The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The Stata Journal (2014)
14, Number 4, pp. 947–964

adjcatlogit, ccrlogit, and ucrlogit: Fitting ordinal logistic regression models

Morten W. Fagerland
Oslo Centre for Biostatistics and Epidemiology
Research Support Services
Oslo University Hospital
Oslo, Norway
morten.fagerland@medisin.uio.no

Abstract. In this article, I present three commands that perform adjacent-category logistic regression (`adjcatlogit`), constrained continuation-ratio logistic regression (`ccrlogit`), and unconstrained continuation-ratio logistic regression (`ucrlogit`) for ordered response data.

Keywords: st0367, adjcatlogit, ccrlogit, ucrlogit, ordinal models, ordered regression, logistic models, adjacent category, continuation ratio

1 Introduction

Ordinal response regression models are used to describe the relationship between an ordered categorical response (dependent) variable and one or more explanatory (independent) variables. The literature on ordinal regression models includes general textbooks such as those by McCullagh and Nelder (1989); Agresti (2010); and Hosmer, Lemeshow, and Sturdivant (2013). It also includes review articles such as those by Greenland (1994) and Ananth and Kleinbaum (1997). There are several models we can use. When choosing a model, we must first select a link function, which describes the functional relationship between the dependent variable and the independent variables. We then must choose which response categories to compare.

The main commands for ordinal regression are `ologit` and `oprobit`. `ologit` fits proportional-odds logistic regression models, also called parallel-lines models. The link function is the logit transformation, which is also used in (ordinary) binary logistic regression (`logit` or `logistic`) and multinomial logistic regression (`mlogit`). `oprobit` fits ordered probit regression models, where the link function is the standard normal cumulative distribution function. Also available is the stereotype logistic regression model (`slogit`), which is a compromise between the multinomial and ordered logistic regression models (see [R] **slogit**).

In this article, I focus on ordinal models that use the logit link, that is, ordinal logistic regression models. Logistic models have several advantages, including mathematical flexibility and ease of use. Also the exponential form of the regression coefficients can be interpreted as odds ratios (ORs) (Hosmer, Lemeshow, and Sturdivant 2013). Several different logistic models are possible. The proportional odds model (`ologit`) compares two sets of response categories: an equal or smaller response versus a larger response. The adjacent category model compares each response category with the next larger response category. The constrained and unconstrained continuation-ratio models compare each response category with all lower response categories. The unconstrained continuation-ratio model defines $c-1$ regression coefficients for each independent variable—where c is the number of response categories—whereas the other models describe the effect of each independent variable using a single regression coefficient. When choosing a model for a particular problem, users should consider which model provides the most informative comparisons for the subject matter as well as the desired amount of model flexibility.

Here I describe three commands for adjacent-category logistic regression (`adjcatlogit`), constrained continuation-ratio logistic regression (`ccrlogit`), and unconstrained continuation-ratio logistic regression (`ucrlogit`).

2 Model definitions

Let the dependent variable Y take on c possible values $1, 2, \ldots, c$, and let $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ denote a vector of p independent variables. The conditional probability of a response equal to category j given $\mathbf{x}$ is denoted by $\pi_j = P(Y = j|\mathbf{x})$ for $j = 1, 2, \ldots, c$. Each model is defined through the $c-1$ logits (logit equations) $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, …, $g_{c-1}(\mathbf{x})$, which relate a set of intercepts (αs) and regression coefficients (βs) to the probability of the response categories. I define the logits for each model in sections 2.1–2.3.

2.1 The adjacent-category logistic regression model

The adjacent category model compares each response category (except the first) with the next larger response category.

$$\begin{aligned} g_j(\mathbf{x}) &= \log\left\{\frac{P(Y = j+1|\mathbf{x})}{P(Y = j|\mathbf{x})}\right\} \\ &= \alpha_j + \boldsymbol{\beta}'\mathbf{x} \qquad\qquad j = 1, \ldots, c-1 \end{aligned}$$

