S. Partial proportional odds model

references: Stokes, Davis, & Koch (2000), p. 533-541

Motivation:

- The proportional odds model is a popular choice for a univariate ordinal response. When the proportional odds assumption is rejected, one course of action is to fit a generalized logits model, treating the response as nominal. Adjacent category or continuation-ratio logits are other options.
- An alternative procedure is to fit a partial proportional odds model. In this extension of the (full) proportional odds model, inference remains directed at the set of cumulative logits. However, it differs from the standard approach in that regression coefficients are allowed to vary for certain covariates across logits. These are the covariates identified as not satisfying proportional odds.
- GEE methodology (for a multivariate outcome) can be used to fit the partial proportional odds model. This is done by specifying for each subject a set of $r-1$ binary outcomes corresponding to the possible cut-offs in the r-category ordinal response variable.

Chronic respiratory disease example

- Consider data from an epidemiological study of chronic respiratory disease, page 252-257, SDK (2000). The assumption of proportional odds is not rejected

$$Q_s = 12.07, 8 \text{ df}, p = .1479$$

- Even though the proportional odds assumption was met, we will fit a partial proportional odds model for the purpose of illustration. (In section 15.13 of SDK, an illustration is given for which the proportional odds assumption is rejected.)
- The response is chronic respiratory disease; subjects were assigned one of four possible categories
  - Level I: no symptoms
  - Level II: cough or phlegm less than 3 mos/yr
  - Level III: cough or phlegm more than 3 mos/yr
  - Level IV: cough and phlegm plus shortness of breath more than 3 mos/yr
- Covariates are air pollution (low,high), job exposore (no,yes) and smoking status (Non, Ex, Current)

Partial proportional odds model

• A full nonproportional odds model specifies separate logits for the cumulative odds

$$\text{logit}(\theta_{ik}) = \alpha_k + \boldsymbol{x_i'}\boldsymbol{\beta}_k \qquad k = 2, 3, 4$$

where $\theta_{ik} = \pi_{ik} + \ldots \pi_{i4}$, and $\boldsymbol{x_i'} = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$.

$$x_1 = \begin{cases} 1 & \text{if air='high'} \\ 0 & \text{if air='low'} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{if exposure='Yes'} \\ 0 & \text{if exposure='No'} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{if smoker='Ex'} \\ 0 & \text{otherwise} \end{cases}$$

$$x_4 = \begin{cases} 1 & \text{if smoker='Current'} \\ 0 & \text{otherwise} \end{cases}$$

• The proportional odds model specifies $\beta_k = \beta$ for all k.

• The partial proportional odds model is

$$\text{logit}(\theta_{ik}) = \alpha_k + \boldsymbol{x_{i1}'}\boldsymbol{\beta}_1 + \boldsymbol{x_{i2}'}\boldsymbol{\beta}_{2k} \quad k = 2, 3, 4$$

where $\boldsymbol{\beta}_1$ are the regression parameters for those covariates for which we assume proportional odds, and $\boldsymbol{\beta}_{22}, \boldsymbol{\beta}_{23}, \boldsymbol{\beta}_{24}$ are the regression parameters for which odds are not proportional.

