - Erosions outcome is of this type.

Ordinal Regression Models

- Generalized Linear Models
  1. The Cumulative Odds Model
  2. The Continuation Ratio Model
  3. Ordered Probit Model
     - Almost identical to the Cumulative Odds Model

- The Stereotype Model
  - Non-linear form of constrained multinomial model

Generalized Linear Models

Model the Cumulative Response Probability γ_j

$$\gamma_j(\mathbf{x}) = \mathbf{pr}(\mathbf{Y} \leq \mathbf{j}|\mathbf{x})$$

$$\eta(\gamma_j) = \theta_j + \beta\mathbf{x}$$

$\eta =$ logit	$\Rightarrow$ Cumulative Odds
$\eta =$ complementary log log	$\Rightarrow$ Continuation Ratio
$\eta =$ probit	$\Rightarrow$ Ordered Probit

All assume that, on some scale, the effect of $\mathbf{x}$ is the same for all levels of Y.

The Cumulative Odds Model

		y		
		1	2	3
x	1	a	b	c
	2	d	e	f

Odds Ratio for being in a higher category if $x = 2$ rather than $x = 1$

$$\theta_1 = \frac{a(e+f)}{d(b+c)}$$

$$\theta_2 = \frac{(a+b)f}{(d+e)c}$$

Assume θ_1 and θ_2 are both estimates of the same population parameter.

Should test that $\theta_1 \approx \theta_2$

Comments

- Motivation: Grouped continuous data
  - Changing groupings does not affect the population parameter being estimated.
- Reversal invariant.
- Stata commands
  - `ologit resp preds`
  - `omodel logit resp preds`

The Continuation Ratio Model

		y		
		1	2	3
x	1	a	b	c
	2	d	e	f

Odds Ratio for category $\geq \; j + 1$ given category $\geq j$ if $x = 2$ rather than $x = 1$

$$\theta_1 = \frac{a(e+f)}{d(b+c)}$$

$$\theta_2 = \frac{bf}{ec}$$

Assume θ_1 and θ_2 are both estimates of the same population parameter.

Should test that $\theta_1 \approx \theta_2$

Comments

- Not reversal invariant.
- Not collapsing invariant.
- Subtables are independent: easy model to fit
- Stata commands
  - `ocratio resp preds`
  - I have not found a test of proportionality of hazards.

The Stereotype Model

- The full multinomial model can be thought of a series of independent logistic regressions
  - category 2 vs category 1
  - category 3 vs category 1
- If we assume that the regression function is the same for all categories, we have a stereotype model.
- Stereotype model has fewer parameters than multinomial, but is nested within it.

Multinomial Model

- Full multinomial model is

$$\text{pr}(y_i = s | x_{i1} \ldots x_{ip}) = \frac{\exp\left(\beta_{0s} + \sum_{j=1}^{p} x_{ij}\beta_{js}\right)}{\sum_{t=1}^{k} \exp\left(\beta_{0t} + \sum_{j=1}^{p} x_{ij}\beta_{jt}\right)}$$

- This is not identified: commonly fix $\beta_{j1} = 0, \quad$ for$j = 0$ to p.

- This compares all groups with group 1.

Stereotype Model

- The stereotype model assumes that for all groups, $\beta_{js} = \phi_s \beta_j$, i.e.

$$\text{pr}(y_i = s | x_{i1} \ldots x_{ip}) = \frac{\exp\left(\beta_{0s} + \phi_s \sum_{j=1}^{p} x_{ij}\beta_j\right)}{\sum_{t=1}^{k} \exp\left(\beta_{0t} + \phi_t \sum_{j=1}^{p} x_{ij}\beta_j\right)}$$

$$\begin{aligned} \beta_{\mathrm{j}} &= \text{Logistic Regression Function} \\ \phi_{\mathrm{s}} &= \text{Distance apart of groups} \end{aligned}$$

- Commonly ϕ_1 is fixed at 0 and ϕ_k fixed at 1.