The regression coefficients, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$, are constant across the logits, whereas the intercepts (α_j) are not. For an independent variable x_k, we may interpret $\exp(\beta_k)$ as the OR for a one-unit increase in x_k, comparing response category $j+1$ with response category j. Thus we have the following equalities,

$$\exp(\beta_k) = \text{OR}(2,1) = \text{OR}(3,2) = \cdots = \text{OR}(c, c-1)$$

where $\text{OR}(a,\, b)$ denotes the OR for comparing category a with category b. As shown in Hosmer, Lemeshow, and Sturdivant (2013, 290–291), we also have that

$$\log\left\{\text{OR}(j,1)\right\} = (j-1) \times \left\{\text{OR}(2,1)\right\} \qquad\qquad j = 3, \ldots, c \tag{1}$$

Equation (1) provides the rationale for fitting the adjacent category model via a constrained multinomial model. The necessary constraints are $\beta_{jk} = (j-1) \times \beta_{2k}$, for

$j = 3, \ldots, c$ and $k = 1, \ldots, p$, where β_{jk} is the multinomial coefficient comparing $Y = j$ with $Y = 1$ (the reference category) for the kth independent variable.

Formulas for the conditional probabilities of the adjacent category model were derived in Fagerland and Hosmer (2014) and are given as

$$\pi_1 = \frac{1}{1+\theta}$$

and

$$\pi_{j+1} = \frac{\theta_j}{1+\theta} \qquad\qquad j = 1, \ldots, c-1$$

where $\theta = \sum_{k=1}^{c-1} \theta_k$ and

$$\theta_j = \exp\left\{\sum_{k=1}^{j} g_k(\mathbf{x})\right\} \qquad\qquad j = 1, \ldots, c-1$$

2.2 The constrained continuation-ratio logistic regression model

The constrained continuation-ratio model compares each response category with all lower response categories.

$$\begin{aligned} g_j(\mathbf{x}) &= \log\left\{\frac{P(Y = j|\mathbf{x})}{P(Y < j|\mathbf{x})}\right\} \\ &= \alpha_j + \boldsymbol{\beta}'\mathbf{x} \qquad\qquad j = 2, \ldots, c \end{aligned} \tag{2}$$

As with the adjacent category model, the regression coefficients are constant across the logits, and we may describe the relationship between an independent variable x_k and the dependent variable Y by a single coefficient or OR. The following interpretations apply:

$$\exp(\beta_k) = \text{OR}(2,1) = \text{OR}(3, 1 \ldots 2) = \cdots = \text{OR}(c, 1 \ldots c-1)$$

Note that we can obtain another version of the constrained continuation-ratio model by substituting the denominator in (2) with $P(Y > j|\mathbf{x})$ and possibly changing the sign of $\boldsymbol{\beta}'\mathbf{x}$. The resulting model can be fit with the user-written program `ocratio` (Wolfe 1998). The two models are, however, not equivalent (Hosmer, Lemeshow, and Sturdivant 2013, 291). Both models can be fit with a generalized linear model formulation.

Fagerland and Hosmer (2014) derived formulas for the conditional probabilities of the constrained continuation-ratio model as defined by `ocratio`. The formulas for the model in (2) are similar and given as

$$\pi_c = \frac{e^{g_c(\mathbf{x})}}{1+e^{g_c(\mathbf{x})}}$$

$$\pi_j = \frac{\gamma_{j+1} \times e^{g_j(\mathbf{x})}}{1+e^{g_j(\mathbf{x})}} \qquad j = 2, \ldots, c-1$$

and

$$\pi_1 = 1 - \sum_{k=2}^{c} \pi_k$$

where

$$\gamma_j = 1 - \sum_{k=j}^{c} \pi_k \qquad j = 3, \ldots, c$$

2.3 The unconstrained continuation-ratio logistic regression model

The unconstrained continuation-ratio model is equal to the constrained model in section 2.2, except that we let the regression coefficients vary across the logits

$$\begin{aligned} g_j(\mathbf{x}) &= \log\left\{\frac{P(Y=j|\mathbf{x})}{P(Y<j|\mathbf{x})}\right\} \\ &= \alpha_j + \boldsymbol{\beta}_j'\mathbf{x} \qquad j = 2, \ldots, c \end{aligned}$$

where $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp})'$ for $j = 2, \ldots, c$. We now have $c-1$ coefficients or ORs to describe the effect of an independent variable on the response. For independent variable x_k, we have the following:

$$\begin{aligned} \exp(\beta_{2k}) &= \mathrm{OR}(2,1) \\ \exp(\beta_{3k}) &= \mathrm{OR}(3, 1 \ldots 2) \\ &\vdots \\ \exp(\beta_{ck}) &= \mathrm{OR}(c, 1 \ldots c-1) \end{aligned}$$

The number of parameters in this model is the same as the number of parameters in the multinomial logistic regression model. The unconstrained continuation-ratio model may be fit using $c - 1$ binary logistic regression models, where the binary response variables are defined as

$$Y_j^* = \begin{cases} 1 & \text{if } Y = j \\ 0 & \text{if } Y < j \\ \text{missing} & \text{if } Y > j \end{cases}$$

for $j = 2, \ldots, c$.

