R Data Analysis Examples: Ordinal Logistic Regression

This page uses the following packages. Make sure that you can load them before trying to run the examples on this page. If you do not have a package installed, run: `install.packages("packagename")`, or if you see the version is out of date, run: `update.packages()`.

```
require(foreign)
require(ggplot2)
require(MASS)
require(Hmisc)
require(reshape2)
```

```
Version info: Code for this page was tested in R version 3.0.2 (2013-09-25)
On: 2014-03-05
With: reshape2 1.2.2; Hmisc 3.12-2; Formula 1.1-1; survival 2.37-4; foreign 0.8-57; knitr 1.5; robustlmm 1.4-2; MASS 7.3-29;
lme4 1.0-6; Matrix 1.1-0; lattice 0.20-24; ggplot2 0.9.3.1
```

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics or potential follow-up analyses.

Examples of ordered logistic regression

Example 1: A marketing research firm wants to investigate what factors influence the size of soda (small, medium, large or extra large) that people order at a fast-food chain. These factors may include what type of sandwich is ordered (burger or chicken), whether or not fries are also ordered, and age of the consumer. While the outcome variable, size of soda, is obviously ordered, the difference between the various sizes is not consistent. The differece between small and medium is 10 ounces, between medium and large 8, and between large and extra large 12.

Example 2: A researcher is interested in what factors influence medaling in Olympic swimming. Relevant predictors include at training hours, diet, age, and popularity of swimming in the athlete's home country. The researcher believes that the distance between gold and silver is larger than the distance between silver and bronze.

Example 3: A study looks at factors that influence the decision of whether to apply to graduate school. College juniors are asked if they are unlikely, somewhat likely, or very likely to apply to graduate school. Hence, our outcome variable has three categories. Data on parental educational status, whether the undergraduate institution is public or private, and current GPA is also collected. The researchers have reason to believe that the "distances" between these three points are not equal. For example, the "distance" between "unlikely" and "somewhat likely" may be shorter than the distance between "somewhat likely" and "very likely".

Description of the Data

For our data analysis below, we are going to expand on Example 3 about applying to graduate school. We have simulated some data for this example and it can be obtained from our website:

```
dat <- read.dta("http://www.ats.ucla.edu/stat/data/ologit.dta")
head(dat)
```

```
##             apply pared public  gpa
## 1     very likely     0      0 3.26
## 2 somewhat likely     1      0 3.21
## 3        unlikely     1      1 3.94
## 4 somewhat likely     0      0 2.81
## 5 somewhat likely     0      0 2.53
## 6        unlikely     0      1 2.59
```

This hypothetical data set has a three level variable called `apply`, with levels "unlikely", "somewhat likely", and "very likely", coded 1, 2, and 3, respectively, that we will use as our outcome variable. We also have three variables that we will use as predictors: `pared`, which is a 0/1 variable indicating whether at least one parent has a graduate degree; `public`, which is a 0/1 variable where 1 indicates that the undergraduate institution is public and 0 private, and `gpa`, which is the student's grade point average. Let's start with the descriptive statistics of these variables.

```
## one at a time, table apply, pared, and public
lapply(dat[, c("apply", "pared", "public")], table)
```

```
## $apply
##
##        unlikely somewhat likely     very likely
##             220             140              40
##
```

```
## $pared
##
##   0   1
## 337  63
##
## $public
##
##   0   1
## 343  57
```

```
## three way cross tabs (xtabs) and flatten the table
ftable(xtabs(~public + apply
```