Distinguishability & Dimensionality

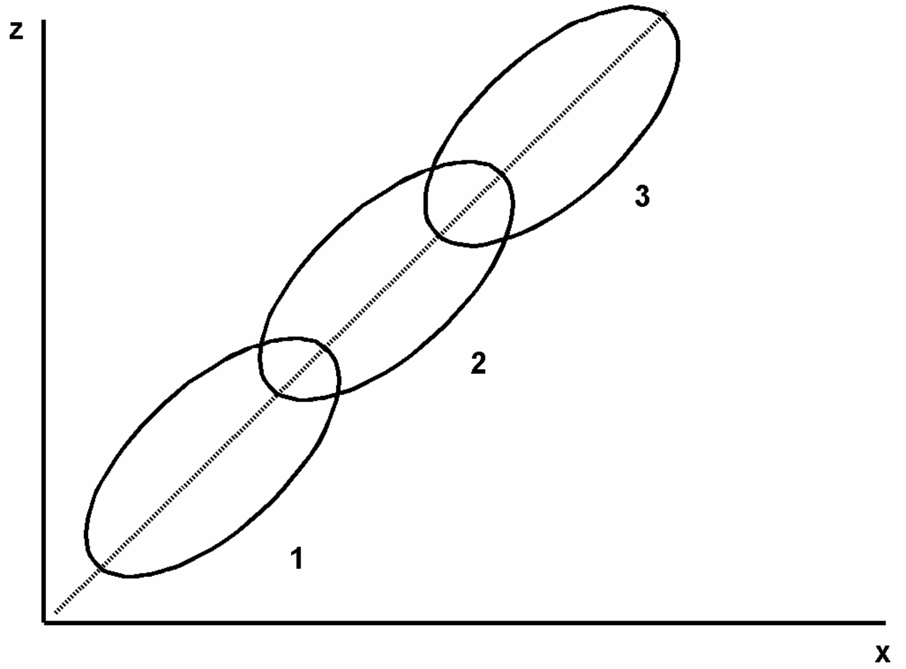
Distinguishability

- If $\phi_i = \phi_j$, then x does not distinguish between groups i and j.
- Can test constrained model with $\phi_i = \phi_j$ for adequacy of fit.

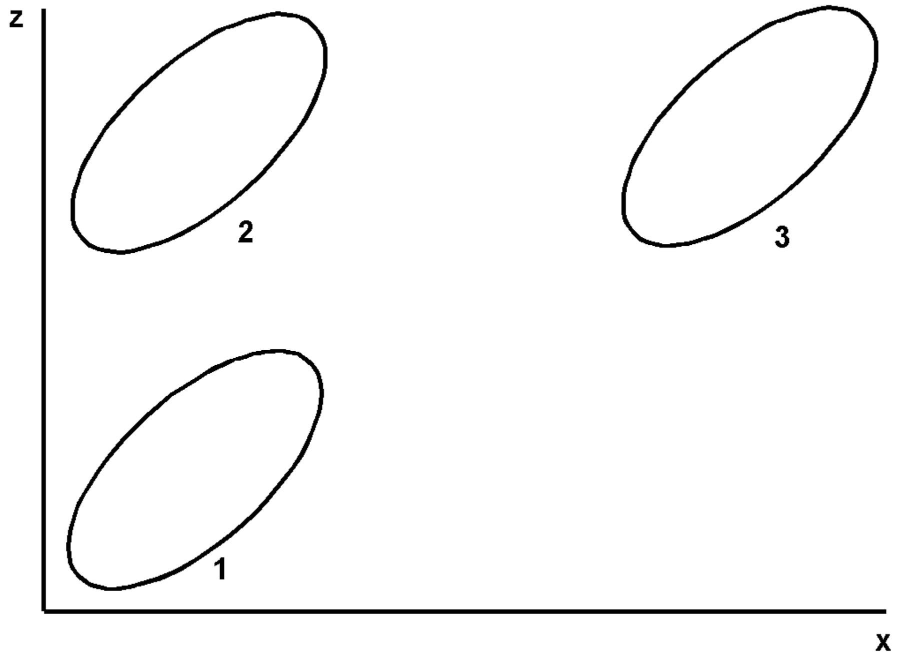
Dimensionality

- If one function of $\mathbf{x}$ discriminates between all groups, relationship is one-dimensional (i.e. ordinal).
- If more than one function is required (i.e. different variables differentiate between different levels) relationship is multi-dimensional.
- In multidimensional models, outcome categories are not strictly ordered with respect to predictors/