The expressions for the conditional probabilities of the unconstrained continuation-ratio model are equal to those for the constrained continuation-ratio model in section 2.2 for given values of the logits $g_2(\mathbf{x})$, ..., $g_c(\mathbf{x})$; however, the evaluations of $g_j(\mathbf{x}_i)$ for a particular observation i are different for the two models.

3 The estimation commands

3.1 adjcatlogit (adjacent-category logistic regression)

Syntax

`adjcatlogit` *depvar* [*indepvars*] [*if*] [*in*] [`, level(`*#*`) or listconstraints`]

indepvars may contain factor variables.

Syntax for predict

`predict` {*newvarname* | *newvarlist*} [*if*] [*in*] [`, pr xb outcome(`*outcome*`)`]

Description

`adjcatlogit` fits adjacent-category logistic regression models of ordinal variable *depvar* on the independent variables *indepvars*. The actual values taken on by the dependent variable are irrelevant, except that larger values are assumed to correspond to "higher" outcomes.

Options for adjcatlogit

`level(`*#*`)` specifies the confidence level, as a percentage, for the confidence interval (CI). The default is `level(95)` or as set by `set level`.

`or` reports the estimated coefficients transformed to ORs, that is, $\exp(\beta)$ rather than β. Standard errors and CIs are similarly transformed. This option affects how results are displayed, not how they are estimated. `or` may be specified at estimation or when replaying previously estimated results.

`listconstraints` requests that a list of the constraints used by `mlogit` to fit the model be displayed.

Options for predict

`pr` calculates the predicted probabilities. This is the default. If you do not also specify the `outcome()` option, you specify one or c new variables, where c is the number of categories of the dependent variable. If you specify one new variable (and no `outcome()` option), `outcome(#1)` is assumed. If you specify the `outcome()` option, you must specify one new variable.

`xb` calculates the linear prediction. You specify one new variable (and no `outcome()` option). The contributions of the estimated constants are ignored in the calculations.

`outcome(`*outcome*`)` specifies the outcome for which the predicted probabilities are to be calculated. `outcome()` should contain either one value of the dependent variable or one of `#1`, `#2`, ... with `#1` meaning the first category of the dependent variable, `#2` meaning the second category, etc.

Remarks

`adjcatlogit` fits the adjacent category model using constrained multinomial logistic regression (`mlogit`), where the lowest category of the dependent variable is used as the reference category. The `constraint free` command is used to select free (unused) constraint numbers. The constraints are not dropped after estimation, so the number of free constraints is reduced each time an adjacent category model is estimated. The maximum number of constraints in Stata is 1,999, which is sufficient for a large number of estimations with `adjcatlogit`. If there are not enough free constraints, `adjcatlogit` will exit and give the error `no free constraints`. If that happens, `constraint drop` can be used to increase the number of free constraints.

Stored results

`adjcatlogit` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(k_cat)`	number of categories
`e(k_exp)`	number of auxiliary parameters
`e(df_m)`	model degrees of freedom
`e(df_0)`	degrees of freedom, constant-only model
`e(r2_p)`	pseudo-R-squared
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model
`e(chi2)`	χ^2
`e(p)`	significance
Macros	
`e(cmd)`	`adjcatlogit`
`e(cmdline)`	command as typed
`e(constraints)`	list of constraints
`e(depvar)`	name of dependent variable
`e(title)`	title in estimation output
`e(chi2type)`	`Wald` or `LR`; type of model chi-squared test
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
Matrices	
`e(b)`	coefficient vector
`e(cat)`	category values
`e(V)`	variance–covariance matrix of the estimators
Functions	
`e(sample)`	marks estimation sample

3.2 ccrlogit (constrained continuation-ratio logistic regression)

Syntax

`ccrlogit` *depvar* [*indepvars*] [*if*] [*in*] [, `level(`*#*`)` `or`]

indepvars may contain factor variables.