SAS code: Proportional odds model

```
data A;
   input iair iexp smk level count @@;
 /* iair =1 if air=high, 0 if low */
 /* iexp =1 if exposure=yes, 0 if no */
 /* smk = 0 if non, 1 if ex, 2 if current */
   ismkex = (smk=1);
   ismkcur = (smk=2);
cards;
 0 0 0 1 158  0 0 0 2 9    0 0 0 3 5   0 0 0 4 0
 0 0 1 1 167  0 0 1 2 19   0 0 1 3 5   0 0 1 4 3
 0 0 2 1 307  0 0 2 2 102  0 0 2 3 83  0 0 2 4 68
 0 1 0 1 26   0 1 0 2 5    0 1 0 3 5   0 1 0 4 1
 0 1 1 1 38   0 1 1 2 12   0 1 1 3 4   0 1 1 4 4
 0 1 2 1 94   0 1 2 2 48   0 1 2 3 46  0 1 2 4 60
 1 0 0 1 94   1 0 0 2 7    1 0 0 3 5   1 0 0 4 1
 1 0 1 1 67   1 0 1 2 8    1 0 1 3 4   1 0 1 4 3
 1 0 2 1 184  1 0 2 2 65   1 0 2 3 33  1 0 2 4 36
 1 1 0 1 32   1 1 0 2 3    1 1 0 3 6   1 1 0 4 1
 1 1 1 1 39   1 1 1 2 11   1 1 1 3 4   1 1 1 4 2
 1 1 2 1 77   1 1 2 2 48   1 1 2 3 39  1 1 2 4 51 ;
run;

proc sort;
  by descending level;
run;
proc logistic order=data;
  weight count;
  model level = iair iexp ismkex ismkcur;
run;
```

SAS output: Proportional odds model

The LOGISTIC Procedure

Response Profile

Ordered Value	level	Total Frequency	Total Weight
1	4	11	230.0000
2	3	12	239.0000
3	2	12	337.0000
4	1	12	1283.0000

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
12.0745	8	0.1479

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 4	1	-3.8938	0.1779	479.2836	<.0001
Intercept 3	1	-2.9696	0.1693	307.7931	<.0001
Intercept 2	1	-2.0884	0.1633	163.5861	<.0001
iair	1	-0.0393	0.0937	0.1758	0.6750
iexp	1	0.8648	0.0955	82.0603	<.0001
ismkex	1	0.4000	0.2019	3.9267	0.0475
ismkcur	1	1.8527	0.1650	126.0383	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
iair	0.961	0.800	1.155
iexp	2.374	1.969	2.863
ismkex	1.492	1.004	2.216
ismkcur	6.377	4.615	8.812

SAS code for separate logistic regressions

```
/* form 3 binary outcomes on same record  */

data b(drop=i count);
   set a;
   z4 = (level=4);
   z3 = (level ge 3);
   z2 = (level ge 2);
   do i=1 to count;
     id = _n_*1000 + i;
     output;
   end;
run;

/* fit separate models to 3 logits */
proc logistic order=data;
  model z4 = iair iexp ismkex ismkcur;
run;
proc logistic order=data;
  model z3 = iair iexp ismkex ismkcur;
run;
proc logistic order=data;
  model z2 = iair iexp ismkex ismkcur;
run;
```

Results of separate logistic regressions vs Proportional odds model

	model $\hat{\boldsymbol{\beta}}$ (s.e.)			
covariate	k=4	k=3	k=2	prop odds
intercept(k=4)	-5.03 (.59)	–	–	-3.89 (.18)
intercept(k=3)	–	-2.84 (.22)	–	-2.97 (.17)
intercept(k=2)	–	–	-2.10 (.17)	-2.09 (.16)
air	-.030 (.15)	-.10 (.11)	-.006 (.10)	-.039 (.09)
exp	.88 (.15)	.86 (.11)	.85 (.10)	.86 (.10)
ex-smoker	1.25 (.65)	.036 (.29)	.44 (.20)	.40 (.20)
current smoker	3.02 (.59)	1.76 (.22)	1.84 (.17)	1.85 (.17)

- The regression coefficients for air and exp are similar across the three logits. This supports proportional odds.
- The regression coefficients for smoking vary, especially for the logit (k=4) comparing the most severe disease (level=IV) with the other levels, suggesting odds are not proportional.
- To fit a partial proportional odds model, we create multiple records per subject.

```
proc print data=b;
   title 'data b has one record per person'; run;
/* create 3 binary outcomes for each person */
data c(drop=z4 z3 z2);
   set b;
   y = z4; odds=4; odds4=1; odds3=0; odds2=0; output;
   y = z3; odds=3; odds4=0; odds3=1; odds2=0; output;
   y = z2; odds=2; odds4=0; odds3=0; odds2=1; output;
run;
proc print data=c;
   title 'data c has three records per person'; run;
```