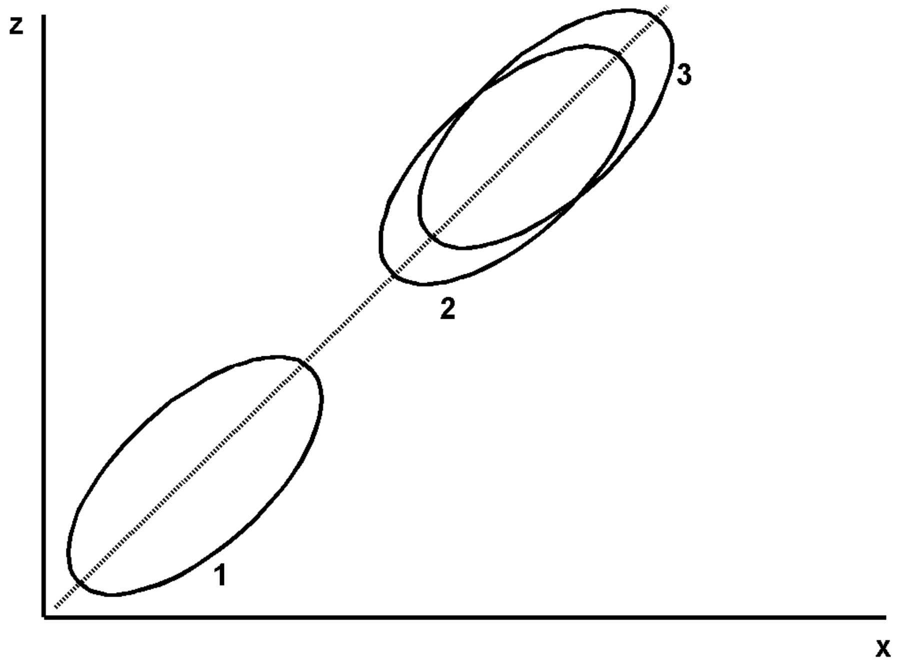
Ordinal relationship



2 dimensional relationship



Indistinguishable categories



Stereotype Regression Strategy

- Determine dimensionality
- Constrain parameters where possible
  - Decide which variables belong to which dimensions if there are more than one
  - Collapse indistinguishable groups together

Model

- Outcome
  - Severity of the most eroded joint (1 - 6)
- Predictors
  - Age (measured in decades,   15 - 85)
  - Rheumatoid factor (present or absent)
  - Shared epitope (0, 1 or 2 copies)

Cumulative Odds Model

```
. omodel logit  erosion age rf epitope
Ordered logit estimates                                 Number of obs   =        251
                                                        LR chi2(3)      =      62.04
                                                        Prob > chi2     =     0.0000
Log likelihood = -385.29669                             Pseudo R2       =     0.0745
------------------------------------------------------------------------------
 erosion |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .3986464   .0879168      4.534   0.000       .2263326    .5709603
      rf |   1.309473   .2594066      5.048   0.000       .8010456    1.817901
 epitope |   .4779209   .1671797      2.859   0.004       .1502548     .805587
---------+--------------------------------------------------------------------
Approximate likelihood-ratio test of proportionality of odds
across response categories:
         chi2(12) =     41.71
       Prob > chi2 =    0.0000
```

Continuation Ratio

```
. ocratio erosion age rf epitope

Continuation-ratio logit Estimates                      Number of obs =      687
                                                        chi2(3)       =    45.26
                                                        Prob > chi2   =   0.0000
Log Likelihood = -393.6903                              Pseudo R2     =   0.0544

------------------------------------------------------------------------------
 erosion |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1749616   .0646592      2.706   0.007       .0482319    .3016914
      rf |   .9110031   .1926102      4.730   0.000       .5334939    1.288512
 epitope |    .392571   .1267941      3.096   0.002        .144059    .6410829
------------------------------------------------------------------------------
Omnibus Test of Proportional Hazards                  LR Chi2(12)   =      53.67
                                                      Prob > chi2   =     0.0000
```