Syntax for predict

`predict` {*newvarname* | *newvarlist*} [*if*] [*in*] [, `pr` `xb` `outcome(`*outcome*`)`]

Description

`ccrlogit` fits constrained continuation-ratio logistic regression models of ordinal variable *depvar* on the independent variables *indepvars*. The actual values taken on by the dependent variable are irrelevant, except that larger values are assumed to correspond to "higher" outcomes.

Options for ccrlogit

`level(#)` specifies the confidence level, as a percentage, for the CI. The default is `level(95)` or as set by `set level`.

`or` reports the estimated coefficients transformed to ORs, that is, $\exp(\beta)$ rather than β. Standard errors and CIs are similarly transformed. This option affects how results are displayed, not how they are estimated. `or` may be specified at estimation or when replaying previously estimated results.

Options for predict

`pr` calculates the predicted probabilities. This is the default. If you do not also specify the `outcome()` option, you specify one or c new variables, where c is the number of categories of the dependent variable. If you specify one new variable (and no `outcome()` option), `outcome(#1)` is assumed. If you specify the `outcome()` option, you must specify one new variable.

`xb` calculates the linear prediction. You specify one new variable (and no `outcome()` option). The contributions of the estimated constants are ignored in the calculations.

`outcome(`*outcome*`)` specifies the outcome for which the predicted probabilities are to be calculated. `outcome()` should contain either one value of the dependent variable or one of `#1`, `#2`, ... with `#1` meaning the first category of the dependent variable, `#2` meaning the second category, etc.

Remarks

`ccrlogit` fits the constrained continuation-ratio model using a generalized linear model (`glm`).

Stored results

`ccrlogit` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(k_cat)`	number of categories
`e(k_exp)`	number of auxiliary parameters
`e(df_m)`	model degrees of freedom
`e(df_0)`	degrees of freedom, constant-only model
`e(r2_p)`	pseudo-R-squared
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model
`e(chi2)`	χ^2
`e(p)`	significance
Macros	
`e(cmd)`	`ccrlogit`
`e(cmdline)`	command as typed
`e(depvar)`	name of dependent variable
`e(title)`	title in estimation output
`e(chi2type)`	`Wald` or `LR`; type of model chi-squared test
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
Matrices	
`e(b)`	coefficient vector
`e(cat)`	category values
`e(V)`	variance–covariance matrix of the estimators
Functions	
`e(sample)`	marks estimation sample

3.3 ucrlogit (unconstrained continuation-ratio logistic regression)

Syntax

`ucrlogit` *depvar* [*indepvars*] [*if*] [*in*] [, `level(`#`)` `or`]

indepvars may contain factor variables.

Syntax for predict

`predict` {*newvarname* | *newvarlist*} [*if*] [*in*] [, `pr` `xb` `outcome(`*outcome*`)`]

Description

`ucrlogit` fits unconstrained continuation-ratio logistic regression models of ordinal variable *depvar* on the independent variables *indepvars*. The actual values taken on by the dependent variable are irrelevant, except that larger values are assumed to correspond to "higher" outcomes.

Options for ucrlogit

`level(#)` specifies the confidence level, as a percentage, for the CI. The default is `level(95)` or as set by `set level`.

`or` reports the estimated coefficients transformed to ORs, that is, $\exp(\beta)$ rather than β. Standard errors and CIs are similarly transformed. This option affects how results are displayed, not how they are estimated. `or` may be specified at estimation or when replaying previously estimated results.

Options for predict

`pr` calculates the predicted probabilities. This is the default. If you do not also specify the `outcome()` option, you specify one or c new variables, where c is the number of categories of the dependent variable. If you specify one new variable (and no `outcome()` option), `outcome(#1)` is assumed. If you specify the `outcome()` option, you must specify one new variable.

`xb` calculates the linear predictions. If you do not also specify the `outcome()` option, you specify one or c new variables, where c is the number of categories of the dependent variable. If you specify one new variable (and no `outcome()` option), `outcome(#1)` is assumed. If you specify the `outcome()` option, you must specify one new variable.