Data Structure for partial proportional odds model: 4 subjects listed

data b has one record per person

Obs	iair	iexp	smk	level	ismkex	ismkcur	z4	z3	z2	id
1	0	0	1	4	1	0	1	1	1	2001
231	0	0	0	3	0	0	0	1	1	13001
470	0	0	0	2	0	0	0	0	1	25001
807	0	0	0	1	0	0	0	0	0	37001

data c has three records per person

Obs	iair	iexp	smk	level	ismkex	ismkcur	id	y	odds	odds4	odds3	odds2
1	0	0	1	4	1	0	2001	1	4	1	0	0
2	0	0	1	4	1	0	2001	1	3	0	1	0
3	0	0	1	4	1	0	2001	1	2	0	0	1
691	0	0	0	3	0	0	13001	0	4	1	0	0
692	0	0	0	3	0	0	13001	1	3	0	1	0
693	0	0	0	3	0	0	13001	1	2	0	0	1
1408	0	0	0	2	0	0	25001	0	4	1	0	0
1409	0	0	0	2	0	0	25001	0	3	0	1	0
1410	0	0	0	2	0	0	25001	1	2	0	0	1
2419	0	0	0	1	0	0	37001	0	4	1	0	0
2420	0	0	0	1	0	0	37001	0	3	0	1	0
2421	0	0	0	1	0	0	37001	0	2	0	0	1

- The new response variable is ”y”.
- ”odds2”, ”odds3”, and ”odds4” are indicators for the logit.

Nonproportional odds model

- First, fit the full nonproportional odds models

```
proc genmod data=c  descending;
  class id odds iair iexp smk;
  title 'Nonproportional odds model';
  model y = odds iair iexp smk odds*iair odds*iexp odds*smk
           / dist=bin scale=1 noscale noint type3 wald;
  repeated subject=id/ type = un corrw;
run;
```

- SAS output

```
                    Nonproportional odds model
                       GEE Model Information
          Correlation Structure                    Unstructured
          Subject Effect                       id (2089 levels)
          Number of Clusters                               2089
          Correlation Matrix Dimension                        3
          Maximum Cluster Size                                3
          Minimum Cluster Size                                3

                    Working Correlation Matrix
                         Col1          Col2          Col3
            Row1       1.0000        0.5916        0.3746
            Row2       0.5916        1.0000        0.6418
            Row3       0.3746        0.6418        1.0000
```

- The unstructured R provides a more powerful assessment of logit type interactions than the independence R; it tends to produce smaller standard errors for within-subject effects. It more closely approximates the actual (multinomial) model-based covariance structure which is a function of the means only (given that the response is really univariate).