Stereotype Regression: 1 Dimensional

```
. soreg erosion age rf epitope
Stereotype Logistic Regression                    Number of obs =       251
Comparison to null model                          LR Chi2(7)    =     66.67
                                                  Prob > chi2   =    0.0000
Comparison to full model                          LR Chi2(8)    =     33.56
                                                  Prob > chi2   =    0.0000
------------------------------------------------------------------------------
         |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   phi11 |  (dropped)
   phi21 |   .5441926   .1649698      3.299   0.001       .2208578    .8675274
   phi31 |    .889524   .2373843      3.747   0.000       .4242595    1.354789
   phi41 |   .9523727   .2647397      3.597   0.000       .4334925    1.471253
   phi51 |   .8984605   .2527707      3.554   0.000        .403039    1.393882
   phi61 |          1          .          .       .              .           .
  beta11 |    .832095   .2479933      3.355   0.001       .3460371    1.318153
  beta21 |   1.864144   .6729833      2.770   0.006        .545121    3.183167
  beta31 |   .7770334   .3536477      2.197   0.028       .0838967     1.47017
------------------------------------------------------------------------------
beta1 = age
beta2 = rf
beta3 = epitope
```

Stereotype Regression: 2 Dimensional

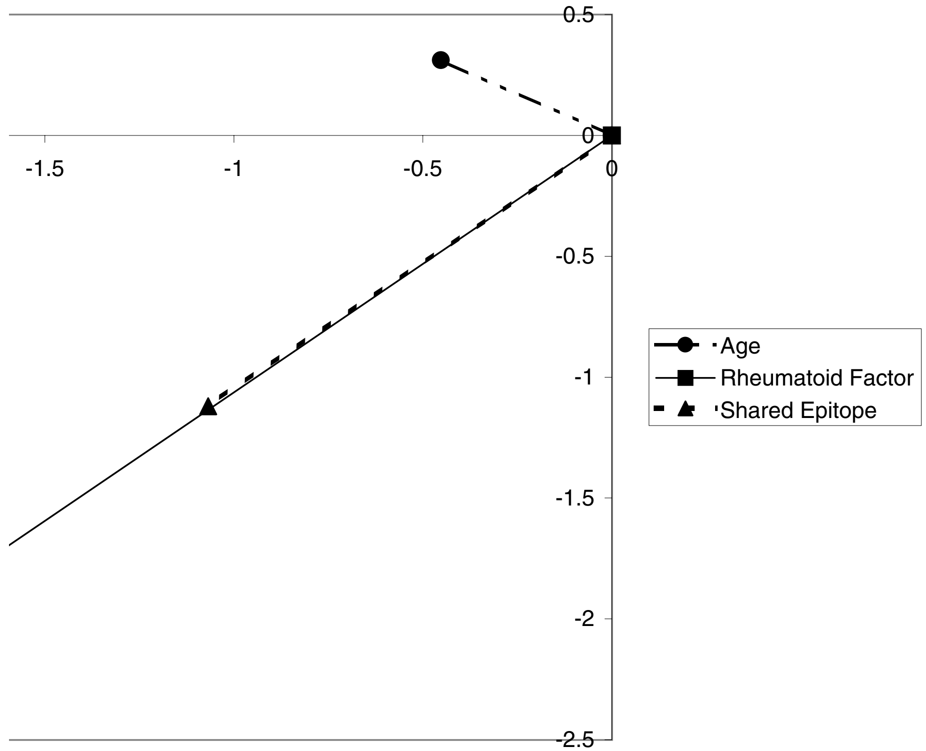
```
. soreg erosion age rf epitope, maxdim(2)

Stereotype Logistic Regression                    Number of obs =        251
Comparison to null model                          LR Chi2(12)   =      97.42
                                                  Prob > chi2   =     0.0000
Comparison to full model                          LR Chi2(3)    =       2.81
                                                  Prob > chi2   =     0.4217


. soreg erosion age rf epitope, maxdim(2) c(1/14)

Stereotype Logistic Regression                    Number of obs =        251
Comparison to null model                          LR Chi2(4)    =      88.34
                                                  Prob > chi2   =     0.0000
Comparison to full model                          LR Chi2(11)   =      11.89
                                                  Prob > chi2   =     0.3719
```