`outcome(`*outcome*`)` specifies the outcome or logit for which the predicted probabilities or linear predictions, respectively, are to be calculated. `outcome()` should contain either one value of the dependent variable or one of `#1`, `#2`, ... with `#1` meaning the first category of the dependent variable or the first logit, `#2` meaning the second category or the second logit, etc.

Remarks

`ucrlogit` fits the unconstrained continuation-ratio model using $c - 1$ binary logistic regression models (`logit`), where c is the number of categories of the dependent variable.

Stored results

`ucrlogit` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(k_cat)`	number of categories
`e(k_exp)`	number of auxiliary parameters
`e(df_m)`	model degrees of freedom
`e(df_0)`	degrees of freedom, constant-only model
`e(r2_p)`	pseudo-R-squared
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model
`e(chi2)`	χ^2
`e(p)`	significance
Macros	
`e(cmd)`	`ucrlogit`
`e(cmdline)`	command as typed
`e(depvar)`	name of dependent variable
`e(title)`	title in estimation output
`e(chi2type)`	`Wald` or `LR`; type of model chi-squared test
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
Matrices	
`e(b)`	coefficient vector
`e(cat)`	category values
`e(V)`	variance–covariance matrix of the estimators
Functions	
`e(sample)`	marks estimation sample

4 Examples

Here we use the well-known low birthweight (`lbw.dta`) dataset accessible in Stata through the `webuse` command.

```
. webuse lbw
(Hosmer & Lemeshow data)
```

This dataset is described in Hosmer, Lemeshow, and Sturdivant (2013, 24). In short, the dataset contains the birthweight and selected risk factors of low birthweight (birthweight less than 2,500 grams) of 189 children–mother pairs. The continuous variable `bwt` contains the birthweight measured in grams. We form an ordinal variable `bwt4` = 1, 2, 3, 4, such that higher values of `bwt4` represent lower birthweight.

```
. generate bwt4 = .
. replace bwt4 = 1 if bwt > 3500
. replace bwt4 = 2 if bwt <= 3500 & bwt > 3000
. replace bwt4 = 3 if bwt <= 3000 & bwt > 2500
. replace bwt4 = 4 if bwt <= 2500
. tabulate bwt4
```

bwt4	Freq.	Percent	Cum.
1	46	24.34	24.34
2	46	24.34	48.68
3	38	20.11	68.78
4	59	31.22	100.00
Total	189	100.00	

It may seem counterintuitive to define the categories of `bwt4` in the opposite direction of the underlying continuous variable. We choose this coding so that higher category values indicate more unfavorable outcomes, which is consistent with the usual way such variables are coded; for example, level of pain: `1` = none, `2` = mild, `3` = moderate, and `4` = severe. It is also consistent with how dichotomous variables are usually coded; for example, `0` = no disease, `1` = disease, `0` = not exposed, and `1` = exposed.

4.1 Examples using adjcatlogit

We start by fitting an adjacent-category logistic regression model of `bwt4` on `smoke` (smoking status during pregnancy: `0` = no, `1` = yes); `race` (`1` = white, `2` = black, `3` = other); `lwt` (mother's weight in pounds at last menstrual period); `ht` (history of hypertension: `0` = no, `1` = yes); and `ui` (presence of uterine irritability: `0` = no, `1` = yes). The option `or` requests that ORs rather than coefficients be displayed.

```
. adjcatlogit bwt4 smoke i.race lwt ht ui, or
Adjacent-category logistic regression             Number of obs =      189
                                                  LR chi2( 6)   =    45.23
                                                  Prob < chi2   =   0.0000
Log likelihood = -237.03515                       Pseudo R2     =   0.0871

------------------------------------------------------------------------------
        bwt4 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
bwt4         |
       smoke |   1.735194   .2870376     3.33   0.001     1.254704    2.399687
             |
        race |
      black  |   2.260407   .5349141     3.45   0.001     1.421523    3.594341
      other  |   1.568366   .2733047     2.58   0.010     1.114593    2.206879
             |
         lwt |   .9918742   .0026479    -3.06   0.002     .9866979    .9970777
          ht |   1.870931   .5829019     2.01   0.044     1.015921    3.445527
          ui |    1.71599    .370994     2.50   0.013     1.123276    2.621459
-------------+----------------------------------------------------------------
_anc         |
       cons1 |    1.95081   .8535413     1.53   0.127     .8275364     4.59878
       cons2 |   1.319242   .8496561     0.43   0.667     .3733441    4.661651
       cons3 |   2.030535   1.623521     0.89   0.376     .4236719     9.73176
------------------------------------------------------------------------------
```