SAS Output for full nonproportional odds model

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept			0.0000	0.0000	0.0000	0.0000	.	.
odds	2		0.5723	0.1023	0.3718	0.7729	5.59	<.0001
odds	3		-0.3261	0.1076	-0.5370	-0.1152	-3.03	0.0024
odds	4		-1.1376	0.1294	-1.3913	-0.8840	-8.79	<.0001
iair	0		0.0227	0.1465	-0.2644	0.3098	0.15	0.8769
iair	1		0.0000	0.0000	0.0000	0.0000	.	.
iexp	0		-0.9042	0.1447	-1.1878	-0.6205	-6.25	<.0001
iexp	1		0.0000	0.0000	0.0000	0.0000	.	.
smk	0		-3.0561	0.6071	-4.2459	-1.8663	-5.03	<.0001
smk	1		-1.7837	0.3062	-2.3840	-1.1835	-5.82	<.0001
smk	2		0.0000	0.0000	0.0000	0.0000	.	.
odds*iair	2	0	-0.0028	0.1407	-0.2786	0.2730	-0.02	0.9841
odds*iair	2	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*iair	3	0	0.0897	0.1157	-0.1371	0.3164	0.78	0.4383
odds*iair	3	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*iair	4	0	0.0000	0.0000	0.0000	0.0000	.	.
odds*iair	4	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*iexp	2	0	0.0567	0.1406	-0.2188	0.3322	0.40	0.6865
odds*iexp	2	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*iexp	3	0	0.0348	0.1150	-0.1905	0.2602	0.30	0.7618
odds*iexp	3	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*iexp	4	0	0.0000	0.0000	0.0000	0.0000	.	.
odds*iexp	4	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*smk	2	0	1.2228	0.5908	0.0649	2.3807	2.07	0.0385
odds*smk	2	1	0.3795	0.2891	-0.1871	0.9461	1.31	0.1893
odds*smk	2	2	0.0000	0.0000	0.0000	0.0000	.	.
odds*smk	3	0	1.2986	0.5681	0.1850	2.4121	2.29	0.0223
odds*smk	3	1	0.0540	0.2397	-0.4158	0.5238	0.23	0.8218
odds*smk	3	2	0.0000	0.0000	0.0000	0.0000	.	.
odds*smk	4	0	0.0000	0.0000	0.0000	0.0000	.	.
odds*smk	4	1	0.0000	0.0000	0.0000	0.0000	.	.
odds*smk	4	2	0.0000	0.0000	0.0000	0.0000	.	.

SAS Output for full nonproportional odds model

```
Partial proportional odds model: all interactions

                The GENMOD Procedure

       Wald Statistics For Type 3 GEE Analysis

                              Chi-
   Source           DF      Square    Pr > ChiSq

   odds              2      187.78        <.0001
   iair              1        0.26        0.6093
   iexp              1       73.96        <.0001
   smk               2      143.36        <.0001
   odds*iair         2        1.45        0.4847
   odds*iexp         2        0.16        0.9218
   odds*smk          4        9.29        0.0542
```

- Proportional odds is not contradicted for air and exp.
- The is some evidence to suggest the proportional odds is not met for smoking
- Fit a partial proportional odds model with proportional odds for air and exp only.

$$\text{logit}(\theta_{i4}) = \alpha_4 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_{34} + x_{i4}\beta_{44}$$

$$\text{logit}(\theta_{i3}) = \alpha_3 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_{33} + x_{i4}\beta_{43}$$

$$\text{logit}(\theta_{i2}) = \alpha_2 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_{32} + x_{i4}\beta_{42}$$

Wald test for proportional odds for smoking

Let $\boldsymbol{\beta}' = (\alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_{32}, \beta_{33}, \beta_{34}, \beta_{42}, \beta_{43}, \beta_{44})$

$H_0 : \beta_{32} = \beta_{33} = \beta_{34}$ and $\beta_{42} = \beta_{43} = \beta_{44}$

is equivalent to

$$H_0 : \begin{pmatrix} \beta_{32} - \beta_{33} \\ \beta_{32} - \beta_{34} \\ \beta_{42} - \beta_{43} \\ \beta_{42} - \beta_{44} \end{pmatrix} = 0 \quad \rightarrow \quad H_0 : C\boldsymbol{\beta} = \mathbf{0}$$

where

$$\underbrace{C}_{4\times 11} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

test statistic

$$Q_W = (C\hat{\boldsymbol{\beta}})^T (C V_{\hat{\boldsymbol{\beta}}} C^T)^{-1} (C\hat{\boldsymbol{\beta}}) \sim \chi^2_4$$

SAS Code for partial proportional odds model

- Using the default parameterization.

```
proc genmod data=c  descending;
  class id odds iair iexp smk;
  title 'Partial proportional odds model: smoking by logit interaction';
  model y = odds4 odds3 odds2 iair iexp smk odds*smk
            / dist=bin scale=1 noscale noint type3 wald;
  repeated subject=id/ type = un corrw;
run;
```

- The Wald test for odds*smk is $Q_W = 9.26$ with p=055. The Score test (not shown) is Q_S=15.85 with p=.0032.