Determining Dimensions



Applying Constraints (1)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
phi11	(dropped)					
phi21	1.692431	1.098211	1.541	0.123	-.4600231	3.84488
phi31	2.272531	1.469143	1.547	0.122	-.6069365	5.151999
phi41	1.517726	.9933333	1.528	0.127	-.4291719	3.464623
phi51	1.273904	.8419309	1.513	0.130	-.3762501	2.924058
phi61	1	.	.	.	.	.
beta11	.4506742	.3014424	1.495	0.135	-.140142	1.04149

phi11	(dropped)					
phi21	1	.	.	.	.	.
phi31	1	.	.	.	.	.
phi41	1	.	.	.	.	.
phi51	1	.	.	.	.	.
phi61	1	.	.	.	.	.
beta11	.7684656	.1288242	5.965	0.000	.5159748	1.020956

beta1 = age

Applying Constraints (2)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
phi12	(dropped)					
phi22	-.0372127	.1810902	-0.205	0.837	-.392143	.3177176
phi32	.3594271	.1656013	2.170	0.030	.0348545	.6839997
phi42	.7134761	.236255	3.020	0.003	.2504249	1.176527
phi52	.7098553	.2355299	3.014	0.003	.2482251	1.171485
phi62	1	.	.	.	.	.
beta22	2.169331	.7273564	2.982	0.003	.7437389	3.594924
beta32	1.120352	.4247061	2.638	0.008	.287943	1.95276
phi12	(dropped)					
phi22	(dropped)					
phi32	.4986462	.1550644	3.216	0.001	.1947256	.8025669
phi42	1	.	.	.	.	.
phi52	1	.	.	.	.	.
phi62	1	.	.	.	.	.
beta22	1.683544	.3560441	4.728	0.000	.9857109	2.381378
beta32	.8272222	.2424987	3.411	0.001	.3519335	1.302511

beta2 = rf
beta3 = epitope

Stereotype Regression: Interpretation(1)

First dimension

- $\boldsymbol{\beta_{age} = 0.77}$
- $\boldsymbol{\phi_1 = 0}$, $\boldsymbol{\phi_2 = \phi_3 = \phi_4 = \phi_5 = \phi_6 = 1}$
- Odds of having some slight damage rather than none increases by $\boldsymbol{e^{0.77}}$ per decade.
- Age does not help to predict how severe the damage is, only that it exists.

Stereotype Regression: Interpretation(2)

Second dimension

- $\boldsymbol{\beta_{rf} = 1.68}$, $\boldsymbol{\beta_{epitope} = 0.83}$
- $\boldsymbol{\phi_1 = \phi_2 = 0}$, $\boldsymbol{\phi_3 = 0.50}$, $\boldsymbol{\phi_4 = \phi_5 = \phi_6 = 1}$
- Odds of being in group 4, 5 or 6 rather than group 1 or 2 is greater by $\boldsymbol{e^{1.68}}$ in the RF+.
- Odds of being in group 3 rather than group 1 or 2 is greater by $\boldsymbol{e^{(1.68 \times 0.50)}}$ in the RF+.
- Odds of being in group 4, 5 or 6 rather than group 1 or 2 is greater by $\boldsymbol{e^{0.83}}$ per copy of the shared epitope.
- Odds of being in group 3 rather than group 1 or 2 is greater by $\boldsymbol{e^{(0.83 \times 0.50)}}$ per copy of the shared epitope.

Relaxing Assumptions

- In theory, can relax the assumptions of the cumulative odds and continuation ratio models.
- Fit a separate β for each level of the outcome.
- But can theoretically produce negative probabilities p$(y \geq 3) \geq$ p$(y \geq 2)$.
- May want to introduce constraints to reduce the number of parameters (partial proportional odds).
- Model fit is similar, parameter interpretations differ.
- May need to choose model on grounds other than goodness of fit.

Conclusions

- Importance of relationship between predictors and outcome.
- An ordinal outcome need not have an ordinal relationship with predictors.
- Several models may fit: ease of interpretation may be the deciding factor.
- Constraints can be used to reduce the number of parameters and simplify interpretation.

References

[1] S. Greenland. Alternative models for ordinal logistic regression. *Statistics in Medicine*, 13:1665–1677, 1994.

[2] J. A. Anderson. Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B.*, 46(1):1–30, 1984.