Smoking during pregnancy, the races "black" or "other", hypertension, and presence of uterine irritability all indicate increased risk of higher values of `bwt4` and thus of lower birthweight. A high value of mother's weight, on the other hand, reduces this risk, although the estimated OR for `lwt` is quite small. The estimated effect of `smoke` on `bwt4` may be interpreted as

$$\widehat{\mathrm{OR}}(2,1) = \widehat{\mathrm{OR}}(3,2) = \widehat{\mathrm{OR}}(4,3) = \exp\left(\widehat{\beta}_k\right) = 1.74 \ (95\%\ \mathrm{CI}\ [1.25,\ 2.40])$$

with similar interpretations for the other estimated effects.

As explained in sections 2.1 and 3.1, `adjcatlogit` uses constrained multinomial logistic regression (`mlogit`) to fit the adjacent category model. A list of the constraints is obtained by specifying the `listconstraints` option, as follows:

```
. adjcatlogit bwt4 smoke i.race lwt ht ui, or listconstraints
Constraints used with mlogit:
  1987:  [3]smoke = 2*[2]smoke
  1986:  [4]smoke = 3*[2]smoke
  1985:  [3]2.race = 2*[2]2.race
  1984:  [4]2.race = 3*[2]2.race
  1983:  [3]3.race = 2*[2]3.race
  1982:  [4]3.race = 3*[2]3.race
  1981:  [3]lwt = 2*[2]lwt
  1980:  [4]lwt = 3*[2]lwt
  1979:  [3]ht = 2*[2]ht
  1978:  [4]ht = 3*[2]ht
  1977:  [3]ui = 2*[2]ui
  1976:  [4]ui = 3*[2]ui
  (output omitted)
```

Factor variables may be specified in the usual manner.

```
. adjcatlogit bwt4 smoke##race lwt ht ui, or
Adjacent-category logistic regression               Number of obs =      189
                                                    LR chi2( 8)   =    49.72
                                                    Prob < chi2   =   0.0000
Log likelihood = -234.79047                         Pseudo R2     =   0.0958

------------------------------------------------------------------------------
        bwt4 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
bwt4         |
       smoke |
      smoker |   2.166554   .4724645     3.55   0.000     1.413014    3.321946
             |
        race |
       black |   2.476267   .7381052     3.04   0.002     1.380633    4.441366
       other |   1.968733   .4246554     3.14   0.002     1.289979    3.004629
             |
  smoke#race |
smoker#black |   .9131809     .44709    -0.19   0.853     .3497934    2.383977
smoker#other |   .4471343   .1705224    -2.11   0.035     .2117471    .9441881
             |
         lwt |   .9923672   .0027256    -2.79   0.005     .9870394    .9977237
          ht |   1.771864   .5593935     1.81   0.070     .9543298    3.289747
          ui |   1.792218   .3952526     2.65   0.008     1.163238    2.761299
-------------+----------------------------------------------------------------
_anc         |
       cons1 |   1.647005   .7491717     1.10   0.273     .6753203      4.0168
       cons2 |   1.068094   .7350177     0.10   0.924     .2772352    4.115009
       cons3 |   1.633601   1.409165     0.57   0.569     .3012234    8.859377
------------------------------------------------------------------------------
```

Following estimation, we can test for the overall effect of the interaction between `smoke` and `race`.

```
. test 1.smoke#2.race = 1.smoke#3.race = 0
 ( 1)  [bwt4]1.smoke#2.race - [bwt4]1.smoke#3.race = 0
 ( 2)  [bwt4]1.smoke#2.race = 0
           chi2(  2) =    4.57
         Prob > chi2 =    0.1020
```

The conditional probabilities of each outcome category can be obtained by `predict` in the usual manner, as follows:

```
. predict p1-p4
(option pr assumed; predicted probability)
```