Wald Statistics For Type 3 GEE Analysis

Source	DF	Chi-Square	Pr > ChiSq
odds4	0	.	.
odds3	0	.	.
odds2	0	.	.
iair	1	0.17	0.6789
iexp	1	79.49	<.0001
smk	2	145.12	<.0001
odds*smk	4	9.26	0.0550

- There is evidence to suggest that the proportional odds assumption is not met smoking.

SAS Code for partial proportional odds model

- Define interactions so that nonsmokers are reference.

```
data d;  set c;
  o4sx = odds4*ismkex;  o4sc = odds4*ismkcur;
  o3sx = odds3*ismkex;  o3sc = odds3*ismkcur;
  o2sx = odds2*ismkex;  o2sc = odds2*ismkcur;
proc genmod data=d  descending;
  class id;
  model y = odds4 odds3 odds2 iair iexp ismkex ismkcur
     o3sx o3sc o2sx o2sc / dist=bin noint covb;
  repeated subject=id/ type = un corrw;
run;
```

- SAS output shows differences in smoking*logit interaction with respect to logit (k=4).

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	0.0000	0.0000	0.0000	0.0000	.	.
odds4	-5.0361	0.5928	-6.1979	-3.8743	-8.50	<.0001
odds3	-2.8681	0.2126	-3.2847	-2.4515	-13.49	<.0001
odds2	-2.0829	0.1605	-2.3975	-1.7683	-12.98	<.0001
iair	-0.0391	0.0943	-0.2240	0.1459	-0.41	0.6789
iexp	0.8623	0.0967	0.6728	1.0519	8.92	<.0001
ismkex	1.2657	0.6608	-0.0295	2.5608	1.92	0.0554
ismkcur	3.0444	0.5959	1.8766	4.2123	5.11	<.0001
o3sx	-1.2317	0.6017	-2.4110	-0.0523	-2.05	0.0407
o3sc	-1.2841	0.5588	-2.3794	-0.1888	-2.30	0.0216
o2sx	-0.8382	0.6394	-2.0914	0.4150	-1.31	0.1899
o2sc	-1.2134	0.5803	-2.3507	-0.0760	-2.09	0.0365

Results of Partial proportional odds model

- Parameter estimates for each logit can be obtained from the previous estimates or with:

```
proc genmod data=d  descending;  class id;
  model y = odds4 odds3 odds2 iair iexp o4sx o4sc o3sx o3sc o2sx o2sc
      / dist=bin noint;   repeated subject=id/ type = un corrw;
```

	model $\hat{\boldsymbol{\beta}}$ (s.e.)			
covariate	k=4	k=3	k=2	prop odds
intercept(k=4)	-5.04 (.59)			-3.89 (.18)
intercept(k=3)		-2.87 (.21)		-2.97 (.17)
intercept(k=2)			-2.08 (.16)	-2.09 (.16)
air	-.039 (.09)	-.039 (.09)	-.039 (.09)	-.039 (.09)
exp	-.86 (.10)	-.86 (.10)	-.86 (.10)	.86 (.10)
ex-smoker	1.27 (.66)	.034 (.29)	.43 (.20)	.40 (.20)
current smoker	3.04 (.60)	1.76 (.22)	1.83 (.17)	1.85 (.17)

- Ex-smokers have approx $e^{1.27} = 3.56$ time higher odds of level IV disease than non-smokers.
- Current smokers have approx $e^{3.04} = 20.9$ time higher odds of level IV disease than non-smokers.
- Ex-smokers have approx $e^{.034} = 1.03$ time higher odds of level III or IV disease than non-smokers.
- Current smokers have approx $e^{1.76} = 5.81$ time higher odds of level III or IV disease than non-smokers.
- Ex-smokers have approx $e^{.43} = 1.54$ time higher odds of level II, III or IV disease than non-smokers.
- Current smokers have approx $e^{1.83} = 6.23$ time higher odds of level II, III or IV disease than non-smokers.