Likewise for the linear prediction, one can type

```
. predict xbeta, xb
```

4.2 Example using ccrlogit

We now fit a constrained continuation-ratio model using the same variables as in section 4.1.

```
. ccrlogit bwt4 smoke i.race lwt ht ui, or
Constrained continuation-ratio logistic regression     Number of obs =      189
                                                        LR chi2( 6)   =    44.22
                                                        Prob < chi2   =   0.0000
Log likelihood = -237.54456                             Pseudo R2     =   0.0851
```

bwt4	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
bwt4						
smoke	2.286725	.5662308	3.34	0.001	1.407481	3.715228
race						
black	4.081572	1.45312	3.95	0.000	2.031342	8.201096
other	2.086064	.5470363	2.80	0.005	1.24771	3.487718
lwt	.9883587	.0040084	-2.89	0.004	.9805335	.9962464
ht	2.41102	1.201645	1.77	0.077	.9077464	6.403789
ui	2.39648	.7904512	2.65	0.008	1.255495	4.574386
_anc						
cons1	2.423381	1.495945	1.43	0.152	.7227326	8.125792
cons2	.7897336	.4673476	-0.40	0.690	.247604	2.518858
cons3	.7136462	.4102232	-0.59	0.557	.2313073	2.201794

The estimated effect of `smoke` in this case is

$$\widehat{\mathrm{OR}}(2,1) = \widehat{\mathrm{OR}}(3,1\ldots 2) = \widehat{\mathrm{OR}}(4,1\ldots 3) = \exp(\widehat{\beta}_k) = 2.29 \text{ (95\% CI [1.41, 3.72])}$$

which means that the odds of smokers are estimated to be 2.29 times that of nonsmokers for higher values of `bwt4`, that is, lower birthweight. The estimated OR is larger than that of the adjacent category model ($\widehat{\mathrm{OR}} = 1.74$) because the comparisons for the constrained continuation-ratio model include several response categories, whereas the adjacent category model compares only adjacent response categories.

4.3 Example using ucrlogit

The unconstrained continuation-ratio model is similar to the multinomial logistic model in that the effect of each independent variable on the response is described by $c - 1$ parameters.

```
. ucrlogit bwt4 smoke i.race lwt ht ui, or

Unconstrained continuation-ratio logistic regression   Number of obs =     189
                                                        LR chi2(18)   =   56.75
                                                        Prob < chi2   =  0.0000
Log likelihood = -231.27942                             Pseudo R2     =  0.1093

-------------------------------------------------------------------------------
         bwt4 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
2             |
        smoke |   2.40644   1.243078     1.70   0.089     .8743305    6.623299
              |
         race |
       black  |  7.091104   6.725815     2.07   0.039     1.104997    45.50581
       other  |  3.540889   1.810036     2.47   0.013     1.300152    9.643407
              |
          lwt |  1.001661   .0083208     0.20   0.842     .9854846    1.018103
           ht |  .2490112   .3461584    -1.00   0.317     .0163281    3.797531
           ui |    2.8845   2.681902     1.14   0.255     .4662888    17.84375
        _cons |  .3330399   .4015545    -0.91   0.362     .0313459    3.538435
--------------+----------------------------------------------------------------
3             |
        smoke |  1.564183   .6990885     1.00   0.317     .6514121    3.755948
              |
         race |
       black  |  3.602625   2.276972     2.03   0.043     1.043838    12.43383
       other  |  .9405772   .4551569    -0.13   0.899     .3643238    2.428295
              |
          lwt |  .9839369   .0078965    -2.02   0.044     .9685811    .9995361
           ht |  1.114572   1.325293     0.09   0.927     .1083873    11.46141
           ui |  2.314626    1.38722     1.40   0.161     .7150425    7.492551
        _cons |  2.246423    2.51009     0.72   0.469     .2514103    20.07244
--------------+----------------------------------------------------------------
4             |
        smoke |  2.817403   1.105908     2.64   0.008     1.305356    6.080917
              |
         race |
       black  |  3.758631   1.959795     2.54   0.011     1.352705    10.44375
       other  |  2.526023   1.087054     2.15   0.031      1.08675    5.871446
              |
          lwt |  .9834361   .0066887    -2.46   0.014     .9704134    .9966336
           ht |  6.490237   4.483259     2.71   0.007     1.676009    25.13302
           ui |  2.471801   1.106213     2.02   0.043     1.028189    5.942297
        _cons |  1.054066   .9884219     0.06   0.955     .1677556    6.623063
-------------------------------------------------------------------------------
Logit 2 compares bwt4==2 with bwt4 < 2
Logit 3 compares bwt4==3 with bwt4 < 3
Logit 4 compares bwt4==4 with bwt4 < 4
```

The estimated effects of `smoke` on `bwt4` are now

$$\begin{aligned}
\widehat{\mathrm{OR}}(2,1) &= \exp\left(\widehat{\beta}_{2k}\right) = 2.41 \text{ (95\% CI [0.87, 6.62])} \\
\widehat{\mathrm{OR}}(3,1\dots 2) &= \exp\left(\widehat{\beta}_{3k}\right) = 1.56 \text{ (95\% CI [0.65, 3.76])} \\
\widehat{\mathrm{OR}}(4,1\dots 3) &= \exp\left(\widehat{\beta}_{4k}\right) = 2.82 \text{ (95\% CI [1.31, 6.08])}
\end{aligned}$$

As with the adjacent category and constrained continuation-ratio models, the odds of higher values of `bwt4` (lower birthweight) are estimated to be higher for smokers than for nonsmokers. Although the confidence intervals of the three ORs for `smoke` are wide and largely overlapping, the point estimates are quite different, as are those for the other independent variables, which supports the idea that the effects are dependent on the category. It may seem that the constrained continuation-ratio model in section 4.2, where the effects are constrained across the logits, does not fit the data well.

Note that the results for logit 4 are identical to the results obtained with a binary logistic regression model using a dichotomized dependent variable defined as `0` = birthweight > 2500 grams, `1` = birthweight $\leqslant 2500$ grams.

The linear predictions obtained by `predict` with the option `xb` after estimation with `ucrlogit` are specific for each logit. Thus you may request either the linear prediction for one logit using the `outcome` option,

```
. predict xb4, xb outcome(4)
```

or the linear predictions for all logits in one command,

```
. predict xb1-xb4, xb
```

The linear prediction for logit 1 (the reference category) is 0.

5 Concluding remarks

I have presented the estimation commands `adjcatlogit`, `ccrlogit`, and `ucrlogit`, which calculate three ordinal logistic regression models: the adjacent category, the constrained continuation-ratio, and the unconstrained continuation-ratio models, respectively. The models can be used as alternatives to the proportional odds model (`ologit`), for instance, when the proportional odds assumption does not hold or when the comparisons between response categories for these models are more informative for the problem at hand. The continuation-ratio models are particularly useful for the analysis of sequential processes (Agresti 2010, 96–97), where Y measures the number of attempts to attain a binary outcome.

Models estimated with `adjcatlogit`, `ccrlogit`, and `ucrlogit` are all equal to the binary logistic regression model (`logit` or `logistic`) if applied to a binary dependent variable.

Further model-building options for ordered response data are provided by the gologit2 command (Williams 2006). gologit2 fits generalized ordered logistic models that include an unconstrained model with the same number of parameters as the multinomial and unconstrained continuation-ratio models, the proportional odds model, and the partial-proportional odds model. The partial-proportional odds model allows for a subset of the regression coefficients to be constrained across the logits, thus providing a compromise between the restrictive constrained models and the unconstrained models, which often estimate more parameters than necessary.

Wolfe (1998) previously published the command (ocratio) for the constrained continuation-ratio model. As discussed in section 2.2, the model implemented by ocratio is not equivalent to the model implemented by ccrlogit. The model formulation in ccrlogit is equal to the recommended version of the constrained continuation-ratio model in Hosmer, Lemeshow, and Sturdivant (2013).

6 References

Agresti, A. 2010. *Analysis of Ordinal Categorical Data*. 2nd ed. Hoboken, NJ: Wiley.

Ananth, C. V., and D. G. Kleinbaum. 1997. Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology* 26: 1323–1333.

Fagerland, M. W., and D. W. Hosmer. 2014. Tests for goodness of fit in ordinal logistic regression models. Unpublished manuscript.

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

Hosmer, D. W., Jr., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall/CRC.

Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6: 58–82.

Wolfe, R. 1998. sg86: Continuation-ratio models for ordinal response data, *Stata Technical Bulletin* 44: 18–21. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 149–153. College Station, TX: Stata Press.

About the author

Morten W. Fagerland is a senior researcher in biostatistics at the Oslo Centre for Biostatistics and Epidemiology. His research interests include the application of statistical methods in medical research, analysis of categorical data and contingency tables, and comparisons of statistical methods using Monte Carlo simulations.